

Setting Anchor in the Minor Groove: *In Silico* Investigation into Formamido *N*-Methylpyrrole and *N*-Methylimidazole Polyamides Bound by Cognate DNA Sequences

Catharine J. Collar,[†] Moses Lee,[‡] and W. David Wilson*,[†]

Department of Chemistry, Georgia State University, Atlanta, Georgia 30303 and Department of Chemistry, Division of Natural and Applied Sciences, Hope College, Holland, Michigan 49423

Received May 14, 2010

Tricyclic *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) containing polyamide monocations are known to bind as stacked dimers within the minor groove of DNA, and those with *N*-terminal formamido (f) substituents bind in a staggered configuration with high specificity over a range of affinities. Although binding constants have been reported, there is not a clear understanding of why such constants vary significantly for polyamide dimers and their respective cognate DNA sequences. By employing computational tools, the following homodimer complexes have been addressed in this study: f-PyPyIm in complex with 5'-d(GAACTAGTTC)^{-3'}, f-ImPyPy in complex with 5'-d(GAATGCGATTC)^{-3'}, and f-ImPyIm in complex with 5'-d(GAACGCGTTC)^{-3'}. These complexes were selected based on their 10- to 100-fold differences in binding constants. From this study, it was possible to determine how polyamides anchor themselves within the minor groove of specific DNA sequences. This is done through several interactions that provide stability for specific recognition: (i) Py groups secure themselves between DNA base pairs, (ii) lone-pair-Π interactions are formed between DNA deoxyribose O4' and Im groups nearest f, (iii) minor groove bases hydrogen bond to Im groups and amides of the polyamide backbone, (iv) f substituents rotate without leaving the minor groove of DNA and with this rotation form specific hydrogen bonds with electron-rich sites on the floor of the minor groove, and (v) flexible charged *N,N*-dimethylaminoalkyl substituents reside favorably in the minor groove of DNA. Results displayed the greatest amount of interactions and stability for dimer f-ImPyIm in complex with 5'-d(GAACGCGTTC)^{-3'} and the least amount in dimer f-PyPyIm in complex with 5'-d(GAACTAGTTC)^{-3'}. Hence, for cognate DNA sequences, the relative binding strength of compounds was determined as f-ImPyIm > f-ImPyPy > f-PyPyIm. This force-field-based computational study is in agreement with experimental results and provides a molecular rational for the binding constant values.

INTRODUCTION

The antibiotics netropsin and distamycin A, along with synthetic, tricyclic *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) polyamides, bind within the minor groove of cognate DNA sequences with high specificity but with a surprisingly wide range of affinities (1×10^5 – 10^8 M⁻¹).^{1,2} The binding sequence specificity of these compounds follows a well-defined set of rules that have been established and confirmed via experimental techniques.^{1,3–5} Polyamide compounds align antiparallel within the minor groove, tail-to-head and head-to-tail, to form stacked dimers able to recognize specific base pairs: Py overlapped with Py (Py–Py) recognizes adenine (A) thymine (T) base pairs or TA base pairs, Im overlapped with Im (Im–Im) recognizes guanine (G) cytosine (C) base pairs or CG base pairs, Im overlapped with Py (Im–Py) recognizes GC, and Py overlapped with Im (Py–Im) recognizes CG. These relationships are displayed visually through experimental findings, including X-ray diffraction^{5–7} and NMR,^{7,8} for complexes of Im- and Py-containing polyamides with duplex oligodeoxyribonucleotides. As a result of their ability to recognize specific DNA sequences, polyamides are being developed as potential gene

control agents with applications in cancer treatment as well as biotechnology.^{9–14}

Experimental studies have uncovered two structural components of polyamide dimers that significantly affect DNA binding affinity: the *N*-terminal formamido group (f) and the combination and order of the pyrrole and imidazole moieties.^{1,2,6,8,15–22} Compared to nonmodified polyamides, those with an *N*-terminal f substituent displayed increased affinity for sequence-specific binding within the minor groove.^{1,2} This general trend can be illustrated with results for ImPyPy and f-ImPyPy. The binding affinity of f-ImPyPy (Figure 1) with cognate DNA increased by approximately 1×10^2 M⁻¹ compared to ImPyPy with cognate DNA when a terminal f was present. The association and dissociation rates were slower for the f derivative, and the polyamide stacking mode for the complex with f compounds were staggered, while the others were overlapped.² Similar results have also been observed for other nonmodified and f-modified polyamides with their respective cognate DNA sequences.^{1,2}

In silico docking and molecular dynamics studies have also provided valuable insight into DNA–polyamide, and other compound, interactions.^{23–27} For example, the flexible β-Dp tails of the ImH_pPyPy-β-Dp polyamides contributed to binding through water-mediated contact with phosphate oxygen.²⁸ Docking studies have also aided in the construction

* Corresponding author. E-mail: wdw@gsu.edu. Telephone: +1-404-413-5503.

[†] Georgia State University.

[‡] Hope College.

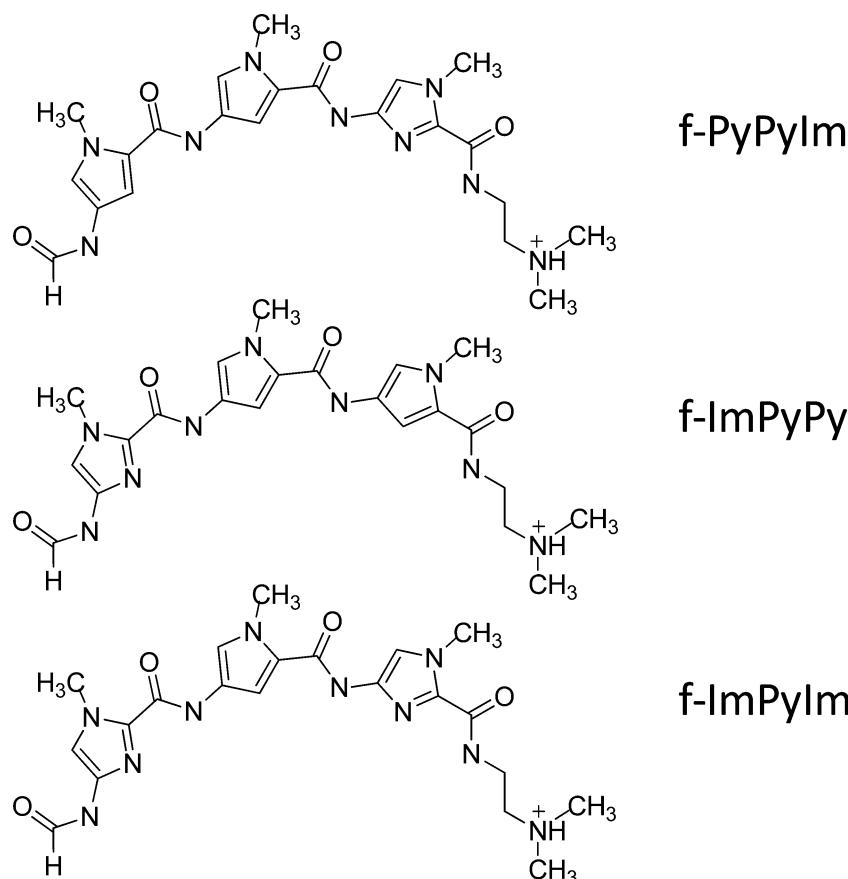


Figure 1. Two-dimensional illustration of polyamide structures (left) with abbreviations (right): formamido (f), *N*-methylpyrrole (Py) and *N*-methylimidazole (Im). Dimer complexes of these compounds are shown docked into cognate DNA sequences in Figures 3–5.

of DNA–polyamide complexes to examine the movements and interactions of individual bases, such as the roll of base pairs when computationally constructed polyamides were examined in complex with DNA.^{24,29} Experimental thermodynamic data have been examined through limited docking of f-ImPyIm (Figure 1) polyamides.¹⁷

The unanswered fundamental question in the experimental studies is why the combination and arrangement of Py and Im moieties have such a significantly different effect on binding affinity with cognate DNAs.^{1,18} Given these observed differences, analyzing experimental binding constants with regard to molecular structure interactions of DNA–polyamide complexes can provide valuable insight. The goal of this study is to use in-depth docking methods to compare and examine how polyamides interact with DNA structure and groups in the minor groove. Through the use of computational tools, several complexes were examined: f-PyPyIm in complex with cognate sequence 5'-d(GAACTAGTTC)^{-3'}, f-ImPyPy in complex with cognate sequence 5'-d(GAATGCATTC)^{-3'}, and f-ImPyIm in complex with cognate sequence 5'-d(GAACCGCGTTC)^{-3'}. These complexes have experimentally determined binding constants of 1×10^6 , 1×10^7 , and 2×10^8 M⁻¹, respectively.^{1,18} This study represents the first in-depth docking approach to examine these polyamides and their cognate sequences to address the experimental affinity variations. Significant differences were found between the strongest and weakest binding polyamide dimers.

METHODS

Polyamides f-PyPyIm, f-ImPyPy, and f-ImPyIm were previously synthesized and examined in complex with 5'-d(GAACTAGTTC)^{-3'}, 5'-d(GAATGCATTC)^{-3'}, and 5'-d(GAACCGCGTTC)^{-3'}, respectively. Surface plasmon resonance (SPR) was employed to determine binding constants.^{1,18}

Docking Preparation. The three polyamide-DNA complexes were evaluated by employing SYBYL 8.1³⁰ software on a Fedora Core 5 Linux workstation. Solution nuclear magnetic resonance (NMR) structure 1B0S³¹ was obtained from the protein data bank; this structure was used as a template and as a reference complex. 1B0S, an f-ImImIm dimer in complex with 5'-d(GAACCCGGTTC)^{-3'}, was mutated using the biopolymer and building and editing modules in SYBYL to form: the f-PyPyIm dimer in complex with 5'-d(GAACTAGTTC)^{-3'}, the f-ImPyPy dimer in complex with 5'-d(GAATGCATTC)^{-3'}, and the f-ImPyIm dimer in complex with 5'-d(GAACCGCGTTC)^{-3'}. These modified complexes were then minimized for 100 iterations using the Tripos force field, thus, allowing the somewhat rigid DNA to accommodate the mutated bases and polyamides through slight changes to the width of the minor groove. Polyamide dimers were then moved to second memory locations, separate molecular areas within the SYBYL graphical user interface. The ability to move, or rather extract, the compounds into a separate molecular area allowed the compounds to explore torsional angles, translation, and rotational angles independent of the DNA when the FlexiDock genetic algorithm was employed. The two compounds of the dimer

Table 1. Energies (E_{MM}) Gained From FlexiDock Docking Studies^a

	E_{MM}
f-PyPyIm with 5'-d(ACTAGT)-3'	-594
f-ImPyPy with 5'-d(ATGCAT)-3'	-670
f-ImPyIm with 5'-d(ACGCGT)-3'	-765

^a These values are in kcal/mol.

were given torsional, translational, and rotational freedom independent of each other; yet, they were docked simultaneously into the DNA. The structures of the polyamides are displayed in Figure 1.

The FlexiDock module of the SYBYL software suite was then implemented. Ten different random starting locations were assigned and employed by the genetic algorithm one at a time for a total of 10 docking trials. Calculated and assigned as in previous studies, the large amount of generations ensured that lowest-energy conformations were obtained.^{32,33} Each docking trial consisted of 516 000 generations. The dimers and the DNA were permitted torsional, rotational, and translational flexibility throughout the docking process. Atomic charges for the DNA were calculated using the Kollman all-atom protocol, while the dimer was assigned Gasteiger–Hückel charges. All possible hydrogen-bonding sites on the dimer and cognate DNA were targeted for function where possible. From each docking, the 20 lowest-energy structures were selected. Hence, 200 structures were produced for each complex examined; 10 random starting locations \times 20 low-energy structures from each docking. The energy values (E_{MM}) for the overall lowest-energy complexes are displayed in Table 1.

Docked Structure Analyses. Interactions were calculated and viewed using modules and tools of the SYBYL software package. The FlexiDock module optimizes torsional angles, translation, and rotational angles to minimize the energy function. Compounds were examined using the FlexiDock scoring function, which is based on the Tripos force field and estimates the energy for the dimer, the receptor, and the complex. The score is evaluated with van der Waals, electrostatic, torsional, and hydrogen-bonding energies; lower energy in the complex state suggests better binding. Hydrogen-bond distances were analyzed using the “display H-bonds” and “measure” options. Values obtained were averaged for the 20 lowest-energy conformations of each complex. All measurements were from heavy atom to heavy atom. The 40 lowest-energy complex conformations displayed variations of the f group. The advanced computation and dock modules, of the SYBYL software suite, were employed to provide further explanation into the Im and Py similarities and differences with respect to electrostatics and dipoles.

Grid Search, an application of the advanced computation module, was used to examine the f groups, of the lowest-energy conformation of each complex, in Cartesian space through systematic rotation about bonds using defined increments of 20° for a total of 360°. At each increment the torsional bond angle was constrained, and the conformation was minimized. The minimization of complexes employed default parameters.³⁰ This allowed for a systematic exploration of torsional freedoms with regard to respective energies. These complexes were then arranged by f group rotation in

increments of 20° and then averaged so that the general trend of the energies could be viewed, analyzed, and compared effectively.

The Dock module of SYBYL 7.3³⁴ was then employed to re-examine structures and calculate energy values for low-energy complexes obtained via FlexiDock and Grid Search. This software was ideal since structures could be observed and energy values could be calculated without changes to DNA–polyamide complex conformations.

Accessible solvent area (ASA) was examined for each lowest-energy complex with Chimera software.³⁵ The module employed was the Area/Volume from Web (StrucTools server) with calculation options Gerstein ASA, surface probe 1.4, and all atoms except water. ASA was acquired for each complex, DNA, and individual polyamide. The ASA was summed for each atom of each residue and polyamide.

The Spartan '04³⁶ software package was employed to examine geometry optimized Py and Im groups using a single point ab initio calculation employing the Hartree–Fock 6-31G** level. This allowed for comparisons of ab initio calculated electrostatic potential maps and dipoles.

RESULTS

The modification of NMR solution structure 1B0S resulted in the construction of the f-PyPyIm dimer in complex with 5'-d(GAACTAGTTC)-3', the f-ImPyPy dimer in complex with 5'-d(GAATGCATTC)-3', and the f-ImPyIm dimer in complex with 5'-d(GAACGCGTTC)-3'. Each of the polyamides underwent extensive docking, as did reference polyamides from 1B0S, within their respective DNAs to yield optimized structures.

Reference complex, 1B0S, displayed only slight deviations from the refined average structure obtained via solution NMR. The calculated root-mean-squared error between the NMR and lowest-energy docked structures was approximately 0.60. Figure 2 displays the alignment of the 10 lowest-energy 1B0S-docked reference structures.

f-PyPyIm in Complex with 5'-d(GAACTAGTTC)-3'. The low-energy structures of f-PyPyIm are hydrogen bonded as a staggered dimer to 5'-d(GAACTAGTTC)-3' (Figure 3). The base pairs involved in hydrogen bonding include those within the center of the DNA, 5'-d(ACTAGT)-3', the AT base pair followed by CG, TA, AT, GC, and TA, respectively. Each image of Figure 3 (right) was taken as the DNA was rotated to the right, so that the bases would take on a view that was as linear as possible.

The hydrogen bond displayed in the top AT base pair is between the upper dimer compound amide NH of the charged polyamide tail and the C2 O of the T base. Both of the hydrogen bonds displayed in the following image of CG also exist between the upper compound of the dimer and the G base subsequent to the T on the same DNA chain. The hydrogen bonds are between the Im N and the G C2 NH₂ and between the following amide NH and the G N3. Because of the staggered stacking, the TA base pair represents the first image in which heterocycles from both compounds of the dimer are present and both are hydrogen bonding. In this image, the upper dimer compound amide NH following the Py is forming a hydrogen bond with the A N3, while the lower compound of the dimer is forming a hydrogen bond between the amide NH above the Py and the T C2 O. The

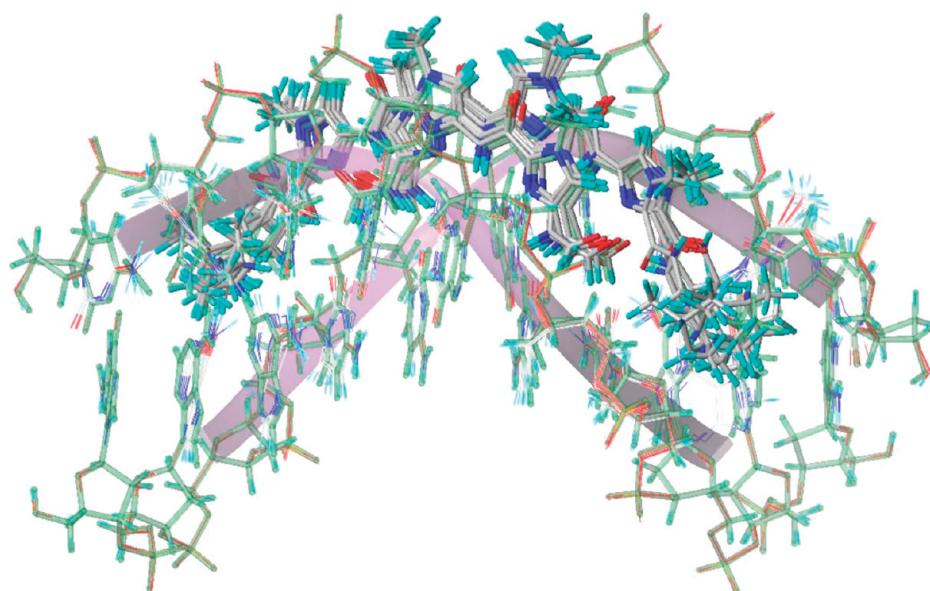


Figure 2. Overlay of the 10 lowest-energy structures for the docking of reference structure, 1B0S,³¹ polyamides into cognate DNA. The refined average structure obtained from the protein data bank is displayed in green, while all other low-energy complexes are displayed by atom type.

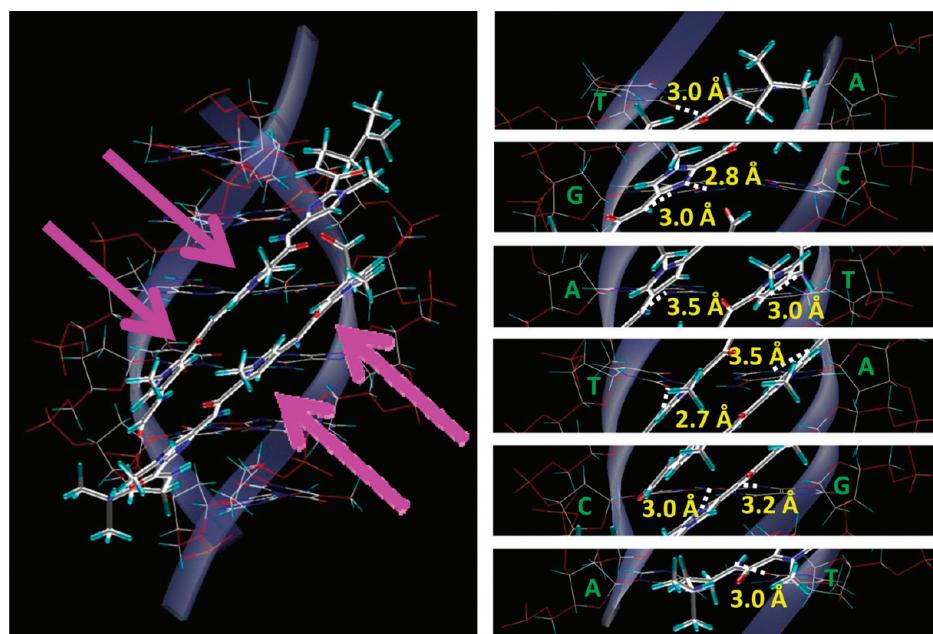


Figure 3. f-PyPyIm in complex with 5'-d(GAACTAGTTC)-3'. For clarity, the terminal bases are not displayed in the images, and the images are of only the lowest-energy conformation. The image on the left displays the complex as a whole, while the segmented images on the right show the individual bases as the polyamide–DNA complex is rotated to the right. Magenta arrows are displayed on the left image; these point to the Py groups. On the right, the bases are labeled in green, the hydrogen bonds are in white, and the average respective hydrogen-bond lengths are in yellow. Notice that the Py groups index themselves with steric complementarity between base pairs, this is pointed out on the left and more clearly viewed in the central TA and AT images on the right.

image of AT and the dimer compounds also displays both compounds forming hydrogen bonds to respective DNA bases. The upper compound amide NH is hydrogen bonded to the T C2 O, while the lower compound amide NH is hydrogen bonded to the A N3. The following image displays a base pair GC which hydrogen bonds with the lower dimer compound. The amide NH prior to the Im is hydrogen bonded to G N3, and the Im N is hydrogen bonded to the G C2 NH₂. The TA base pair, of the last image, is similar to the first base pair AT. In this region, the NH following the Im of the lower dimer compound hydrogen bonds to the T C2 O.

f-ImPyPy in Complex with 5'-d(GAATGCATTC)-3'. Docking results display f-ImPyPy to be hydrogen bonded more favorably to 5'-d(GAATGCATTC)-3' (Figure 4) than the f-PyPyIm dimer described above. Further stabilization is obtained from lone-pair-II interactions between the DNA deoxyribose O4' and the Im groups of the dimer polyamides. Similar to the compounds observed in Figure 3, f-ImPyPy is bound in a staggered dimer conformation, and the base pairs involved in hydrogen bonding include those within the center of the DNA, in this case 5'-d(ATGCAT)-3'. The images of Figure 4 (right) show the AT base pair followed by TA, GC, CG, AT, and TA.

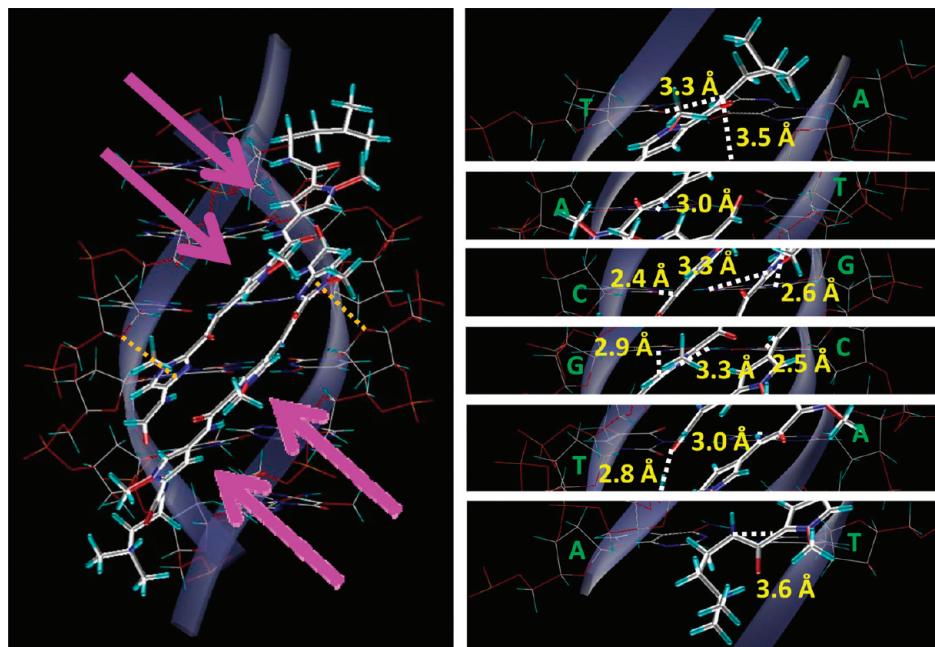


Figure 4. f-ImPyPy in complex with 5'-d(GAATGCATTC)-3'. For clarity, the terminal bases are not displayed in the images, and the images are of only the lowest-energy conformation. The image on the left displays the complex as a whole and identifies lone-pair- Π interactions (orange), while the segmented images on the right show the individual bases as the polyamide-DNA complex is rotated to the right (labeled as in Figure 3).

The hydrogen bonds displayed in the top AT base pair region are between the upper compound amide NH of the charged polyamide tail and the C2 O of the T base and between the upper compound amide NH and the lower compound f O. The hydrogen bond displayed in the following image of TA exists between the upper compound amide NH subsequent to the Py and the A N3. The GC base pair represents the first image in which both compounds of the dimer are visible and both are partaking in hydrogen bonding. In this image, the upper dimer compound amide NH following the Py is forming a hydrogen bond with the C C2 O, while the lower compound of the dimer is forming hydrogen bonds between the f amide NH and the G N3 and the Im N and the G C2 NH₂. The image of CG and the dimer compounds also displays both compounds forming hydrogen bonds to respective DNA bases. The upper compound Im N is hydrogen bonded to the G C2 NH₂, and the following amide NH is hydrogen bonded to the G N3. The lower compound amide NH is hydrogen bonded to the C C2 O. The following image displays base pair AT. Hydrogen bonds displayed in this base pair region exist between the compounds of the dimer and the lower compound amide NH, above the Py, and the A N3. The last image, of the TA base pair, displays a single hydrogen bond between an amide NH and the T C2 O.

f-ImPyIm in Complex with 5'-d(GAACGCGTTC)-3'. Docking results display the f-ImPyIm dimer to be hydrogen bonded to 5'-d(GAACGCGTTC)-3' (Figure 5) even more favorably than either of the previous dimers to their cognate sequences. Similar to what was seen in Figure 4, Figure 5 shows stabilization in lone-pair- Π interactions between the DNA deoxyribose O4' and the Im group nearest the f of the polyamides. As seen in Figures 3 and 4, these compounds take on a staggered dimer conformation, and the base pairs involved in hydrogen bonding include those within the center of the DNA, in this case 5'-d(ATGCAT)-3'. As displayed

in Figure 3 and 4, the images of Figure 5 (right) show the AT base pair followed by TA, CG, GC, CG, GC, and TA.

The hydrogen bond displayed in the top AT base pair region is between the upper compound polyamide charged tail amide NH and the C2 O of the T base. Four hydrogen bonds exist in the following image of CG. These hydrogen bonds are between the upper compound amide NH and the lower compound f O, between the upper compound Im and the G C2 NH₂, between the amide NH subsequent to the Im of the upper compound to the G N3, and between the f O of the lower compound and the G C2 NH₂. The GC base pair represents the first image in which both compounds of the dimer are present and both are partaking in hydrogen bonding to both strands of the DNA. In this image the upper dimer compound amide NH following the Py is forming a hydrogen bond with the C C2 O, while the lower compound of the dimer is forming hydrogen bonds between the amide NH and the G N3 and the Im N and the G C2 NH₂. The image of CG and the dimer compounds also displays both compounds forming hydrogen bonds to respective DNA bases. The upper compound Im N is hydrogen bonded to the G C2 NH₂, and the following amide NH is hydrogen bonded to the G N3. The lower compound amide NH is hydrogen bonded to the C C2 O. The following image displays base pair GC. Similar to the first CG region, four hydrogen bonds are displayed in this base pair region. These hydrogen bonds are between the upper compound f O and the lower compound amide NH, between the lower compound Im N and the G C2 NH₂, between the amide NH prior to the Im of the upper compound to the G N3, and between the f O of the upper compound and the G C2 NH₂. The last image, of the TA base pair, displays a single hydrogen bond between an amide NH and the T C2 O.

Terminal Interactions. Overall docking results for all three complexes display significant flexibility in the polyamide charged tails, as expected, and significant rotation of

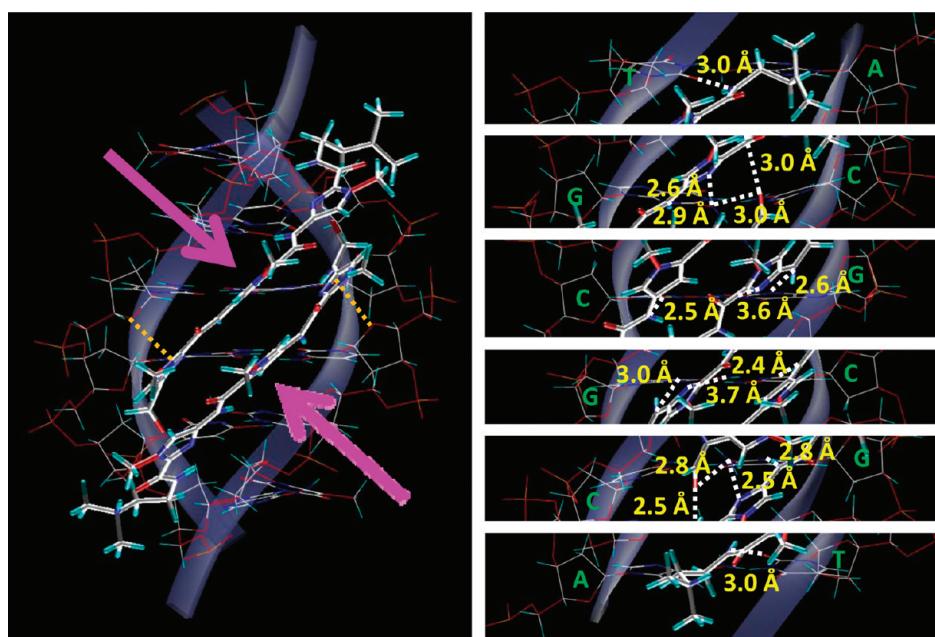


Figure 5. f-ImPyIm in complex with 5'-d(GAACGCGTTC)-3'. For clarity, the terminal bases are not displayed in the images, and the images are of only the lowest-energy conformation. The image on the left displays the complex as a whole, while the segmented images on the right show the individual bases as the polyamide–DNA complex is rotated to the right (labeled as in Figures 3 and 4).

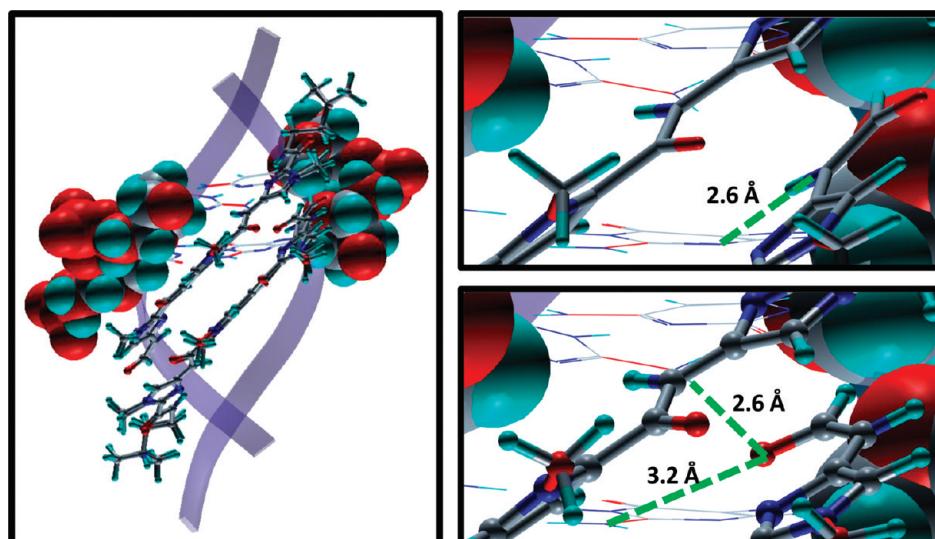


Figure 6. f-ImPyIm in complex with cognate sequence 5'-d(GAACGCGTTC)-3'. The image on the left displays the overlap of the two low-energy dimer conformations in the minor groove of the lowest-energy DNA base pairs affected by f rotation. In an enlarged view for clarity, the images on the right display the two low-energy conformations individually. These conformations consist of different hydrogen-bonding interactions, which are shown in green. The upper right image displays the f N hydrogen bonding to the G N3 of the first GC base pair of the recognition sequence; whereas the lower right image displays the hydrogen bonding of f O to the G C2 NH₂ of the first CG base pair as well as the G C2 NH₂ of the first GC base pair.

the f substituent, which provides unexpected extra insights, while the heterocycles and amide groups exhibit less flexibility. As noted above, the charged tail amide NH forms hydrogen bonds to T C2 O base pairs (Figures 3–45). The remainder of each polyamide charged tail resides favorably within respective cognate DNA minor grooves. Rotation of the f group allows for different hydrogen bonds to form and stabilize the complex. Figure 6 displays such changes within overlaid lowest-energy structures of f-ImPyIm in complex with 5'-d(GAACGCGTTC)-3'. The lowest-energy structure, previously addressed in Figure 5, is shown as capped sticks, while a second low-energy conformation is displayed as ball and stick structures. The f NH of the most common lowest-energy structures obtained from docking displays hydrogen

bonds to G N3 of the first GC base pair of the recognition sequence (Figures 5 and 6, upper right); however, upon rotation of f, this interaction is lost, and new hydrogen bonds are formed between the f O and the G C2 NH₂ of the first CG base pair and the G C2 NH₂ of the first GC base pair (Figure 6, lower right). When this rotation occurs the hydrogen bond between the dimer polyamides, f O and charged tail NH, can no longer form. The rotation of f can occur at either end of the dimer formation. These formations, obtained via FlexiDock, were analyzed further through Grid Search. Figure 7 displays the energy fluctuations as rotation occurs in the f group. In the plot, the lowest-energy complex conformations, with respect to f torsional angles, are for the two hydrogen-bonded conformations in Figure 6.

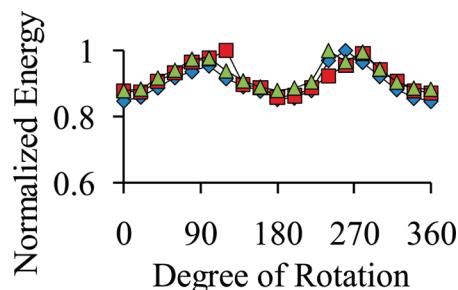
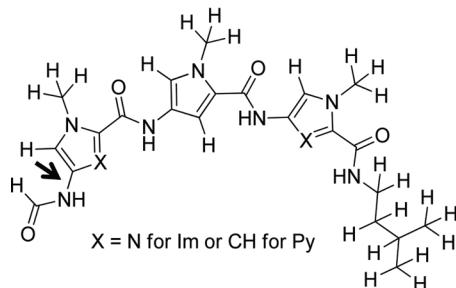


Figure 7. Polyamide structure (left) with an arrow pointing to the bond rotated via Grid Search. X represents CH for Py or N for Im depending on the complex employed for f bond rotation. The graph (right) displays the averaged, normalized energy values for each structure obtained after a 20° rotation of one or both f bonds within the dimer. Data for complexes with f-PyPyIm, f-ImPyPy, and f-ImPyIm, are displayed in blue, red, and green, respectively. The lowest-energy conformations are at 0° and 180°. At 0° the f is positioned as in Figure 6, upper right, and position 180° is shown in Figure 6, lower right.

Table 2. Total Energies, Reported as E_{MM} Values, Gained from Dock for the Lowest-Energy Complexes Obtained via FlexiDock and Grid Search^a

	FlexiDock	Grid Search 1	Grid Search 2
f-PyPyIm with 5'-d(ACTAGT)-3'	-80.1 (-23.5/-56.6)	-103 (-75.3/-28.0)	-99.4 (-72.6/-26.8)
f-ImPyPy with 5'-d(ATGCAT)-3'	-119 (-73.9/-44.9)	-118 (-75.3/-42.6)	-117 (-72.4/-44.8)
f-ImPyIm with 5'-d(ACGCGT)-3'	-130 (-77.9/-52.4)	-126 (-72.7/-53.3)	-126 (-73.6/-52.6)

^a The E_{MM} values for both low-energy structures acquired via Grid Search are labeled 1 and 2, respectively. Grid Search 1 relates to the structures with the f positioned as in the top right image of Figure 6, while Grid Search 2 relates to the structures with the f positioned as in the bottom right image of Figure 6. The steric and electrostatic E_{MM} values are shown, respectively, to the right of the total energies. All values are in kcal/mol.

Complex Energies. Relative total energies for lowest-energy complexes subsequent to FlexiDock and Grid Search, calculated via Dock, are reported in Table 2. Total energies are the sum of steric and electrostatic energies. For complexes from FlexiDock, the steric and electrostatic energies are, respectively, -23.5 and -56.6 kcal/mol for the complex with dimer f-PyPyIm, -73.9 and -44.9 kcal/mol for the complex with f-ImPyPy, and -77.9 and -52.4 kcal/mol for the complex with dimer f-ImPyIm. For each complex acquired via Grid Search, the steric and electrostatic energies are, respectively, -75.3 and -28.0 kcal/mol for the complex with dimer f-PyPyIm, -75.3 and -42.6 kcal/mol for the complex with f-ImPyPy, and -72.7 and -53.3 kcal/mol for the complex with dimer f-ImPyIm. Since the 180° rotation of f occurred in the top 20% of lowest-energy complexes, as viewed in Figure 6, energies for these complexes were also calculated subsequent to Grid Search (Table 2). For each complex acquired steric and electrostatic energies are, respectively, -72.6 and -26.8 kcal/mol for the complex with dimer f-PyPyIm, -72.4 and -44.8 kcal/mol for the complex with f-ImPyPy, and -73.6 and -52.6 kcal/mol for the complex with dimer f-ImPyIm.

Accessible Surface Area (ASA). Buried surface on complex formation was addressed through ASA calculations. Figure 8 displays the surfaces for the complexes, DNA, and polyamides; blue surfaces encompass positively charged regions, while red cover those that are negatively charged. Positive and negative regions of the polyamides can align with those of the DNA minor groove to maximize electrostatic interactions. This was further supported with the ASA values for the complexes, DNA, and polyamide (Figure 9). The red area represents the DNA. Notice that the DNA ASA is fairly similar for all three complexes, as are the three polyamide areas displayed in green. The blue areas show differences related to the respective DNAs binding their specific polyamides for complex formation: (i) the complex with f-PyPyIm displays more ASA at the T bases of the recognition sequence than the other two complexes, (ii) the complex with f-ImPyPy displays a decreased ASA near the A bases of the recognition sequence, and (iii) the complex with f-ImPyIm is the most uniform and consists of the most buried ASA.

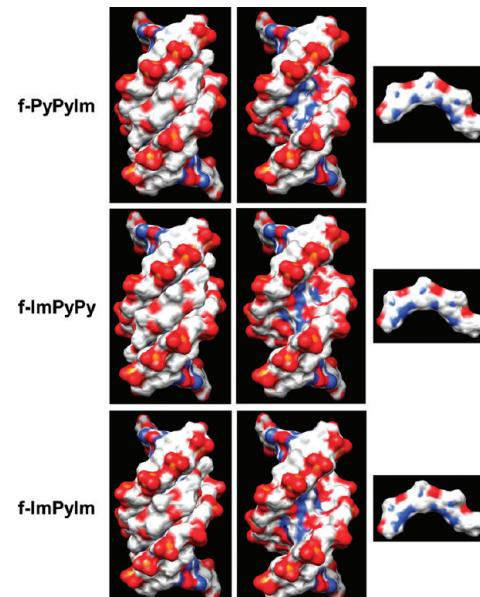


Figure 8. Surfaces displaying electrostatic potentials with respect to Coulombic coloring for the complexes (left), DNA (center), and polyamides (right); blue surfaces encompass positively charged regions, while red cover those that are negatively charged.

with f-PyPyIm displays more ASA at the T bases of the recognition sequence than the other two complexes, (ii) the complex with f-ImPyPy displays a decreased ASA near the A bases of the recognition sequence, and (iii) the complex with f-ImPyIm is the most uniform and consists of the most buried ASA.

Ab Initio Electrostatic Potential Maps. To understand the energy contributions from polyamide structures, it is informative to compare the ab initio calculated electrostatic potential maps for the Py, Im, and amide units of the polyamide dimers (Figure 10) with each other as well as with the low-energy stacked complexes shown in Figure 11.

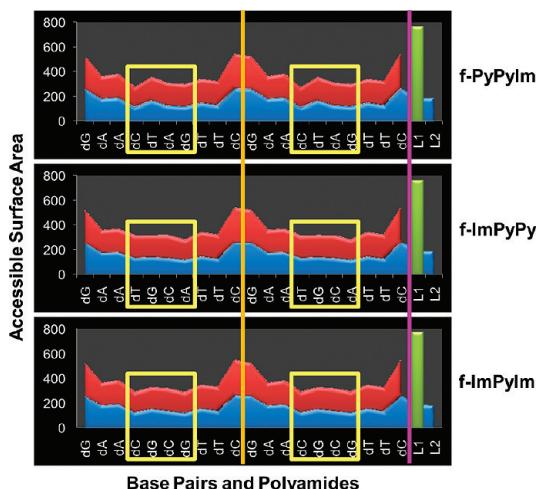


Figure 9. ASA calculated for each base pair and polyamide in complex (blue) and alone (red for DNA and green for single polyamide). The ASA is reported in Å², and the DNA bases are denoted as dA, dT, dG, and dC for adenine, thymine, guanine, and cytosine, respectively. Orange and purple lines spanning the three ASA graphs separate the two DNA strands and the polyamides. Both DNA strands are shown from 5'-3', displaying the differences in each strand with (blue) and without (red) compound interaction. Yellow boxes highlight the specific recognition sites for each complex. The polyamides are displayed as L1 and L2. When analyzing only a single polyamide (green), this compound is L1.

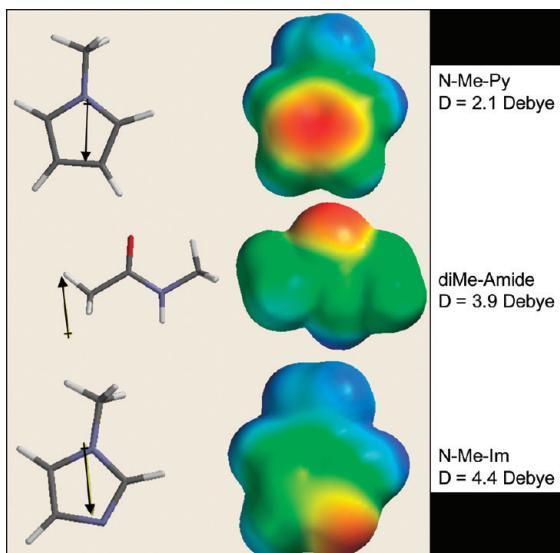


Figure 10. Ab initio calculated electrostatic potential maps for the Py, Im, and amide units of the polyamide dimers, respectively. These units are shown on the left with their dipole moments. The electrostatic potentials are shown in the center, blue is positive and red is negative, and the magnitudes of the dipoles are displayed on the right.

Although the dipole moments of the Py and Im heterocycles point in the same direction, the magnitude of the dipole is larger for the Im, and the electrostatic potential maps clearly show a significantly different distribution of molecular electrostatic potential. With both the Py and Im, the positive potential is distributed on the N-Me group and close vicinity. With the Py, the highest negative potential is on the Py Π -system, while in the Im, it is on the unprotonated Im–N (Figure 10). As expected, the negative potential on the amide is highest on the carbonyl O, while the positive potential is on the –NH. With this distribution, the dipole moment of

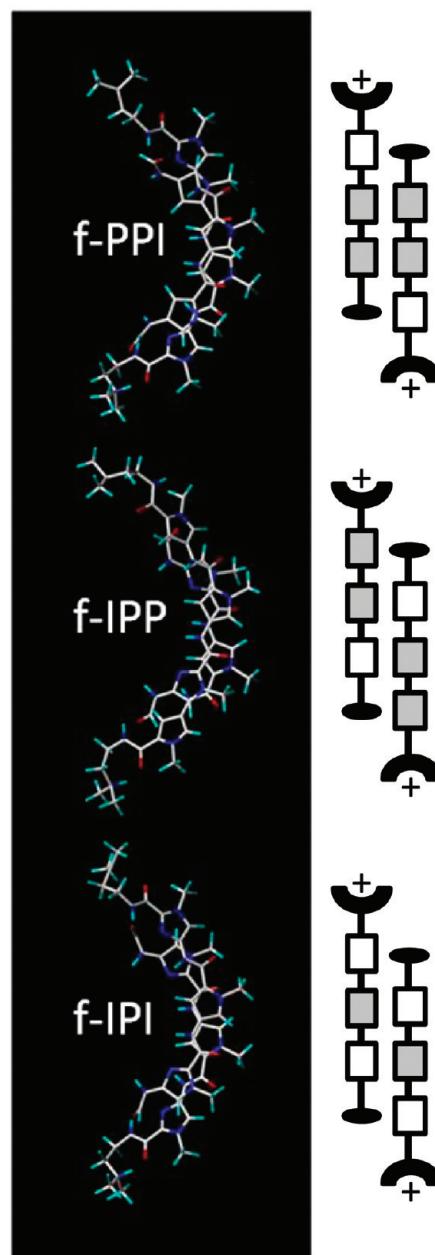


Figure 11. Top view of dimers formed during docking (left) and schematic representation (right) with Py in gray and Im in white. The DNA has been removed so that preferred staggered conformations can be viewed. From top to bottom, the dimers come from f-PyPyIm in complex with 5'-d(GAAACTAGTTC)-3', f-ImPyPy in complex with 5'-d(GAATGCATTC)-3', and f-ImPyIm in complex with 5'-d(GAACGCGTTC)-3'. Notice the spacing of the dimers.

the amide points in the opposite direction to the heterocycles (Figure 10) in the orientation of DNA binding (Figure 11).

Each of the stacked polyamides has six heterocycles that can be evaluated in terms of the maps in Figure 10. Starting with the weakest binder, f-PyPyIm (Figure 11), the heterocycles interact as follows: at top of the figure, (i) the first Im is relatively unstacked; (ii) the next heterocycle (lower molecule of the dimer) is stacked favorably with a positive amide $-NH$ over the negative area of the pyrrole; (iii) the next two pyrroles are stacked such that their positive regions are near negative carbonyl O atoms, a fairly favorable orientation; (iv) The next Py has its most positive region closely stacked with a positive $-NH$, an unfavorable interaction; and (v) the last Im is not well stacked. In the

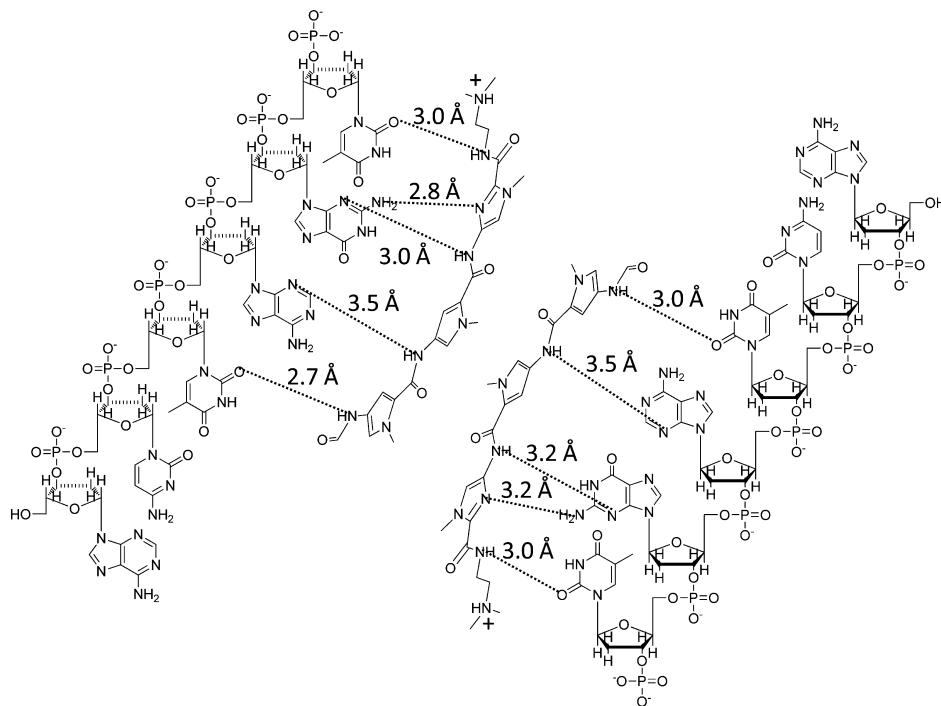


Figure 12. Two-dimensional illustration of f-PyPyIm in complex with cognate sequence 5'-d(GAACTAGTTC)-3'. Hydrogen bonds are displayed by a dashed line with respective distances.

strongest binding f-ImPyIm complex (Figure 11), (i) the first Im is not strongly stacked; (ii) the next Im (lower molecule) is favorably stacked with an amide with the negative carbonyl O of the amide near the most positive region of the Im (Figure 10), and the amide positive H of the $-NH$ is stacked near the negative Im-N; (iii) the next two Py groups are also favorable stacked with amide negative O atoms near their positive N-Me groups; (iv) the next Im is in a similar favorable orientation with an amide O near the positive N-Me of the imidazole, while the positive H of the amide $-NH$ is near the Im negative N; and (v) the last Im is not as strongly stacked as the internal heterocycles. It should be noted, however, that the terminal two Im groups are stacked with their most negative regions near the most positive regions of the adjacent Ims, and given the larger dipole moment of the Im versus Py groups (Figure 10), this should be a favorable contribution. In summary, the electrostatic interactions between the stacked heterocycles appear to make favorable contributions to dimer binding of both f-PyPyIm and f-ImPyIm, but there are more and stronger favorable interactions in the f-ImPyIm dimer.

DISCUSSION

Experimental results for simple tricyclic polyamides, such as those in Figure 1, have a puzzling, large variation in energies when bound by their cognate DNA sequences.^{1,2} We have conducted a docking study for the polyamides of Figure 1 and respective cognate DNA to provide some initial molecular level information on the different complexes. Three components that contribute to polyamide dimer-DNA interactions were investigated in the docked structures: (i) hydrogen bonding, (ii) buried surface on complex formation, and (iii) electrostatic interactions of the polyamide units in the stacked dimer. All of these interactions have been evaluated, and these results provide insight into the large variations in binding constants.

When analyzing the dimer of f-PyPyIm in complex with 5'-d(GAACTAGTTC)-3' and its resulting position due to interactions with the minor groove, it is important that the dimer overlap is a staggered conformation of central Py-Py/Py-Py (Figures 3 and 11). The stacked Py groups form a stable motif with the ability to anchor these polyamides into stable dimer conformations in the minor groove (Figures 2). Figure 12 illustrates the models from Figure 3 in two-dimensions and shows the interactions of the Py-Py stacked motif with the central base pairs of the cognate binding sequence. The Py groups fit between the bases and aid in position indexing so that amide NH groups and the Im N can form favorable hydrogen bonds with base functional groups. The electrostatic potentials also play a significant role in both the specific interactions and complex stabilization (Figures 8 and 9).

The f-ImPyPy dimer polyamides also overlap in a staggered conformation of Py-Im/Im-Py (Figures 4 and 11). As in Figure 3, steric interactions of the stacked Py groups appear to play an important role in anchoring these polyamides within their recognition sequence. Py groups index themselves with steric complementary between the base pairs. The Im groups form favorable hydrogen bonds, due in part to the added stability provided by the steric positioning interactions of the Py groups. The optimum positioning of the Py and Im groups allows the dimer to form hydrogen bonds between the ends of the upper and lower stacked compounds in Figures 4 and 13. Polyamide f-ImPyPy, in complex with 5'-d(GAATGCATTC)-3', displays stacking differences that vary from those of f-PyPyIm in complex with 5'-d(GAACTAGTTC)-3', and these appear to be due to electrostatic interactions between the polyamides and the DNA (Figures 8, 10, and 11). DNA and polyamides were mobile throughout the docking process; the f-ImPyPy polyamides moved into minor groove regions that are more optimal than those of the f-PyPyIm complex (Figures 4 and

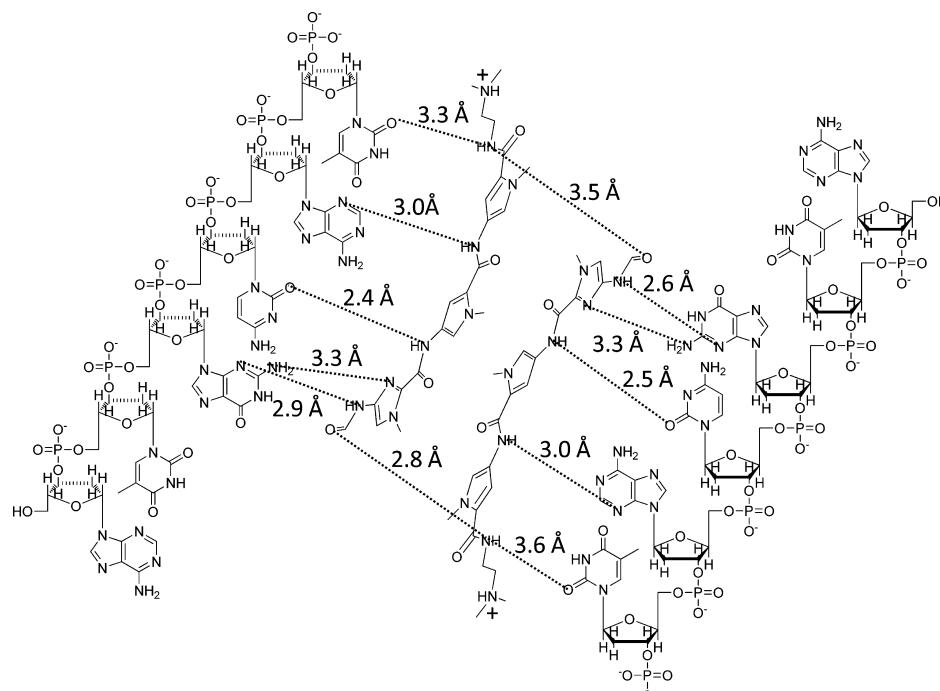


Figure 13. Two-dimensional illustration of f-ImPyPy in complex with cognate sequence 5'-d(GAATGCATTC)-3'. Hydrogen bonds are displayed by a dashed line with respective distances.

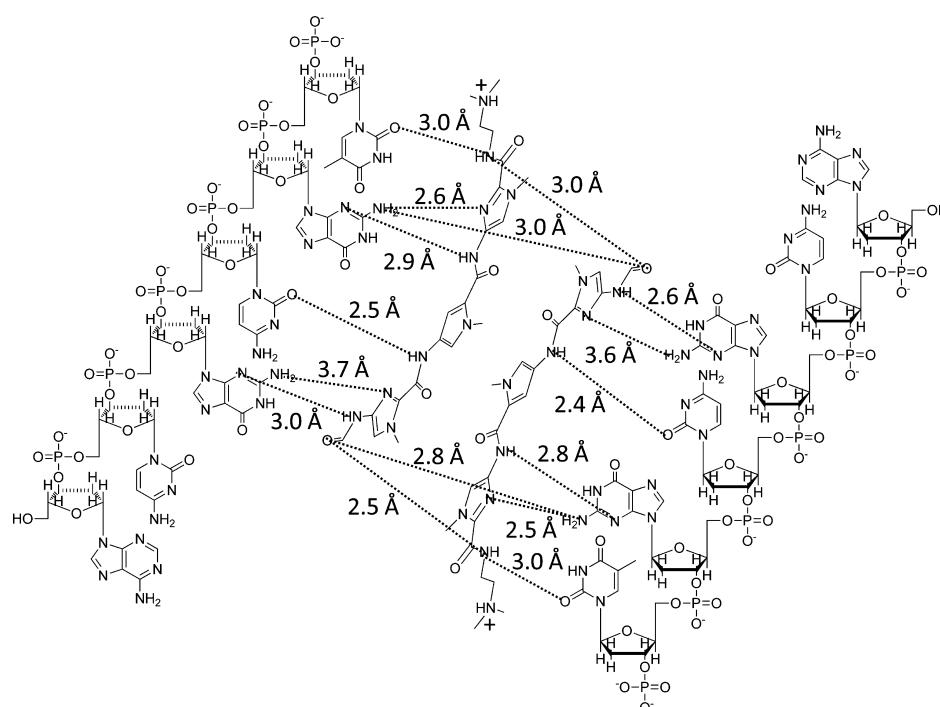


Figure 14. Two-dimensional illustration of f-ImPyIm in complex with cognate sequence 5'-d(GAACCGCGTTC)-3'. Hydrogen bonds are displayed by a dashed line with respective distances.

13). This allows for more favorable interactions and a larger negative calculated energy, E_{MM} (Table 2).

Similar to f-PyPyIm and f-ImPyPy dimers, f-ImPyIm polyamide dimers overlap in their minor groove location in a staggered conformation of Py—Im/Im—Py (Figures 5 and 14). As seen in Figures 3 and 4, the Py groups index themselves between base pairs with steric complementary and play an important role in anchoring the compound into stable low-energy docked conformations with the best possible positioning. The Im groups contribute to the GC base pair recognition and general affinity. The optimum

positioning of the Py and Im groups also allows the dimer to form hydrogen bonds between the charged *N,N*-dimethyl-aminoalkyl tail NH of the upper compound and the f O of the lower compounds. The Im groups on the ends also contribute significantly to the amount of hydrogen bonding. This is by far the most stable of the three structures evaluated, as shown by the wealth of hydrogen bonding and the positioning of the compounds. The terminal Im, not involved in the Py—Im/Im—Py stacking, forms tight hydrogen bonds, and the compounds are pulled in close to the DNA. These interactions are enhanced by favorable electrostatic interac-

tions that reduce the ASA (Figures 8 and 9). The DNA and compounds form tight favorable interactions, and the E_{MM} value for the complex with f-ImPyIm is more negative than the values obtained for complexes with compounds f-PyPyIm or f-ImPyPy.

Given that the compounds of the dimer are binding to the same sequences on opposite DNA strands one may expect the hydrogen bonds to be quite similar. The small differences in observations of individual structures are as expected for flexible docking. The complex with f-PyPyIm does not exhibit hydrogen bonding between the two compounds of the dimer, and these compounds exhibit more mobility within the DNA minor groove (Figures 3 and 12). The hydrogen-bond length similarities in compound binding to respective DNA strand are only found in two locations, at the terminal T and the center A of recognition sequence 5'-d(ACTAGT)-3'. The complex with f-ImPyPy exhibits hydrogen bonding within the dimer, at both ends of the compounds, and this results in a greater amount of consistent hydrogen-bond length similarities between compounds and their respective DNA strands (Figures 4 and 12). The hydrogen-bond length similarities in compound binding to the respective DNA strand are found in three locations central the recognition site, at the G, C, and A of recognition sequence 5'-d(ATGCAT)-3'. The complex with f-ImPyIm exhibits hydrogen bonding within the dimer and to the bases of the parallel DNA strand (Figures 5 and 14). The hydrogen-bond length similarities in compound binding to respective DNA strands are found throughout the recognition site, at the G, C, G, and T of recognition sequence 5'-d(ACGCGT)-3'.

The terminal groups of the polyamides are flanked by a flexible charged tail and a small f substituent. The movements of the charged *N,N*-dimethylaminoalkyl tail were minor in comparison to a previous molecular dynamics study examining polyamides with longer tails.²⁸ The charged N resided toward the center of the minor groove between the phosphates of the DNA backbone. Perhaps, possible interactions with phosphates are limited by the size of the tail and the stable hydrogen bonding of the tail NH with T C2 O.

The rotation of the f group in the stacked complexes is a significant observation in these experiments. The terminal f substituents are small enough to rotate in the dimer widened minor groove without the polyamide leaving the minor groove, and energies obtained suggest that the two orientations shown in Figure 6 contribute to binding (Figures 6, 7 and Table 2). The steric contributions to energy are similar, most likely due to the similar size of all three polyamides and the areas occupied; electrostatics differ much more. It is also important to note that stability of structures resulted in consistency of E_{MM} values, even when f rotated 180°. These results suggest that when a single polyamide begins to deviate from its recognition site, a rotation of f occurs, and new bonds are formed; thus, keeping the complex longer than if the f substituent was absent. This discovery explains our recent observations that modifications of the f group with other acyl groups results in diminished binding affinity.²² Specifically, the order of binding constants was f-ImPyIm ≫ Acetyl-ImPyIm > *N*-methylureidoacetyl-ImPyIm > trifluoroacetyl-ImPyIm. This is consistent with the suggestion that small and planar N-terminus substituents promote favorable binding with DNA. Furthermore, consistent with the role of the f or acyl group, ImPyIm analogs bearing an

NH₂ at the N-terminus and nonformamido-ImPyIm gave the weakest binding indicating the importance of having an f group to form favorable hydrogen bonds with sites on the floor in the minor groove.

CONCLUSIONS

Previously, experimental studies employing surface plasmon resonance (SPR) acquired binding constants for f-PyPyIm in complex with 5'-d(GAACTAGTTTC)-3', f-ImPyPy in complex with 5'-d(GAATGCATTC)-3', and f-ImPyIm in complex with 5'-d(GAACCGTTTC)-3'. These constants are approximately 1×10^6 , 1×10^7 , and $2 \times 10^8 \text{ M}^{-1}$, respectively.^{1,18} In our studies, E_{MM} values were calculated for the lowest-energy complexes obtained via FlexiDock and Grid Search (Table 2); the more negative the value, the stronger the binding. Both the experimental and the in silico data are in agreement. The ranked binding from strongest to weakest is: f-ImPyIm in complex with 5'-d(GAACCGTTTC)-3' > f-ImPyPy in complex with 5'-d(GAATGCATTC)-3' > f-PyPyIm in complex with 5'-d(GAACTAGTTTC)-3'.

This in-depth docking approach provides useful new molecular information about polyamide complexes and how they are anchored within the minor groove. Hydrogen bonding and steric and electrostatic interactions all play a role, along with compound conformation, to determine how a compound will recognize specific DNA sequences. Specifically, f-ImPyIm binds better than the other dimers as a result of the greater amount of intradimer and intracomplex hydrogen bonds and lone-pair-Π and optimum dipole interactions as well as excellent steric fit and electrostatic interactions. We are currently employing these findings to improve compound design. Findings suggest that dimer spacing provided by Py groups and hydrogen-bonding interactions of Im groups can be employed to recognize even longer DNA sequences. This of course is given that recognition compounds: (i) keep a curvature that parallels that of DNA, (ii) stack efficiently maintaining electrostatic interactions, and (iii) have the ability to form hydrogen bonds on both ends. Insights from this study suggest that compounds, such as f-ImPyImPyIm, should bind and recognize 5'-d(-ACGCGCGT)-3 with similar affinity and greater specificity than f-ImPyIm binds and recognizes 5'-d(GAACCGTTTC)-3'. These studies are in progress, and the results will be reported in due course.

ACKNOWLEDGMENT

We thank the National Science Foundation (grants CHE 0550992 and 0809162) for support (to M.L. and W.D.W.), along with the Georgia State University Molecular Basis of Disease Fellowship (to C.J.C.).

REFERENCES AND NOTES

- (1) Buchmueller, K. L.; Staples, A. M.; Uthe, P. B.; Howard, C. M.; Pacheco, K. A.; Cox, K. K.; Henry, J. A.; Bailey, S. L.; Horick, S. M.; Nguyen, B.; Wilson, W. D.; Lee, M. Molecular recognition of DNA base pairs by the formamido-pyrrole and formamido/imidazole pairings in stacked polyamides. *Nucleic Acids Res.* **2005**, *33* (3), 912–921.
- (2) Lacy, E. R.; Le, N. M.; Price, C. A.; Lee, M.; Wilson, W. D. Influence of a terminal formamido group on the sequence recognition of DNA by polyamides. *J. Am. Chem. Soc.* **2002**, *124* (10), 2153–2163.
- (3) White, S.; Baird, E. E.; Dervan, P. B. On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* **1997**, *4* (8), 569–578.

- (4) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **1998**, *391* (6666), 468–471.
- (5) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. Structural effects of DNA sequence on T.A recognition by hydroxypyrrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* **2000**, *295* (3), 557–567.
- (6) Chenoweth, D. M.; Dervan, P. B. Allosteric modulation of DNA by small molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (32), 13175–13179.
- (7) Yang, X. L.; Hubbard, R. B.; Lee, M.; Tao, Z. F.; Sugiyama, H.; Wang, A. H. Imidazole-imidazole pair as a minor groove recognition motif for T:G mismatched base pairs. *Nucleic Acids Res.* **1999**, *27* (21), 4183–4190.
- (8) Zhang, Q.; Dwyer, T. J.; Tsui, V.; Case, D. A.; Cho, J.; Dervan, P. B.; Wemmer, D. E. NMR structure of a cyclic polyamide-DNA complex. *J. Am. Chem. Soc.* **2004**, *126* (25), 7958–7966.
- (9) Hochhauser, D.; Kotecha, M.; O'Hare, C.; Morris, P. J.; Hartley, J. M.; Taherbhai, Z.; Harris, D.; Forni, C.; Mantovani, R.; Lee, M.; Hartley, J. A. Modulation of topoisomerase IIalpha expression by a DNA sequence-specific polyamide. *Mol. Cancer Ther.* **2007**, *6* (1), 346–354.
- (10) Chou, C. J.; Farkas, M. E.; Tsai, S. M.; Alvarez, D.; Dervan, P. B.; Gottesfeld, J. M. Small molecules targeting histone H4 as potential therapeutics for chronic myelogenous leukemia. *Mol. Cancer Ther.* **2008**, *7* (4), 769–778.
- (11) Lai, Y. M.; Fukuda, N.; Ueno, T.; Matsuda, H.; Saito, S.; Matsumoto, K.; Ayame, H.; Bando, T.; Sugiyama, H.; Mugishima, H.; Serie, K. Synthetic pyrrole-imidazole polyamide inhibits expression of the human transforming growth factor-beta1 gene. *J. Pharmacol. Exp. Ther.* **2005**, *315* (2), 571–575.
- (12) Matsuda, H.; Fukuda, N.; Ueno, T.; Tahira, Y.; Ayame, H.; Zhang, W.; Bando, T.; Sugiyama, H.; Saito, S.; Matsumoto, K.; Mugishima, H.; Serie, K. Development of gene silencing pyrrole-imidazole polyamide targeting the TGF-beta1 promoter for treatment of progressive renal diseases. *J. Am. Soc. Nephrol.* **2006**, *17* (2), 422–432.
- (13) Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G., Jr.; Dervan, P. B. Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (48), 16768–16773.
- (14) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. Improved nuclear localization of DNA-binding polyamides. *Nucleic Acids Res.* **2007**, *35* (2), 363–370.
- (15) Bremer, R. E.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. Recognition of the DNA minor groove by pyrrole-imidazole polyamides: comparison of desmethyl- and N-methylpyrrole. *Bioorg. Med. Chem.* **2000**, *8* (8), 1947–1955.
- (16) Brown, T.; Mackay, H.; Turlington, M.; Sutterfield, A.; Smith, T.; Sielaff, A.; Westrate, L.; Bruce, C.; Kluza, J.; O'Hare, C.; Nguyen, B.; Wilson, W. D.; Hartley, J. A.; Lee, M. Modifying the N-terminus of polyamides: PyImPyIm has improved sequence specificity over f-ImPyIm. *Bioorg. Med. Chem.* **2008**, *16* (9), 5266–5276.
- (17) Buchmueller, K. L.; Bailey, S. L.; Matthews, D. A.; Taherbhai, Z. T.; Register, J. K.; Davis, Z. S.; Bruce, C. D.; O'Hare, C.; Hartley, J. A.; Lee, M. Physical and structural basis for the strong interactions of the -ImPy- central pairing motif in the polyamide f-ImPyIm. *Biochemistry* **2006**, *45* (45), 13551–13565.
- (18) Buchmueller, K. L.; Staples, A. M.; Howard, C. M.; Horick, S. M.; Utche, P. B.; Le, N. M.; Cox, K. K.; Nguyen, B.; Pacheco, K. A.; Wilson, W. D.; Lee, M. Extending the language of DNA molecular recognition by polyamides: unexpected influence of imidazole and pyrrole arrangement on binding affinity and specificity. *J. Am. Chem. Soc.* **2005**, *127* (2), 742–750.
- (19) Mackay, H.; Brown, T.; Utche, P. B.; Westrate, L.; Sielaff, A.; Jones, J.; Lajiness, J. P.; Kluza, J.; O'Hare, C.; Nguyen, B.; Davis, Z.; Bruce, C.; Wilson, W. D.; Hartley, J. A.; Lee, M. Sequence specific and high affinity recognition of 5'-ACCGCT-3' by rationally designed pyrrole-imidazole H-pin polyamides: thermodynamic and structural studies. *Bioorg. Med. Chem.* **2008**, *16* (20), 9145–9153.
- (20) Melander, C.; Herman, D. M.; Dervan, P. B. Discrimination of A/T sequences in the minor groove of DNA within a cyclic polyamide motif. *Chemistry* **2000**, *6* (24), 4487–4497.
- (21) O'Hare, C. C.; Utche, P.; Mackay, H.; Blackmon, K.; Jones, J.; Brown, T.; Nguyen, B.; Wilson, W. D.; Lee, M.; Hartley, J. A. Sequence recognition in the minor groove of DNA by covalently linked formamido imidazole-pyrrole-imidazole polyamides: effect of H-pin linkage and linker length on selectivity and affinity. *Biochemistry* **2007**, *46* (42), 11661–11670.
- (22) Westrate, L.; Mackay, H.; Brown, T.; Nguyen, B.; Kluza, J.; Wilson, W. D.; Lee, M.; Hartley, J. A. Effects of the N-terminal acylamido group of imidazole- and pyrrole-containing polyamides on DNA sequence specificity and binding affinity. *Biochemistry* **2009**, *48* (24), 5679–5688.
- (23) Holt, P. A.; Chaires, J. B.; Trent, J. O. Molecular docking of intercalators and groove-binders to nucleic acids using Autodock and Surflex. *J. Chem. Inf. Model.* **2008**, *48* (8), 1602–1615.
- (24) Ge, W.; Schneider, B.; Olson, W. K. Knowledge-based elastic potentials for docking drugs or proteins with nucleic acids. *Biophys. J.* **2005**, *88* (2), 1166–1190.
- (25) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* **1998**, *282* (5386), 111–115.
- (26) Wellenzohn, B.; Loferer, M. J.; Trieb, M.; Rauch, C.; Winger, R. H.; Mayer, E.; Liedl, K. R. Hydration of hydroxypyrrrole influences binding of ImHpPyPy-beta-Dp polyamide to DNA. *J. Am. Chem. Soc.* **2003**, *125* (4), 1088–1095.
- (27) Cashman, D. J.; Buscaglia, R.; Freyer, M. W.; Dettler, J.; Hurley, L. H.; Lewis, E. A. Molecular modeling and biophysical analysis of the c-MYC NHE-III1 silencer element. *J. Mol. Model.* **2008**, *14* (2), 93–101.
- (28) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. Complex of B-DNA with polyamides freezes DNA backbone flexibility. *J. Am. Chem. Soc.* **2001**, *123* (21), 5044–5049.
- (29) Correa, B. J.; Canzio, D.; Kahane, A. L.; Reddy, P. M.; Bruice, T. C. DNA sequence recognition by Hoechst 33258 conjugates of hairpin pyrrole/imidazole polyamides. *Bioorg. Med. Chem. Lett.* **2006**, *16* (14), 3745–3750.
- (30) SYBYL Molecular Modeling Software, 8.1 ed.; Tripos Inc.: St. Louis, MO, 2008.
- (31) Yang, X. L.; Kaenzig, C.; Lee, M.; Wang, A. H. Binding of AR-1–114, a tri-imidazole DNA minor groove binder to CCGG sequence analyzed by NMR spectroscopy. *Eur. J. Biochem.* **1999**, *263* (3), 646–655.
- (32) Munde, M.; Ismail, M. A.; Arafa, R.; Peixoto, P.; Collar, C. J.; Liu, Y.; Hu, L.; David-Cordonnier, M. H.; Lansiaux, A.; Baily, C.; Boykin, D. W.; Wilson, W. D. Design of DNA minor groove binding diamidines that recognize GC base pair sequences: a dimeric-hinge interaction motif. *J. Am. Chem. Soc.* **2007**, *129* (44), 13732–43.
- (33) Liu, Y.; Collar, C. J.; Kumar, A.; Stephens, C. E.; Boykin, D. W.; Wilson, W. D. Heterocyclic diamidine interactions at AT base pairs in the DNA minor groove: effects of heterocycle differences, DNA AT sequence and length. *J. Phys. Chem. B* **2008**, *112* (37), 11809–18.
- (34) SYBYL Molecular Modeling Software, 7.3 ed.; Tripos Inc.: St. Louis, MO, 2006.
- (35) UCSF Chimera; Regents of the University of California: San Francisco, CA, 2007.
- (36) Spartan '04, Wavefunction Inc.: Irvine, CA, 2004.

CI100191A