AD_____

Award Number: DAMD17-99-1-9323

TITLE: Antigene Strategy in Breast Cancer Therapy: Rationales for Direct Targeting of erbB2/Her2 DNA with Polymides

PRINCIPAL INVESTIGATOR: Vsevolod Y. Katritch, Ph.D.

CONTRACTING ORGANIZATION: The Scripps Research Institute
La Jolla, California 92037

REPORT DATE: September 2000

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20010330 085

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2000	Annual Summary (1 Sep 99 - 31 Aug 00)

4. TITLE AND SUBTITLE
Antigene Strategy in Breast Cancer Therapy: Rationales for Direct Targeting of erbB2/Her2 DNA with Polymides

5. FUNDING NUMBERS
DAMD17-99-1-9323

6. AUTHOR(S)
Vsevolod Y. Katritch, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Scripps Research Institute
La Jolla, California 92037

E-MAIL:
sevak@scripps.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

We have developed a complex computational approach to identify the best fragments within erbB2/HER2/neu promoter suitable for targeting with a novel class of Pyrrole-Imidazole containing polyamide molecules and to design the most potent polyamide binders to these targets. The target identification takes into account the most important regulatory elements in the promoter; whole-genome specificity of the targets and possible mutations in the target sequences. We have designed a variety of conformationally feasible polyamide molecules, matching the best candidate 11-14 bp DNA target sequences.

We have developed an algorithm, based on the ICM molecular modeling package, to build all-atom 3D models of various polyamide-DNA complexes with different sequences and polyamide topology, and to predict polyamide-DNA binding constants. We continue to improve and benchmark the algorithm with new experimental results, such as a series of experimentally measured affinity data and NMR structural data. This fast and reliable algorithm will help to identify the best polyamide candidate antigene inhibitors to be synthesized and tested during the next stage of the project.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
34

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

Tyrosine kinase receptor erbB2/HER2/neu oncogene, a key component in the epidermal growth factor (EGF) signaling pathway, is amplified and upregulated in 25-30% of human breast cancers and is associated with poor clinical prognoses [Perou et al. 2000]. Specific inhibition of the gene on the transcriptional level (antigene strategy) would have a high therapeutic potential. We suggest using a novel class of Pyrrole-Imidazole (Py-Im) containing polyamides to bind specific DNA sequences in the erbB2 promoter region in order to disrupt formation of the transcription complex. The polyamides have been demonstrated to be highly effective and sequence specific dsDNA binders with decent cell permeability and recently tested as erbB2 inhibitors [Chiang et al. 2000]. The major aims of our research are (i) to apply sequence analysis tools to identify the most promising short targets within erbB2 DNA promoter sequence and (ii) to design optimal polyamide molecules that bind these dsDNA targets.

Body

Task1: Optimization of target sequences in gene Her2/erbB-2 promoter.

The sequence of the erbB2 gene promoter contains well-characterized TATAA and CCAAT boxes, repetitive GGA motif and putative SP1 binding sequences in the region upstream to the major transcription start site, see Figure 1. Despite TATA presence, multiple transcription start sites have been found, the major ones being 21 and 70 bp down from the TATA box. It was shown that the 500bp region upstream of the major starting site is sufficient for both basal and inducible transcription activity, the most proximal 125bp DNA stretch being responsible for about 30-fold overexpression in most cancer cell lines [Scott et al. 1994].

a. We performed a comprehensive database analysis, based on the specialized MatInspector tool [Quandt et al 1995], to find putative regulatory elements in the 500 bp promoter. Table 1 lists the results of this search for the most important 150 bp proximal region. Most sites, found and characterized previously, were identified in the search (these entries are emphasized both in Table 1 and Figure 1). For example, the ETS response element next to the TATAA box [Scott et al 1994], as well as AP-2 binding site [Bosher et al., 1995], CCAAT box, were identified.

Based on the analysis presented in Table 1 we listed 6 short 16 bp sequences, flanking transcription factor binding sites, see Figure 1. Note that four of these sequences overlap with more that one major activation site, which makes them the most interesting targets for antigene therapy.

b. Recent availability of the human genome sequence gives us an opportunity to predict the specificity of a polyamide binder on a whole genome level. We designed a specialized program to perform exhaustive BLAST-based searches in the human genome draft to assign sequence specificity of a particular binding pattern. We performed both searches for exact sequence matches, as well as a simple sequence profile search with low penalty for A-T substitution. The latter approach was devised to take into account full degeneracy of Py-Py recognition of A-T pair and partial degeneracy of Pyrrole-Hydrohypyrrole (Py-Hp) recognition of A-T. Using this program, we assigned the

specificity to all possible 11, 12, 13 and 14 bp fragments within preselected target sequences. Figure 2 demonstrates an example result of our analysis in the case of 13 bp fragments.

c. Analysis of the 5 available versions of the erbB2 promoter sequences from different sources demonstrated that the region from -120 to 0 is completely identical in all sequences, while some deletions-insertions are possible in the farther upstream sequence. Conservation of the target sequence is crucial for development of effective antigene inhibitors, so we plan to repeat this analysis when more cell culture- and tissue-specific sequences of erbB2 promoter are available.

d. We sorted all the short fragments (~ 130 of them) based on the sequence specificity score, length and overlap with core activation sites. This analysis gave several nontrivial insights. First, the regions around TATAA box (sequences **5** & **6**), though very important for regulation of gene activity, may not be the best targets for polyamide binding, since they both have very poor specificity profile. In addition, sequence **6** is very AT rich, which further lowers its polyamide specificity score. On the other hand, sequences **1**, **2**, and **4** contain 13 bp fragments with almost unique whole-genome specificity, and each of them overlap with more than one activation site.

Task 2: Overall design and evaluation of complimentary polyamides.

a-b. Using a set of polyamide elements and polyamide-DNA pairing rules [Wemmer & Dervan 1997; de Clairac et al., 1999; Herman et al, 1999], see Table 1, we devised an algorithm to build all matching polyamide sequences for each target dsDNA. The algorithm starts by building a "perfect match" sequence that contains Py, Im and Hp rings only and performs all possible substitutions to allow various types of topology suggested in the proposal. Additional empirical rules are also applied to eliminate unfeasible designs, e.g. only 2 to 4 subsequent rings are allowed, β-alanines should be isolated, only 4 γ-links are allowed, and so on. With these restrictions applied, the program automatically generates as many as ~30-50 different polyamides for each 13 bp DNA sequence or ~20-30 polyamides for 12 bp DNA. We performed this procedure with the best 50 DNA targets from our target list and stored the results in a database. The feasibility of chemical synthesis was checked for the resulting structures.

c. The central part of our project is 3D modeling of the resulting DNA-polyamide complexes and evaluation of their relative affinity. Our original algorithm uses the fact that polyamide complexes with DNA are very modular in structure. This allows us to build initial conformations of new complexes, based on known X-ray geometries of previously characterized complexes [Kielkopf et al. 1998ab]. The program tethers DNA and ligand residues to the respective residues in the X-ray structure. These initial conformations are subsequently optimized by restrained energy minimization, where energy terms include bonded, van der Waals, electrostatic and hydrogen bonding terms. The application of geometry restraints enforces DNA-DNA base-pairing and DNA-polyamide pairing rules in the initial stage of the optimization, forcing the model to

follow the "canonical" pattern of polyamide-DNA recognition [Kielkopf et al. 1998ab]. In the final stage, the restraints are removed and free global energy minimization is applied. The deviation between restrained and free energy minimized models is usually within all-atom RMSD < 1.5 A for "match" polyamide-DNA complexes, which suggest high quality of the modeling. Single polyamide mismatches increase this RMSD to ~2-3A, thus reflecting big deviations of the fully energy-optimized model from the "canonical" recognition pattern.

The polyamide-DNA binding energy of the models was estimated in terms of van der Waals, hydrogen bonding, electrostatic and solvation contributions. The accuracy of relative binding energy predictions is about 1.5 kcal/mol, estimated by comparison with more than 50 published experimental measurements. This accuracy is satisfactory for the preliminary assignment of the affinity of newly designed polyamides, though we plan further improvement by using a more elaborated molecular force field.

The polyamide-DNA modeling algorithm was presented at the Program in Mathematics and Molecular Biology meeting last year (see the abstract attached) and was significantly upgraded recently to accommodate new variants of polyamide topology and improve affinity estimations.

A manuscript on target identification and polyamide design will be prepared for publication by November 2000.

Task 3 : **Detailed modeling and selection of candidate structures**

Recently we started collaboration with Prof. David Wemmer (UC Berkeley) and his structural biology group who specialize in polyamide synthesis and NMR studies of polyamide-DNA complexes [Wemmer & Dervan 1997]. A modified version of our algorithm, accounting for NOESY distance restraints was used to study novel polyamide-DNA complexes. This work confirmed the quality of our model, which fully satisfies most NMR restraints (some expected deviations were found only in the flexible "tail" region of the polyamide) and its usefulness in fast NMR-based 3D structure determination. The manuscript, describing this joint modeling-NMR study will be submitted for publication in October 2000, a draft version is attached here.

We plan to continue this collaboration with Prof. David Wemmer group to synthesize and test the affinity of the best candidate polyamide inhibitors of erbB2 transcription.

Key Research Accomplishments

- We have found the most important candidate targets for antigene therapy within the proximal erbB2 promoter
- We have estimated the whole-genome specificity of all possible short fragments within this promoter region
- We have designed an automatic algorithm to list all possible polyamide topologies matching a given DNA sequence.
- We have written a program, generating 3D model of a polyamide-DNA complex from its "sequence", based on the known pattern of polyamide-DNA recognition

and on global geometry optimization
- We have benchmarked and optimized our predictions of polyamide-DNA binding affinity, using available experimental data
- We have tested the quality of our 3D models in a joint modeling-NMR study

Reportable outcomes

- Meeting Presentation and Abstract:
Katitch, V., Abagyan, R.A. and Olson, W.K. (1999).
Structural Modeling of Polyamide-DNA Recognition.
Mathematics and Molecular Biology VI, Santa Fe, NM.

- Article:
The modularity of DNA recognition by polyamide molecules persists for a ten-ring hairpin in complex with an eight base pair binding site.
Bernhard H. Geierstanger, Colin J. Loweth, Vsevold Katritch, Ruben Abagyan, Peter G. Schultz & David E. Wemmer.
Manuscript to be submitted before October 15, 2000.

Conclusions

During the first year of our effort, we have mostly accomplished Tasks 1 and 2 (months 1-15) of the approved Statement of Work, and started to obtain some interesting result for Task 3. We have identified the best candidate targets for polyamide binding within the most important proximal region of the erbB2 gene promoter and sorted them according to their whole-genome specificity and overlap with transcription activation sites. We have also devised a procedure to find all cell-culture and tissue-specific mutations in these sequences, this work to be continued upon the availability of new erbB2 data in genomic databases.

Using an original automated procedure we have built all conformationally and chemically possible polyamides matching the target dsDNA sequences, according to the polyamide-DNA pairing rules.

Finally, we have designed a fast and reliable algorithm to build 3D models of these polyamides-DNA complexes, based on the known modular structure of the complexes and all-atom conformational energy minimization. The affinity of the DNA - polyamide binding can be predicted by our method with an accuracy of ~1.5 Kcal/mol, which significantly narrows the search for the best candidate polyamides for future *in vitro* and *in vivo* experiments. The accuracy of our modeling has also been confirmed by experimental NMR restraints.

References:

Bosher, J.M. and Williams, T.Hurst, H. The developmentally regulated transcription factor AP-2 is involved in c-erbB-2 overexpression in human mammary carcinoma. *Proc Natl Acad Sci U S A* **92**, 744-7 (1995).

Chen, Y. and Fischer, W.H.Gill, G. Regulation of the ERBB-2 promoter by RBPJkappa and NOTCH. *J Biol Chem* **272**, 14110-4 (1997).

Chiang, S.Y. and Burli, R.W.Benz, C.C.Gawron, L.Scott, G.K.Dervan, P.B.Beerman, T. Targeting the ets binding site of the HER2/neu promoter with pyrrole-imidazole polyamides. *J Biol Chem* **275**, 24246-54 (2000).

de Clairac, R.P.L., Seel, C.J., Geierstanger, B.H., Mrksich, M., Baird, E.E., Dervan, P.B., Wemmer, D.E. NMR Characterization of the Aliphatic / Pairing for Recognition of A·T/T·A Base Pairs in the Minor Groove of DNA *J. Am. Chem. Soc.* **121**, 2956-2964 (1999).

Herman,D.M., James M. Turner, Eldon E. Baird, and Peter B. Dervan Cycle Polyamide Motif for Recognition of the Minor Groove of DNA *Journal of American Chemical Society* **121**, 1121-1129 (1999).

Kielkopf, C.L., Baird, E.E., Dervan, P.B. and Rees, D.C. Structural basis for G.C recognition in the DNA minor groove. *Nat Struct Biol* **5**, 104-9 (1998).

Kielkopf, C.L., White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., Dervan, P.B. and Rees, D.C. A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* **282**, 111-5 (1998).

Perou, C.M. and Sorlie, T.Eisen, M.Bvan, de, Rijn, M.Jeffrey, S.S.Rees, C.A.Pollack, J.R.Ross, D.T.Johnsen, H.Akslen, L.A.Fluge, O.Pergamenschikov, A.Williams, C.Zhu, S.X.Lonning, P.E.Borresen-Dale, A.L.Brown, P.O.Botstein, . Molecular portraits of human breast tumours. *Nature* **407**, 747-52 (2000).

Quandt, K. Frech, K. Karas, H. Wingender, E. and Werner, T. MatInd and MatInspector - New fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Research 23, 4878-4884 (1995)

Wemmer, D.E. and Dervan, P.B. Targeting the minor groove of DNA. *Curr Opin Struct Biol* **7**, 355-61 (1997).

Name of family/matrix	Further Information	Position	Strand	Core sim.	Matrix sim.	Sequence
V$SP1F/GC_01	GC box elements	-148 - -135	(+)	0.876	0.790	gctgGGAGttgccg
V$LYMF/TH1E47_01	Thing1/E47 heterodimer	-134 - -119	(-)	1.000	0.910	aacgaagtCTGGgagt
V$CMYB/CMYB_01	c-Myb	-120 - -103	(+)	1.000	0.949	ttggaatgcaGTTGgagg
V$VMYB/VMYB_02	v-Myb	-113 - -105	(-)	0.819	0.899	tccAACTgc
V$COMP/COMP1_01	COMP1	-89 - -66	(-)	1.000	0.781	tcctgtgATTGggagcaagcg cgc
V$PCAT/CAAT_01	cellular and viral CCAAT box	-82 - -71	(+)	1.000	0.890	tgctcCCAAtca
V$ECAT/NFY_01	nuclear factor Y (Y-box binding factor)	-82 - -67	(+)	1.000	0.920	tgctcCCAAtcacagg
V$VDRF/VDR_RXR_B	VDR/RXR heterodimer site	-69 - -55	(+)	1.000	0.906	aggagaagGAGGagg
V$VDRF/VDR_RXR_B	VDR/RXR heterodimer site	-57 - -43	(+)	1.000	0.892	aggtggagGAGGagg
V$AP2F/AP2_Q6	activator protein 2	-51 - -40	(-)	0.857	0.772	agCCCTcctcct
V$ETSF/ETS1_B	c-Ets-1 binding site	-36 - -22	(+)	1.000	0.910	tgaGGAAgtataaga
V$TBPF/TATA_C	Retroviral TATA box	-30 - -21	(+)	0.843	0.779	agTATAAGAa
V$NFKB/NFKB_Q6	NF-kappaB	-8 - 5	(-)	1.000	0.830	agGGGAatctcagc
V$NOLF/OLF1_01	olfactory neuron-specific factor	-1 - 20	(-)	1.000	0.822	ctccggTCCCaatggagggga a

Table 1. Sequence analysis for 600 bp promoter fragment containing the major transcriptional start site (position 0), CCAAT and TATAA boxes, ETS response element and other potential targets for antigene therapy.

	G•C	C•G	T•A	A•T
Im/Py, Im/β	+	-	-	-
Py/Im, β/Im	-	+	-	-
Hp / Py	-	-	+	-
Py / Hp	-	-	-	+
Py/Py, β/Py, Py /β	-	-	+	+
γ-linker (R)^{H2N} γ-linker	-	-	+	+
β, β/β	-	-	+	+

Table 2. Polyamide-DNA pairing rules. Along with Pyrrole (Py), Imidazole (Im) and Hydrohypyrrole (Hp) rings, other elements include □-alanine, which can stack with any ring or with itself to provide some flexibility, as well as two types of □-links, used as flexible "connectors" linking opposite polyamide strands.

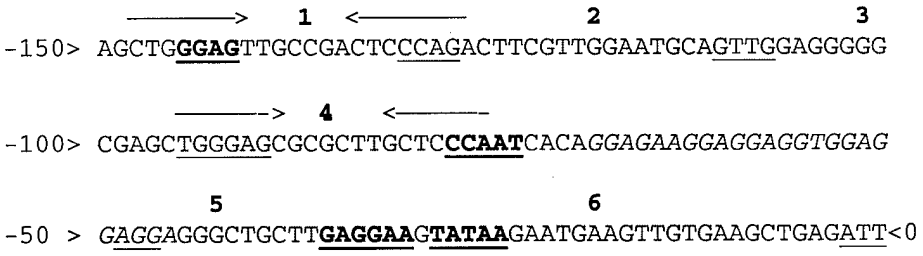

Figure 1. Sequence of the proximal region of erbB2 promoter. Core activation sites are underscored, arrows show two palindromic sequences [Chen et al., 1997] involved in transcription activation. We have highlighted and numbered 16-bp sequences, chosen as putative targets for further analysis.

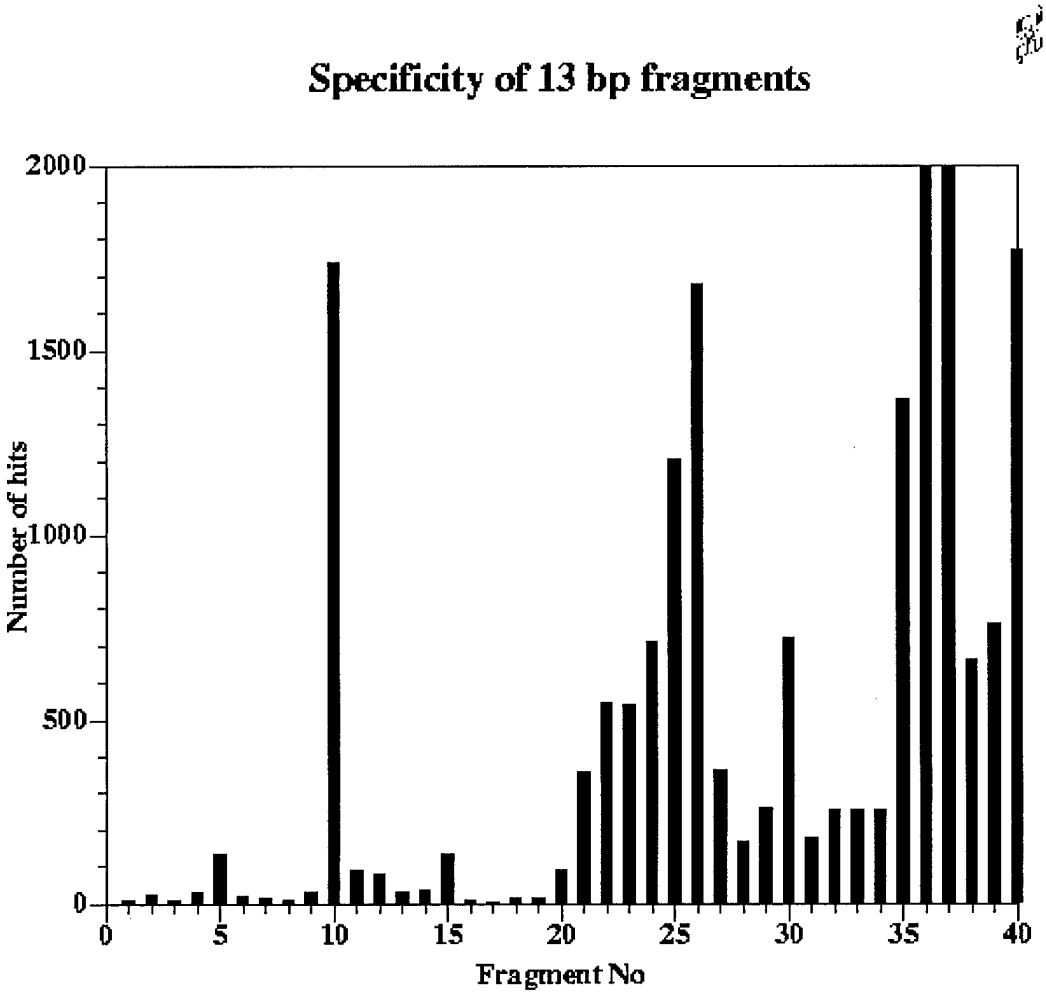

Figure 2. Whole-genome specificity analysis for 13 bp fragments of the proximal erbB2 promoter sequence. Note that the most rare fragments 1-3, 8, 16-17 correspond to sequences **1**, **2** and **4** respectively (see Figure 1), while fragments in the region flanking TATA box 21-30 have very poor specificity, comparable to the specificity of the control fragment with a GGA repeat (35-40)

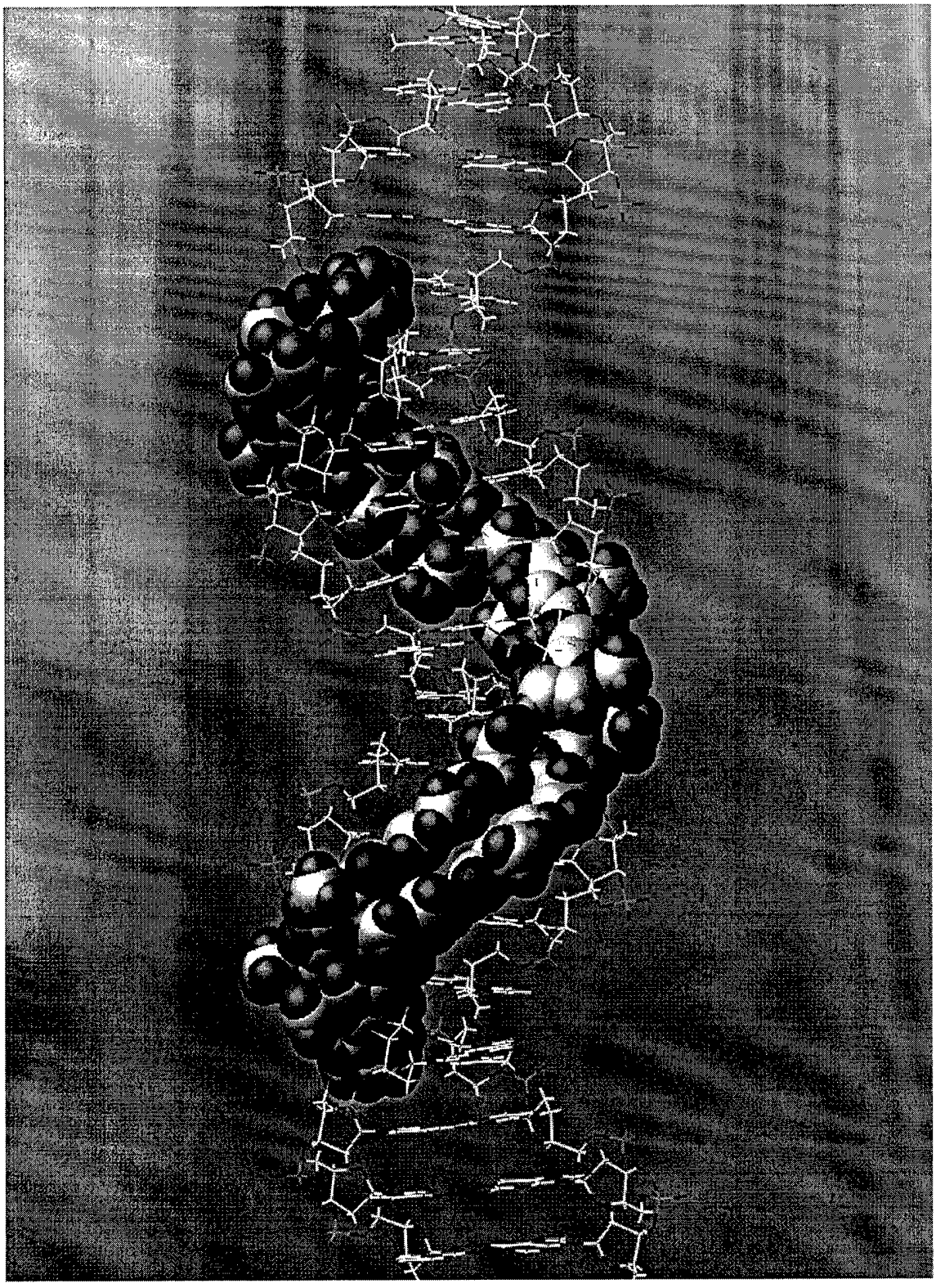

Figure 3. Recognition of a target DNA sequence **AGCGCGCTTGCT** by two sequence-specific polyamide hairpins, each containing 8 Im-Py rings .

Structural Modeling of Polyamide-DNA Recognition

V. Katritch[1], R. A. Abagyan[2], and W. K. Olson[1]

[1] Department of Chemistry, Rutgers University, Piscataway, NJ 08854.

[2] Skirball Institute of Biomolecular Medicine, NYU Medical Center, NY, NY 10016.

A novel generation of synthetic compounds, pyrrole-imidazole containing polyamides, use an effective base-pair recognition code to bind the B-DNA minor grove with affinity and specificity comparable to native transcription factors [1]. Further improvements in the rational design of polyamide drugs rely on understanding the structural details of the drug-DNA interactions.

Here we report a comprehensive procedure for all-atom molecular mechanics modeling of polyamide-B-DNA complexes, build on the basis of the ICM software package [2]. The program provides a means to manipulate polyamide building blocks ("residues"), search effectively for the global energy minimum of the DNA-polyamide complexes and evaluate binding energy accurately in terms van der Waals, hydrogen bonding, electrostatic and solvation contributions. The fine-tuning of the model parameters has been performed with the currently available polyamide-DNA structures (*PDB*: 365d, *NDB*: bdd002, bdd003). The X-ray data are also used as templates for the initial conformations of complexes with various DNA and polyamide sequences. The calculated energy of drug binding is compared with the corresponding binding constants, measured experimentally. (Supported by NIH grants GM20861 and CA77433 and Burrough Wellcome funding from PMMB).

1. Kielkopf CL, White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB, Rees DC (1998). A structural basis for recognition of A·T and T·A base pairs in the minor groove of B-DNA. *Science* **282**: 111-5.

2. Abagyan R.A., Totrov M.M. and Kuznetsov D.N. (1994). ICM- a new method for protein modeling and design. Applications to docking and structure prediction from the distorted native conformation. *J.Comp.Chem*; **15**:488-506.

The modularity of DNA recognition by polyamide molecules persists for a ten-ring hairpin in complex with an eight base pair binding site

Bernhard H. Geierstanger[1], Colin J. Loweth[2], Vsevold Katritch[2], Ruben Abagyan[1,2], Peter G. Schultz[1,2] & David E. Wemmer[3]*

[1]Genomics Institute of the Novartis Research Foundation, 3115 Merryfield Row, San Diego, CA 92121-1125

[2]Department of Chemistry and Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037

[3]Department of Chemistry, University of California, Berkeley, CA 94720

Abstract: Polyamides containing *N*-methylimidazole (Im), *N*-methylpyrrole (Py) and *N*-methylhydroxypyrrole (Hp) amino acids recognize DNA through specific contacts in the minor groove. In a side-by-side arrangement of polyamide ring residues one pair of stacked residues specifically interacts with a single base pair. Pairing rules to specifically recognize all four base pairs have been developed. Commonly used polyamide ligands consist of three or four ring residues linked via a hairpin residue to a second set of three or four rings followed by two tail residues. We use 2D NOESY data combined with restrained molecular modelling to, for the first time, characterize the binding of a ten-ring hairpin polyamide to its eight base pair target site. The high modularity of the polyamide-DNA complexes allowed us to develop a computer script for the molecular modelling program ICM to quickly generate starting models for NMR refinements from the geometry of polyamide residues in previously studied complexes. This is illustrated for the case of the ten-ring hairpin ligand Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp bound to d(GGAATAGTCTGC) ·d(GCAGACTATTCC): NOE restrained molecular modelling indicates a complex consistent with the rules discovered previously. Broadening of NMR resonance lines of the first and the tenth ring residue that are stacked on top of each other indicate conformational exchange in this part of the complex. However, overall the geometric complementarity of ligand and DNA seems to be preserved.

Introduction

Polyamides containing *N*-methylimidazole (Im), *N*-methylpyrrole (Py) and *N*-methylhydroxypyrrole (Hp) amino acids have emerged as designed DNA ligands of high affinity and specificity.[1-12] These molecules recognize the minor groove of DNA through an antiparallel side-by-side arrangement of pairs of polyamide ring residues.[3a] Pairing rules to specifically recognize all four base pairs have been developed:[2-5,12a-d] Im opposite Py targets a G•C base pair while a Py/Im pair targets G•C.[2a,12a,b] A Py/Py combination is selective for A/T base pairs but can not distinguish between T•A and A•T base pair.[2,3] An Hp/Py pair however, can discriminate T•A from A•T.[5] As demonstrated by high-resolution NMR[12] and X-ray[13] structural studies hydrogen bonding between the imidazole ring nitrogen and the amino group of guanosine or between the OH of hydroxypyrrole and the carbonyl of thymidine form the molecular basis for base-specific DNA recognition by polyamides. There is a strict one pair of polyamide residues per one base pair correlation and this modularity has allowed for the successful design of ligands that recognize a variety of sequences.[1-12] The affinity of polyamide ligand in side-by-side dimeric complexes increases from three to four to five ring pairs.[6a] Six and seven ring pairs have similar binding affinities as five ring pairs but the specificity of the complexes is reduced significantly.[6a] If the ligand is extended further the affinity decreases dramatically because the curvature of the ligand does not perfectly match the canonical geometry of B-form DNA.[6a,13c] The resulting size limitation of the DNA target site can be overcome by replacing ring residues with flexible β-alanine residues allowing the ligand geometry to fall back into register with the DNA geometry.[6d,14]

Commonly used polyamide ligands consist of three or four ring residues linked via an γ-aminobutyric residue (γ) to a second set of three or four rings followed by two tail residues. In this side-by-side "hairpin" motif[8,13e] a ligand with N rings will bind to a 0.5*N+3 base pair site with affinities of up to 10^9 M^{-1}.[6b,1,15] Eight-ring hairpin polyamides have been shown to be cell permeable and to compete with natural transcription factors in cellular assays.[15] When coupled to a peptide derived from the activation domain of Gcn4 eight-ring hairpin ligands can act as small molecule transcriptional activators.[16] The molecular structures of hairpin polyamide DNA

complexes has so far been only probed by NMR spectroscopy and these studies were limited to six-ring hairpins.[12e,17] Here we combine NMR and molecular modelling to characterize the binding of a ten-ring polyamide hairpin to its eight base pair DNA target site.

Material and Methods

Synthesis of polyamide molecule. The polyamide molecule Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp was synthesized using solid-phase chemistry and purified as described previously.[11] The γ-aminobutyric acid-Im and Py-Im dimers were synthesized in solution prior to being used in solid phase synthesis. d(GGAATAGTCTGC) and d(GCAGACTATTCC) were purchased from Operon, Inc. and used without further purification.

NMR sample preparation. Equimolar amounts of DNA oligonucleotides d(GGAATAGTCTGC) and d(GCAGACTATTCC) were mixed and annealed. Aliquots of a polyamide ligand stock solution in water were stepwise added to a DNA duplex solution in 10 mM sodium phosphate buffer in 95% H_2O/5% D_2O at pH 7. The progress of the titration was monitored by 1D NMR spectroscopy. The final concentration of the 1:1 hairpin/DNA complex sample was approximately 2 mM (in 500 μl). For experiments in D_2O the sample was repeatedly lyophilized from 99.9% D_2O and finally redisolved in 100% D_2O.

NMR spectroscopy. All NMR spectra were acquired on a DPX 400 Bruker NMR instrument (Bruker Instruments, Billerica, MA) equipped with a ^1H-^{13}C SEI probe. 1D proton and 2D NOESY spectra in 95% H_2O/5% D_2O were acquired using a 1-1-jump-and-return-echo pulse sequence[18] with Z-gradients for water suppression. For 2D NOESY spectra typically 512 t_1 experiments with 128 scans were accumulated with a recycling delay of 2 s. For assignment purposes a NOE mixing time of 200 ms was used with a 70 μs 1-1-jump-and-return delay for maximum excitation of the imino proton region. To resolve assignment problems because of overlapping resonances NOESY spectra were acquired at 15, 25 and 35° C. Ligand and DNA proton resonances were assigned according to established procedures[19,20] or as previously described.[12] Peak

volumes were measured in a 100 ms mixing time NOESY acquired in 95% H_2O/5% D_2O with a 140 μs 1-1-jump-and-return delay τ using XWINNMR and corrected for the $\sin^3(2\pi\Delta\nu*\tau)$ excitation profile of the pulse sequence[18] (where Δν is the frequency offset in Hz from the center of the spectrum). Cross-peaks were classified relative to cytosine H5/H6 cross-peaks into five categories: 1.7-2.7 Å, 2.2-3.2 Å, 2.7-3.7 Å, 3.2-4.2 Å and 3.7-5.0 Å. Additional distance restraints were derived from 100 ms NOESY data in D_2O. Restraints involving ligand *N*-methyl groups or strongly overlapped cross-peaks were set to 1.5-5.0 Å.

Molecular modelling script. Standard geometries of polyamide residues in DNA complexes were derived from the X-ray structures of the polyamide Im-Hp-Py-Py-β-Dp (Pdb entry 407D) in complex with $d(CCAGTACTGG)_2$.[5b] Each ring residue and additional residues for hairpin linker and tail regions of polyamide ligands were parameterized for the molecular modelling program ICM 2.8[21] (Molsoft, L.L.C., Metuchen, NJ). After defining DNA target sequence and hairpin ligand sequence the ICM script tethers DNA and ligand residues to the respective residues in the X-ray structure of the model and overlays the two models followed by energy minimization.

Molecular modelling using restraint energy minimization with NMR-derived distance restraints. Standard B-form DNA and a starting structure for Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp was built in ICM (Molsoft) running on a Windows-NT personal computer. The ligand was energy-minimized with selected intramolecular distance restraint, then manually docked into the binding site followed by cartisian energy minimization and 10000 steps of ICM torsion minimization with 81 intermolecular DNA-ligand and 45 intramolecular ligand-ligand restraints. Additional distance restraints were introduced for the Watson-Crick base pairing hydrogen bonds. This NMR-derived model was compared with a model derived with the modelling script described above and a third model in which the modelling script-derived model was subjected to 10000 steps of ICM torsion energy minimization using the same NMR-derived restraints.

Results and Discussion

NMR characterization of a ten-ring hairpin complex. The polyamide Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp and d(GGAATAGTCTGC)·d(GCAGACTATTCC) form a well-defined hairpin-DNA complex with a 1:1 ligand/duplex stoichiometry as indicated by one-dimensional NMR spectra acquired during a titration (data not shown). The complex dissociates slowly on the NMR time scale and was further characterized by two-dimensional NOESY spectroscopy. NOE contacts between ligand amide NH and *N*-methylpyrrole ring protons with ribose H1' and adenine H2 protons place the polyamide ligand into the minor groove of DNA (Figure 1, Figure 2 and Table 1). The orientation of the ligand and the stacking arrangement (Figure 1) follows the rules previously established for polyamide-DNA recognition:[1-12] Starting with the N-terminal *N*-methylpyrrole residue Py1 abut of T5, the polyamide ligand extends toward its C-terminal contacting DNA in a 5' to 3' direction. The *N*-methylimidazole Im3 specifically recognizes G7 via a hydrogen bond between the imidazole nitrogen and the guanine amino group. Py4 and Py5 contact T8 and C9 respectively, followed by the γ-aminobutyric acid linker γ6 adjacent to the T10·A15 base pair. Completing the hairpin arrangement, the guanine amino group of G16 (base-paired with C9) is recognized by Im7, and Py8 through Py11 contact A17 through A20. NOE contacts by protons of β-alanine β12 and of the dimethylaminopropyl tail residue Dp13 with DNA resonances of A4·T21 and A3·T22 (Figure 3 and Table 1) indicate that the ten-ring hairpin polyamide covers an eight base-pair DNA binding site. The H5 proton of one *N*-methylpyrrole or imidazole ring shows a strong NOE to the *N*-methyl protons of the respective ring residue it stacks on top (Figure 3).[12d,e] Nine of the ten possible interresidue H5 to *N*-methyl NOEs (Figure 3, Supplementary Material: Table 2) verify the stacking of all five pyrrole and imidazole residue pairs as schematically drawn in Figure 1. Additional intramolecular ligand-ligand NOE contacts (Supplementary Material: Table 2) further confirm the overall structure of the polyamide-DNA hairpin complex.

NMR evidence for ligand-DNA hydrogen bonds. Intermolecular ligand-DNA hydrogen bonds play an important role in the recognition of DNA by polyamides. Most

polyamide amide NH protons form hydrogen bonds with DNA hydrogen bond acceptor groups on the minor groove edge of the nucleobases.[12,13] These hydrogen bonds are reflected in the wide range of amide proton chemical shifts observed in these complexes. Hydrogen bonds typically result in a downfield shift of amide proton resonances.[20] In the ten-ring hairpin complex all amide resonances other than in tail and linker residues have chemical shifts lager than 9.4 ppm (Figure 2 and Supplementary Material: Table 3), at least 0.5 ppm higher than in the unbound ligand (amide proton chemical shift in the parent compound distamycin: 8.86 ppm). The sequence specificity of polyamide complexes can be engineered and *N*-methylimidazole residues specifically recognize the guanine amino group exposed in the minor groove of G•C base pairs by forming a hydrogen bond between the imidazole nitrogen and the amino proton not participating in Watson-Crick base pairing.[2a-c,12a-d] For the currently studied ten-ring hairpin complex the two guanine amino protons of each of the two guanines are magnetically identical and at 25° C resonate at 7.91 ppm for G7 and 8.73 ppm for G16, respectively. These chemical shift values are in the range found for the hydrogen-bonded amino proton of cytosines in G•C base pairs.[22] This suggests that for G16 and for G7 both amino protons are involved in hydrogen bonds, one in Watson-Crick base pairing, the other with the imidazole nitrogen of the polyamide ligand. Chemical shifts are not only determined by hydrogen bonding and must be interpreted with caution. For example, as in previous complexes DNA H4' close to the ligand ring residues are upfield shifted, some to chemical shifts as low as 1.74 ppm (Supplementary Material: Table 4), because of ring current effects.[12d,e] However, DNA guanine amino groups are typically not observed at all because of exchange with solvent and because of line broadening caused by rotation around the N-C bond.[22b-c] The chemical shift and the detection of the guanine amino groups in the ten-ring hairpin complex, therefore, strongly support the presence of specific hydrogen bonds to the imidazole nitrogens. For the G7 amino group additional evidence is obtained from NOE contacts to Py4-HN, Py3-HN as well as Py9-HN, Py9-H3 and Py10-HN (Figure 2). Similar NOE contacts were however not observed for the amino group of G16 at 25° C. Weak NOE peaks between G16 amino protons (at a single resonance line) and Py7-HN and Py5-HN were observed in a 2D NOESY spectra acquired at 15° C. As discussed

below, the position and distance of the interacting ligand imidazole nitrogen may explained these observations.

NMR evidence for molecular motions and differences in polyamide residue stacking. Resonance lines of *N*-methylpyrrole residue Py1 and Py11 are significantly broader that corresponding resonance lines of other residues (Figure 2 and 3). The amide proton lines of Py2, Py11, β12, Dp13 and γ6 are significantly broadened also (Figure 2). Compared to the previously studied Im-Py-Py-γ-Py-Py-Py-Dp hairpin complex[12e] intraresidue NOE contacts in the hairpin linker γ6 are broader than expected. At the other end of the ligand, intraresidue NOEs between the aliphatic protons of Dp13 and NOE contacts of these protons to DNA resonances are not observed presumably because of line broadening. Tendative assignments of the aliphatic Dp13 protons was only possible because of broad NOE cross-peaks to the Dp13-HN amide proton. Line broadening of selected resonance in the ten-ring polyamide hairpin complex suggests conformational exchange at both ends of the ligand.

We have previously observed line broadening of selected resonance in polyamide complexes because of conformational exchange. This has been particularly apparent for the tail residue[12d], and was also observed for the pairing of two glycine residues in the side-by-side Im-Py-Py-gly-Py-Py-Py-Dp dimer.[14] In the current complex conformational exchange involves not only the β-alanine residue and the tail group but propagates to the last ring residue that pairs with the first ring of the ten-ring hairpin ligand. This resembles observations made very recently for the DNA complexes of Im-Py-Py-γ-Py-Py-Py-gly-Dp and Ac-Im-Py-Py-γ-Py-Py-Py-β-Dp hairpin ligands:[17] In these complexes conformational exchange on the millisecond time scale is observed between the standard arrangement of the Im1/Py7 ring pair and a conformation in which the ring of Py7 is flipped by 180°. While the tail group looses all contacts with the DNA minor groove in the latter conformation the Py7-NCH_3 protons give rise to NOE contacts with minor groove DNA protons.[17]

For the ten-ring hairpin complex it is clearly the Py11-H5 resonance line that is most exchange broadened (Figure 3) suggesting that the magnetic environment of the Py11 ring protons is changed most dramatically in the exchange process as would be expected

if the outer edge of the Py11 ring now faces toward the bottom of the DNA groove. A ring flip of Py1 can be excluded because the expected cross-peaks between Py1-NCH_3 and Py11-H3 and between Py1-NCH_3 and A20 H2 are not observed. However, weak cross-peaks of Py11-NCH_3 to A4 H2 as well as to Py1-H3 are detectable suggesting that for a small ligand population Py11 is flipped by 180° as in the Im-Py-Py-γ-Py-Py-Py-gly-Dp and Ac-Im-Py-Py-γ-Py-Py-Py-β-Dp hairpin complexes.[17] The strong line broadening of Py2-HN would also be consistent with this interpretation since in the flipped conformation Py11-NCH_3 would be located right next to Py2-HN and could cause a large chemical shift change necessary to explain large line broadening in the intermediate to fast NMR exchange regime.

Compared to previously characterized complexes[12e] the H3, H4 and H5 resonances of Py1 are upfield shifted (Figure 2, Figure 3 and Supplementary Material: Table 3). The unusual chemical shifts and line broadening for Py1 and Py11 is particularly apparent in the NOE connectivities between *N*-methyl and H5 protons of stacked ring residues (Figure 3). Compared to all other ring pairs Py1-H5 is shifted upfield by at least 0.7 ppm and the Py11-H5 to Py11-NCH_3 cross-peak is broadened substantially. While the later observation can only be explained by conformational exchange, the unusual chemical shifts of Py1-H3, H4 and of H5 in particular, cannot be the result of exchange alone, but instead suggest that the stacking of Py1 on top of Py11 may differ significantly from that of the other residue pairs.

Previously, ring flipping was only observed for hairpin ligands with either a glycine residue in the tail or an acetyl group on the N-terminal residue.[17] When both modifications are present, in addition to a flipped terminal pyrrole ring, the ligand prefers a binding orientation opposite to that generally observed for polyamide molecules. There is no indication for binding of the ten-ring hairpin in the opposite direction and it is therefore surprising, to see a flip of a terminal pyrrole ring to occur in ten-ring hairpin that lacks a glycine as well as an N-terminal acetyl group. Clearly, one could speculate that in the ten-ring hairpin complex DNA contacts are not as tight as in shorter complexes allowing for the observed local exchange processes to occur.

Molecular modelling of the ten-ring hairpin complex with and without NOE distance restraints. Polyamide complexes with DNA are very modular in structure. This allowed us to build a molecular model of the ten-ring hairpin complex based on known geometries of previously characterized complexes. The X-ray structure of two polyamide Im-Hp-Py-Py-β-Dp molecules in a side-by-side complex with $d(CCAGTACTGG)_2$ (PDB entry 407D)[5b] was used to define a polyamide residue library for the molecular modelling package ICM. Additional residues for hairpin linker and tail regions of polyamide ligands were defined and parameterized. After defining the DNA target sequence and hairpin ligand sequence a script written for ICM tethers DNA and ligand residues to the respective residues in the X-ray structure. The script overlays the two models followed by energy minimization. The model generated for Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp in complex with d(GGAATAGTCTGC) •d(GCAGACTATTCC) is shown in Figure 4. This model was used as starting structure for restrained energy minimization using 81 intermolecular ligand-DNA and 45 intramolecular ligand-ligand restraint derived from NOE data. The restraint minimized model is overlayed on the starting structure (Figure 4) and the RMSD of all atoms is 1.22 Å. When just comparing the ligand the respective RMSD is only 1.04 Å. The main differences in the two models appear to be in the tail region that turns into the minor groove because of NOE restraints to protons of the A4•T21 and A3•T22 base pairs (Table 1). Minor adjustments also occur because Watson-Crick base pairing was enforced using distance restraints during the energy minimization protocol.

A third model was generated by first building a model for the hairpin ligand using the ICM polyamide library. This hairpin model was than energy minimized with a subset of intramolecular distance restraints, and than manually docked into a B-form DNA model of the target sequence. The model was than subjected to the energy minimization with the same NOE derived distance restraints as for the other model. The two NOE restrained models of the ten-ring hairpin complex deviate from each other with an RMSD of 1.90 Å. Again the ligands overlay better (RMSD 1.34 Å). All three models agree well with the overall features of the hairpin complex. Including NOE derived distance restraints improves the agreement between observed intermolecular DNA-ligand and intramolecular ligand-ligand contacts and the model.

Molecular modelling of the ten-ring hairpin complex and discussion of hydrogen bonding and polyamide stacking. From the molecular models hydrogen bonds between all NH amide protons and respective DNA acceptor groups can be inferred: In the NOE restrained models the following ligand amide nitrogens are within 3.0 Å of a DNA hydrogen bond acceptor.

For the unrestrained model all but Dp13 HN, Py11 HN, Im3 HN and Im7 HN are in hydrogen bonding distance (acceptor to nitrogen distance smaller or equal to 3.0 Å) of the respective DNA acceptor groups.

All three models position the imidazole nitrogen of Im3 slightly above the plane of the G7•C18 base pair (toward T8·A17). Although the NH···N alignment is far from linear which would be ideal for strong hydrogen bonding interactions, the N-N distance of approximately 3.1 Å is well with in the range found for hydrogen bonds. The situation is different for the Im7/G16 interaction. Here the geometry of seems optimal for a hydrogen bond yet the imidazole nitrogen is about 3.8 Å away from the G7 amino nitrogen. This suggests that the hydrogen bonding interaction at Im7/G16 is weaker than for Im3/G7, and this may explain why the amino group of G16 does not give rise to NOE contacts with neighboring protons. The reason for the larger distance in the case of Im7/G16 is not apparent from the model. It may be speculated that steric clashes of the γ6 hairpin linker prevent the ligand from sitting deeper in the groove.

The NMR data clearly indicates that resonance of the Py1/Py11 residue pair are affected by conformational exchange. The data is most consistent with the ring of Py11 flipped for a small population of ligands. When observed previously, ring flipping was attributed to steric clashes of acetyl groups on the N-terminal residue with the opposing tail groups or to poor contacts of the glycine tail residue with the DNA minor groove.[17] Neither argument can be made for the presently studied ten-ring ligand that lacks both groups. In addition, the molecular models do not suggest an unusual stacking arrangement of the Py1/Py11 residue pair that may explain the unusual chemical shifts observed. However, one could speculate that the surface complementarity of the ten-ring hairpin ligand with DNA must already be suboptimal to allow for occasional ring flipping of Py11.

Conclusions

For the first time, we characterized the structure of a ten-ring hairpin polyamide complexed to its DNA target site using NMR and molecular modelling. Ligand-DNA contacts are consistent with the recognition rules previously established. The ten-ring hairpin ligand binds N- to C-terminal in the 5' to 3' direction of the contacting DNA strand; the opposite orientation is not observed. The complementarity of ligand curvature to the DNA groove surface seems to persist. No major distortions of DNA or ligand geometry are observed. We do however, detect molecular motions affecting the proton resonances of the first and the last ring residue that stack on top of each other. One likely explanation is a conformational change involving a ring flip of the last *N*-methylpyrrole residue of the ten-ring hairpin for a small population of ligands. This may suggest that the contacts of the last ring pair with each other or with DNA are not as energetically favorable as at other positions of polyamide ligands.

Acknowledgement. We thank Prof. Tammy Dwyer and Cheryl Hawkins for discussions and for input files for the Insight/Discover modelling.

Supplementary Material. Tables of intramolecular ligand NOE contacts, chemical shift values of ligand protons and of selected DNA protons in the ten-ring hairpin complex.

Figure 1. (A) Structure of the polyamide hairpin Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp. Selected NOE contacts to H1' and adenine H2 protons in the minor groove of d(GGAATAGTCTGC)·d(GCAGACTATTCC) are shown. (B) Schematic representation of the ten-ring hairpin complex indicating orientation and residue stacking. Shaded circles represent *N*-methylimidazole ring residues while open circles are drawn for *N*-methylpyrrole rings.

Figure 2. Expansion of a NOESY spectra (in 95% H_2O/ 5% D_2O, 400 MHz, 25° C, τ_{mix} = 200 ms) of Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp in complex with d(GGAATAGTCTGC)·d(GCAGACTATTCC). Sequential aromatic to H1' connectivities for the DNA duplex are shown as solid lines with nucleotide numbers indicating the intraresidue aromatic to H1' cross-peaks. Dashed lines indicate resonance lines of ligand amide and *N*-methylpyrrole protons, and of DNA protons in NOE contact with ligand protons. Ligand protons are labelled according to Figure 1A.

Figure 3. Expansion of a NOESY spectra (in 100% D_2O, 400 MHz, 25° C, τ_{mix} = 200 ms) of Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp in complex with d(GGAATAGTCTGC)·d(GCAGACTATTCC). *N*-methylpyrrole or imidazole H5 to *N*-methyl proton connectivities characteristic for the residue stacking arrangement shown in Figure 1B are drawn as solid squares. Ring residue numbers indicate the intraresidue *N*-methyl proton to H5 cross-peaks. Dashed lines indicate resonance lines of ligand or DNA protons in NOE contact. Ligand protons are labelled according to Figure 1A.

Figure 4. Molecular model of Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp in complex with d(GGAATAGTCTGC)·d(GCAGACTATTCC). Stereo diagram of the complex. Overlayed are the models derived from standard geometries before (black lines) and after energy minimization with semiquantitative distance restraints derived from NOE data (gray lines) and an energy minimized NOE restrained model in which the ligand was docked manually as described in the Method section. Hydrogens have been omitted for clarity.

Figure 5. Molecular model of Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp in complex with d(GGAATAGTCTGC) ·d(GCAGACTATTCC). Stereo diagram of the ligand in the model derived from standard geometries after NOE restrained energy minimization illustrating the similarity of the stacking for the various polyamide ring residue pairs.

Molecular modelling of polyamide residue stacking and interpretation of NMR observations. Figure 5 highlights the stacking of the different polyamide ring residue pairs in the NOE restrained model. Only the stacking of Im7 on top of Py5 appears different and the H5 proton chemical shift values (Figure 3) for this ring pair are upfield of all but one H5 proton of the Py2/Py10, Im3/Py9 and Py4/Py8 pairs. However, the general trend for H5 chemical shifts is similar than for the Im-Py-Py-γ-Py-Py-Py-Dp hairpin ligand[12e]: H5 protons of the ring pair adjacent to the γ-linker are somewhat upfield of the next pair but for the N-terminal ring pair the H5 protons are most upfield shifted. The same is also true for the side-by-side dimer of Im-Py-Im-Py-Dp complexed with its target site[12d]. The chemical shift of Py1-H5 in the ten-ring hairpin complex currently investigated falls outside the range of previously observed values but the molecular model do not suggest an obvious explanation.

Table 1. Intermolecular ligand-DNA contacts in the hairpin complex identified in 2D NOESY spectra (100 ms mixing time, in 95 %/5 % H_2O/D_2O or 100 % D_2O)

Ligand	**DNA**	**Ligand**
	A3 H2	Dp13 NCH_3-1/2, Dp13-HN
	T22 H1'	β12 CH_2-H292[a,b]
	A4 H1'	Dp13 HN
	A4 H2	β12 CH_2-H301/291/292[a], Dp13 HN
	T21 H1'	β12 HN, Py11-H3
	T21 H4'	Py11-NCH_3
Py1-H3, Py1-H4, Py2-H3	A20 H2	Py11-H3, Py11-HN, Py10-H3
	A20 H1'	Py11-HN, Py10-H3
	A20 H4'	Py10-H5, Py10-H3
Py1-H3, Py2-HN, Py2-H3, Im3-HN	A6 H2	Py10-H3
Py2-HN	A6 H1'	
Py1-H5	A6 H4'	
	T19 H1'	Py10-HN, Py9-H3
	T19 H4'	Py9-H5, Py9-H3, Py9-NCH_3
Py2-H3, Im3-HN	G7 H1'	
Py2-H5, Py2-H3, Py2-NCH_3	G7 H4'	
Im3-HN, Py4-HN	G7 NH_2	Py9-HN, Py9-H3, Py10-HN
	C18 H1'	Py9-HN, Py8-H3
	C18 H4'	Py8-H5, Py8-H3, Py9-HN, Py8-NCH_3
Py4-HN	T8 H1'	
Im3-H5, Im3-NCH_3, Py4-HN	T8 H4'	
Py4-H3, Py4-HN, Py5-HN	A17 H2	Py8-H3, Py8-HN, Py9-HN
	A17 H1'	Py8-HN
	A17 H4'	Im7-H5, Im7-NCH_3
Py4-H3, Py5-HN	C9 H1'	
Py4-H5, Py4-H3, Py4-HN, Py4-NCH_3	C9 H4'	
	G16 NH_2	Im7-NH (low temp. only)
	G16 H1'	Im7-HN
Py5-H3	T10 H1'	γ6-HN
Py5-H3[c]	T10 H2'	
Py5-H3[c]	T10 H2"	
Py5-H5, Py5-H3, Py5-HN, Py5-NCH_3	T10 H4'	
Py5-H3	A15 H2	γ6-HN, γ6-H291/H302[a]

Py1-H3 to A6 H1' is not observed because chemical shift of the two protons is almost the same. [a] Protons not stereospecifically assigned. [b]Not used as restraint because only one of four expected peaks is observed. [c]Not used as restraint; can only be explained by spin diffusion.

Supplementary Material:

Table 2. Intramolecular ligand-ligand contacts in the hairpin complex identified in 2D NOESY spectra

Sequential:				Non-sequential:			
Py1	H3	Py2	HN	Py1	H4	β12	H302/1[c]
Py2	HN	Py2	H3	Py1	H3	Py11	HN
Py2	H3	Im3	HN	Py1	$(N)CH_3$	Py11	$(N)CH_3$
Py4	HN	Py4	H3	Py1	H5	Py11	$(N)CH_3$
Py4	H3	Py5	HN	Py2	H3	Py10	H3
Py5	HN	Py5	H3	Py2	$(N)CH_3$	Py10	$(N)CH_3$
Py5	H3	γ6	HN	Py2	$(N)CH_3$	Py10	H5
γ6	H301	Im7	HN	Py2	H5	Py10	$(N)CH_3$
γ6	H311	Im7	HN	Im3	$(N)CH_3$	Py9	$(N)CH_3$
γ6	HN	γ6	H291	Im3	$(N)CH_3$	Py9	H5
γ6	HN	γ6	H292	Im3	H5	Py9	$(N)CH_3$
γ6	HN	Im7	HN	Im3	HN	Py10	HN
Py8	HN	Py8	H3	Py4	HN	Py9	HN
Py8	H3	Py9	HN	Py4	HN	Py8	H3
Py9	HN	Py9	H3	Py4	H3	Py8	H3
Py9	H3	Py10	HN	Py4	$(N)CH_3$	Py8	$(N)CH_3$
Py10	HN	Py10	H3	Py4	$(N)CH_3$	Py8	H5
Py10	H3	Py11	HN	Py4	H5	Py8	$(N)CH_3$
Py11	HN	Py11	H3	Py5	HN	Py8	HN
Py11	HN	β12	HN[b]	Py5	$(N)CH_3$	Im7	$(N)CH_3$
Py11	H3	β12	HN	Py5	$(N)CH_3$	Im7	H5
β12	HN	β12	H292	Py5	H5	Im7	$(N)CH_3$
Dp13	HN	β12	H302[a]	Im7	HN	Py5	H3
Dp13	HN	β12	H301[a]				

[a]Not used as restraints because of conformational exchange and ambiguous assignment.
[b]Not used as restraint because cross-peak must be due to spin diffusion.
[c]Peak overlap makes quantitation impossible; assigned 1.5-5.0 Å as restraint.

Table 3. Chemical shift assignments of ligand proton resonances in the hairpin complex relative to H_2O at 4.76 ppm at 25° C)*

Ligand [ppm]	HN	H3	H4	H5	NC H_3	CH_2 291/2	CH_2 301/2	CH_2 311/1	Dp NCH_3
Py1	---	5.23	5.58	6.37	3.69	---	---	---	---
Py2	9.86	5.65	---	7.59	3.93	---	---	---	---
Im3	10.25	---	---	7.74	3.99	---	---	---	---
Py4	10.20	6.60	---	7.58	3.74	---	---	---	---
Py5	9.42	6.26	---	7.08	3.49	---	---	---	---
γ6	6.98	---	---	---	---	3.56/2.52	1.74/1.70	2.38/?	---
Im7	10.07	---	---	7.31	3.95	---	---	---	---
Py8	11.23	6.24	---	7.75	3.89	---	---	---	---
Py9	9.43	6.04	---	7.55	3.77	---	---	---	---
Py10	9.46	6.53	---	7.28	3.69	---	---	---	---
Py11	9.66	6.02	---	7.08	3.60	---	---	---	---
β12	7.73	---	---	---	---	3.12/3.78	2.34/2.61	---	---
Dp13	9.00	---	---	---	---	3.24/3.60	2.00/2.62	?	3.22/2.97

*γ6 protons were assigned based on similarity of NOE cross-peaks in the Im-Py-Py-γ-Py-Py-Py-Dp hairpin[12e]

Table 4. Chemical shift assignments of DNA proton resonances in the hairpin complex relative to H_2O at 4.76 ppm at 25° C)

DNA [ppm]	H8/H6	H2/H5/ CH_3	H1'	H2'	H2"	H3'	H4'	NH_2
G1	7.81	---	5.64	2.61	2.42	4.81	4.15	---
G2	7.87	---	5.44	2.79	2.70	5.02	4.35	---
A3	8.26	7.45	6.01	2.96	2.83	5.12	4.47	---
A4	8.31	8.00	6.06	2.72	2.66	5.12	4.45	---
T5	7.24	1.41	5.60	1.98	1.96	4.81	3.85	---
A6	8.09	7.92	5.17	2.76	2.56	?	3.48	---
G7	7.72	---	5.37	2.68	2.16	?	3.17	7.91 (both)
T8	7.00	1.37	5.59	2.39	1.65	4.56	2.38	---
C9	7.50	5.53	5.66	2.48	1.70	4.52	2.32	8.73/7.09
T10	7.07	1.74	5.47	1.99	1.79	4.52	2.18	---
G11	7.74	---	5.83	2.62	2.48	4.09	4.25	---
C12	7.44	5.45	6.21	2.18	2.18	4.51	4.07	---
G13	7.99	---	6.02	2.83	2.66	4.86	4.25	---
C14	7.45	5.45	5.88	2.57	2.10	4.93	4.24	8.58/6.62
A15	8.33	7.73	5.96	2.77	2.69	5.08	4.33	---
G16	7.96	---	5.19	2.75	2.68	4.99	4.17	8.73 (both)
A17	8.15	8.29	5.62	2.76	2.30	5.06	3.33	---
C18	7.06	5.06	5.48	2.29	1.45	4.54	1.74 check	8.57/6.36
T19	6.96	1.66	5.32	2.25	1.65	4.55	2.10	---
A20	8.48	7.96	5.70	2.73	2.32	4.66	2.94	---
T21	6.84	1.62	5.25	2.05	1.62	4.49	2.43	---
T22	7.19	1.51	5.96	2.37	2.04	4.83	3.94	---
C23	7.56	5.73	6.03	2.45	2.29	4.84	4.21	8.54/6.90
C24	7.68	5.82	6.25	2.29	2.29	4.54	4.05	---

(1) For subnanomolar binding, see: (a) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953-6961. (c) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636-7644. (d) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 1421-1423. (e) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.
(2) For specificity of Im/Py pairings, see: (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794. (b) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385-11389. (c) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576. (d) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756-8765.
(3) For Py/Py pairing, see: (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723-5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Biomol. Struct. Dyn.* **1990**, *8*,81-97. (c) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393-1399. (d) Wemmer, D. E.; Fagan, P. A.; Pelton, J. G. In *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions*; Pullman, B., Jortner, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1990; pp 95-101. (e) Fagan, P. A.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 1080-1081. (f) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532-12537.
(4) For Im/Im pairing, see: (a) Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5911-5919. (b) Singh, S. B.; Ajay; Wemmer, D. E.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7673-7677. (c) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569-578.
(5) For Hp/Py pairing, see: (a) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471. (b) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *288*, 111-115.
(6) For extended dimers, see: (a) Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981-6985. (b) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166. (c) Geierstanger, B. H.; Mrksich M.; Dervan, P. B.; Wemmer, D. E. *Nature, Struct. Biol.* **1996**, *3*, 321-324. (d) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600-1607. (e) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534-3535.
(7) For central-ring-bridged dimers, see: (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892-9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900-9906. (c) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663-3664. (d) Chen, Y. H.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 6995. (e) Alsaid, N. H.; Lown, J. W. *Tetrahedron Lett.* **1994**, *35*, 7577. (f) Alsaid, N. H.; Lown, J. W. *Synth. Comm.* **1995**, *25*, 1059 (g) Chen, Y. H.; Yang, Y. W.; Lown, J. W. *J. Biomol. Struct.*

Dyn. **1996**, *14*, 341. (h) Singh, M. P.; Wylie, W. A.; Lown, J. W. *Magn. Reson. Chem.* **1996**, *34*, S55. (i) Greenberg, W. A.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1998**, *4*, 796-805.
(8) For hairpin motifs, see: (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.*, **1994**, *116*, 7983-7988. (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147-6152. (c) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153-6159. (d) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198-8206. (e) Pilch, D. S.; Poklar, N. A.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306-8311.
(9) For chiral hairpins and tandem hairpins, see: (a) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391. (b) Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1999**, *5*, 975-983.
(10) For cyclic polyamides, see: (a) Cho, J. Y.; Parks, M. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10389. (b) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121-1129.
(11) For solid-phase polyamide syntheses, see: Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
(12) For NMR studies, see: (a) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586-7590. (b) Geierstanger, B. H.; Dwyer, T. J.; BatHini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474-4482. (c) Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055-3062. (d) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646-650. (e) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909-7916. (f) Wemmer, D. E.; Geierstanger, B. H.; Fagan, P. A.; Dwyer, T. J.; Jacobsen, J. P.; Pelton, J. G.; Ball, G. E.; Leheny, A. R.; Chang, W. H.; Bathini, Y.; Lown, J. W.; Rentzeperis, D.; Marky, L.; Singh, S. B.; Ajay; Kollman, P. A. In *Structural Biology: The State of the Art. Proceedings of the Eighth Conversation on Biomolecular Stereodynamics*; Sarma, S. M., Ed.; Adenine Press: Guilderland, NY, 1994; Vol. 2, pp 301-323.
(13) For X-ray studies, see: (a) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingham, M. *Nature, Struct. Biol.* **1994**, *1*, 169. (b) Chen, X.; Ramakrishnan, B.; Sundaralingham, M. *J. Mol. Biol.* **1997**, *267*, 1157. (c) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature, Struct. Biol.* **1998**, *5*, 104-109.
(14) de Clairac, R. P. L.; Seel, C. J., Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1999**, *121*, 2956-2964.
(15) For cell culture and applications of eight-ring hairpins, see: (a) Gottesfield, J. M.; Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205. (b) Dickenson, L. A.; Guzilia, P.; Trauger, J. W.; Baird, E. E.; Mosier, D. M.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12890-12895.

(c) McBryant, S. J.; Baird, E. E.; Trauger, J. W.; Dervan, P. B.; Gottesfeld, J. M. *J. Mol. Biol.* **1999**, *286*, 973-981. (d) Lenzmeier, B. A.; Baird, E. E.; Dervan, P. B.; Nyborg, J. K. *J. Mol. Biol.* **1999**, *291*, 731-744.
(16) Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97,* 3930-3935.
(17) Hawkins, C. A.; de Clairac, R. P. L.; Dominey, R. N.; Baird, E. E.; White, S.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **2000**, *122*, 5235-5243.
(18) Sklenàr, V.; Bax, A. *J. Mag. Res.* **1987**, *74*, 469-479.
(19) Hare, D. R.; Wemmer, D. E.; Chou, S.-H.; Drobny, G.; Reid, B. R.
J. Mol. Biol. **1983**, *171*, 319-336.
(20) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; J. Wiley &
Sons: New York, 1986.
(22) (a) Chou, S. H.; Hare, D. R.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1983**, *22*, 3037-3041. (b) Boelens, R.; Scheek, R. M.; Dijkstra, K.; Kaptein, R. *J. Mag. Reson.* 1985, *62*, 378-386. (c) Sklenàr, V.; Brooks, B. R.; Zon, G.; Bax, A. *FEBS Lett.* **1987**, *216*, 249-252.