Combinatorial Determination of Sequence Specificity for Nanomolar DNA-Binding Hairpin Polyamides†

Y. N. Vashisht Gopal and Michael W. Van Dyke*

Department of Molecular and Cellular Oncology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030-4009

Received December 18, 2002; Revised Manuscript Received March 28, 2003

ABSTRACT: Development of sequence-specific DNA-binding drugs is an important pharmacological goal, given the fact that numerous existing DNA-directed chemotherapeutic drugs rely on the strength and selectivity of their DNA interactions for therapeutic activity. Among the DNA-binding antibiotics, hairpin polyamides represent the only class of small molecules that can practically bind any predetermined DNA sequence. DNA recognition by these ligands depends on their side-by-side amino acid pairings in the DNA minor groove. Extensive studies have revealed that these molecules show extremely high affinity for sequence-directed, minor groove interaction. However, the specificity of such interactions in the presence of a large selection of sequences such as the human genome is not known. We used the combinatorial selection method restriction endonuclease protection, selection, and amplification (REPSA) to determine the DNA binding specificity of two hairpin polyamides, ImPyPyPy-γ-PyPyPyPy-β-Dp and ImPyPyPy-γ-ImPyPyPy-β-Dp, in the presence of more than 134 million different sequences. These were verified by restriction endonuclease protection assays and DNase I footprinting analysis. Our data showed that both hairpin polyamides preferentially selected DNA sequences having consensus recognition sites as defined by the Dervan pairing rules. These consensus sequences were rather degenerate, as expected, given that the stacked pyrrole−pyrrole amino acid pairs present in both polyamides are unable to discriminate between A·T and T·A base pairs. However, no individual sequence within these degenerate consensus sequences was preferentially selected by REPSA, indicating that these hairpin polyamides are truly consensus-specific DNA-binding ligands. We also discovered a preference for overlapping consensus binding sites among the sequences selected by the hairpin polyamide ImPyPyPy-γ-PyPyPyPy-β-Dp, and confirmed by DNase I footprinting that these complex sites provide higher binding affinity. These data suggest that multiple hairpin polyamides can cooperatively bind to their highest-affinity sites.

Hairpin polyamides are a class of DNA-binding molecules chemically related to the natural product DNA minor groove binder distamycin (reviewed in ref *1*). They consist of two polyamide oligomers tethered in an antiparallel orientation that contain *N*-methylpyrrole (Py),[1] *N*-methylimidazole (Im), and 3-hydroxypyrrole (Hp) amino acids. Additional amino acids such as β-alanine (β) and (*N*,*N*-dimethylamino)-propylamide (Dp) form the linker groups, while γ-aminobutyric acid (γ) forms the curved tether for the hairpin structure. These molecules bind to DNA in a hairpin conformation similar to that of the side-by-side 2:1 distamycin−DNA complex (*2*). Some of the well-characterized molecules of this class are the eight-ring hairpin polyamides (four rings on each side of the hairpin), which bind to 6 bp target sequences at sub-nanomolar concentrations (*3*, *4*). Two examples of hairpin polyamides and their binding to DNA are shown schematically in Figure 1. These molecules can distinguish G·C base pairs from C·G and A·T or T·A base pairs depending on the side-by-side pairings of the amino acids in the DNA minor groove. For example, an antiparallel pairing of a Py opposite an Im targets a C·G bp, whereas an Im/Py pairing recognizes a G·C bp (*5*). However, a Py/Py pairing is degenerate and recognizes either A·T or T·A base pairs (*2*). Introduction of the Hp amino acid into the hairpin structure removes this degeneracy and slightly decreases binding affinity (*6*). Other five-membered heterocyclic rings are known to enhance the biological activity of the hairpin polyamides. For example, substitution of the amino acid desmethylpyrrole for Py yields hairpin polyamides with increased water solubility, though decreased binding specificity (*7*). Pairing 3-pyrazolecarboxylic acid with Py enhances the binding affinity and specificity for G·C base pairs compared to that of Im/Py pairs (*8*). Likewise, an *N*-

† This research was supported by grants from the American Cancer Society (RPG-97-028-03-LBC) and the Robert A. Welch Foundation (G-1199). Y.N.V.G. was supported by a postdoctoral fellowship from the Department of Defense (DAMD17-01-0305).

* To whom correspondence should be addressed: Department of Molecular and Cellular Oncology, Unit 079, University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030-4009. Phone: (713) 792-8954. Fax: (713) 794-0209. E-mail: mvandyke@mdanderson.org.

[1] Abbreviations: β, β-alanine; C_{50}, apparent binding constant; Dp, (*N*,*N*-dimethylamino)propylamide; EDTA, ethylenediaminetetraacetic acid; γ, γ-aminobutyric; Hp, 3-hydroxypyrrole; HPA, hairpin polyamide; IISRE, type IIS restriction endonuclease; Im, *N*-methylimidazole; N, random nucleotide; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; Py, *N*-methylpyrrole; REPA, restriction endonuclease protection assay; REPSA, restriction endonuclease protection, selection, and amplification; ST-2, selection template 2; W, adenine or thymine.

10.1021/bi027373s CCC: $25.00
Published on Web 05/16/2003

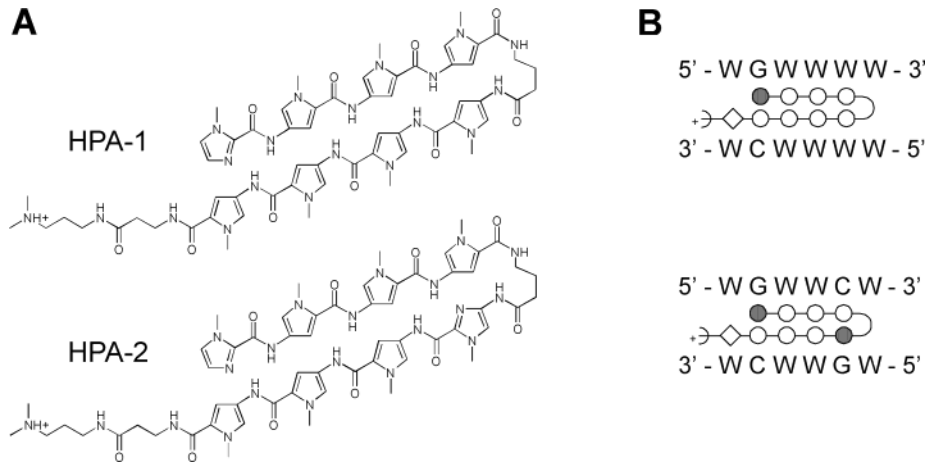

FIGURE 1: (A) Chemical structures of the hairpin polyamides ImPyPyPy-γ-PyPyPyPy-β-Dp (HPA-1, top) and ImPyPyPy-γ-ImPyPyPy-β-Dp (HPA-2, bottom). (B) Schematic representations of HPA-1 (top) and HPA-2 (bottom) binding to their consensus DNA sequences. Hairpin polyamide amino acids include *N*-methylpyrrole (○), *N*-methylimidazole (●), γ-aminobutyric acid (⊂), β-alanine (◇), and (*N*,*N*-dimethylamino)propylamide (∈). W represents nucleic acid bases adenine and thymine.

methylpyrazole/Py pair demonstrates greater specificity than a Py/Py pair for A·T and T·A base pairs than for C·G and G·C base pairs (*9*).

In principle, hairpin polyamides can be designed to bind sequence specifically to any duplex DNA, and they can differentiate between preferred and nonpreferred sites. These characteristics give them immense pharmacological potential. For example, regulation of specific gene expression has been demonstrated in which the downregulation of RNA polymerase III- and RNA polymerase II-dependent transcription was achieved using hairpin polyamides (*10*−*16*). These studies showed that this class of molecules would avidly bind to the transcription factor response elements in genes and either dissociate the bound transcription factors or occlude their DNA-binding sites, thus interfering with transcription. Transcription activation has also been investigated using hairpin polyamides to occlude the promoter binding of transcriptional repressors and through the design of bifunctional hairpin polyamides tethered with small activation peptides, both with promising results (*17*−*19*). In addition, hairpin polyamide−camptothecin conjugates have been used sequence specifically to trap topoisomerase I at nanomolar concentrations and induce DNA cleavage (*20*).

Hairpin polyamides possess DNA binding affinity comparable to that of transcription factors but may show less sequence specificity because these molecules interact differently with DNA than do transcription factors. Transcription factors have large, modular DNA recognition domains that undergo conformational changes when they bind DNA, sometimes also inducing DNA conformational changes for the best fit, and that establish three-dimensional spatial and chemical interactions with DNA (*21*). Hairpin polyamides, on the other hand, are small molecules whose DNA binding affinity depends on how well they fit into the curvature of the minor groove and form hydrogen bonds with the nucleotide bases. Sequence specificity may thus be governed by hydrogen bonding between a snugly fit hairpin polyamide (in the minor groove) and the DNA bases, with steric disturbances between the hydrogen-bonding groups playing an important role (*22*, *23*). Some studies of small molecule−DNA interactions have indicated that sequence specificity may also depend on other factors, like flanking sequences in the region of binding and cooperative binding effects (*24*, *25*). Although no conclusive findings have been made for the hairpin polyamides, such factors are nonetheless worth exploring. In addition, hairpin polyamides with Py/Py pairings recognize a degenerate code of either A·T or T·A base pairs, but whether a recognition preference exists for either bp in relation to adjacent amino acid pairings is not completely known. Therefore, to realize the full therapeutic potential of hairpin polyamides, a broader understanding of DNA specificity vis-à-vis base preference, flanking sequences, and cooperativity is required. Well-designed studies would reveal the best sequences that can be targeted and show the extent of unwanted interactions with other DNA sequences.

The purpose of this study was to determine the DNA sequence specificity of two eight-ring hairpin polyamides, ImPyPyPy-γ-PyPyPyPy-β-Dp (HPA-1) and ImPyPyPy-γ-ImPyPyPy-β-Dp (HPA-2) (Figure 1A). According to the hairpin polyamide pairing rules, the Py and Im pairing combination in HPA-1 dictates a consensus binding sequence of 5′-WGWWWW-3′/5′-WWWWCW-3′ (where W is A or T), whereas that in HPA-2 dictates a consensus of 5′-WGWWCW-3′ (Figure 1B). Earlier studies showed that HPA-1 and HPA-2 have their highest affinity for the sequences 5′-AGTATT-3′ and 5′-AGTACT-3′, respectively (*3*). These findings were determined through standard DNase I footprinting assays on plasmid or genomic DNA-derived restriction fragments of small length and limited sequence selection. In the study presented here, specificity for a new set of high-affinity recognition sequences was determined through the combinatorial procedure restriction endonuclease protection, selection, and amplification (REPSA). REPSA is a highly sensitive, PCR-based, enzymatic procedure that simultaneously searches very large numbers of different DNA sequences for ligand binding and extracts those sequences with the highest binding affinity. This combinatorial method has been successfully tested on macromolecular DNA-binding ligands, including the purine motif triple-helical DNA-forming oligodeoxyribonucleotide ODN 1 and the TATA box-binding subunit of the human class II general transcription factor TFIID (*26*, *27*). However, unlike conventional combinatorial methods, which rely on the physical

separation of ligand-bound from free nucleic acids, REPSA has also been successfully used to identify the preferred duplex DNA binding sites of small molecules such as distamycin A and actinomycin D in their native state (*28*, *29*). The results of our REPSA assay showed that the hairpin polyamide pairing rules held true for both molecules under study, with ~60% of the selected DNA sequences containing the consensus sequences expected for its hairpin polyamide. However, within a particular degenerate consensus, no individual sequences exhibited significantly higher affinity than other consensus members. One hairpin polyamide, HPA-1, preferentially selected DNA sequences containing multiple, overlapping consensus binding sites. These findings were verified by DNase I footprinting assays to reveal the best binding sequences for the two hairpin polyamides.

MATERIALS AND METHODS

Oligonucleotides. Polyacrylamide gel electrophoresis (PAGE)-purified oligonucleotides obtained from either Integrated DNA Technologies (Coralville, IA) or Gibco BRL (Rockville, MD) were used in this study. The nucleotide sequences of the oligonucleotides (5′ → 3′) were as follows: 63AL, CTAGGAATTCGTGCAGAGGTGAAT; 63AR, GTCCAAGCTTCTGGAGGGATGGTAA; FPP-L, CAGTGAATTCGTGCAGAGGTGAAT; FPP-R, CACACAGGAAACAGCTATGACCA; MSU, ACGACGTTGTAAACGA; NFP1, CAGTGAATTCGTGCAGAGGTGAATTGTATTGTTAACATTACGAATTCCTGTTTTTATGCTTGTAGTATTTGGTCATAGCTGTTTCCTGTGTG; and NFP2, CAGTGAATTCGTGCAGAGGTGAATAGTATTGTATTCATTACGAATTCCTGTATATATGCTTGTTGTATTTGGTCATAGCTGTTTCCTGTGTG.

The single-stranded starting material 63R14 (GTCCAAGCTTCTGGAGGGATGGTAAN_{14}ATTCACCTCTGCACGAATTCCTAG) used for making the ST-2 REPSA template was synthesized on a Millipore (Bedford, MA) Cyclone Plus DNA synthesizer using standard phosphoramidite chemistry. The 14-base random nucleotide region (N) was made using an equimolar mixture of each phosphoramidite for every incorporated base.

Biochemicals and Chemicals. Restriction enzymes were purchased from New England Biolabs (Beverly, MA) and DNase I from Roche Diagnostics (Mannheim, Germany). Other chemicals and biochemicals used in the study were standard analytical grade. The two hairpin polyamides, GSQ-412X-3 (HPA-1) and GSQ-436X-3 (HPA-2), were synthesized by GeneSoft Inc. (South San Francisco, CA). The hairpin polyamides were dissolved in water before use, and their concentration was determined spectroscopically.

REPSA. The ST-2 selection template (Figure 2A) was synthesized by six rounds of a polymerase chain reaction (PCR) using the oligonucleotide 63R14 as the template and 63AL and 63AR as the primers for the reaction. REPSA, schematically represented in Figure 2B, is a three-step process. In the first step (ligand binding), the ST-2 template (5 nM) was incubated with the hairpin polyamide (20 nM) for 1 h at 37 °C in a 10 μL volume containing 10 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, and 0.05% Nonidet P-40. These conditions were previously determined to be optimal for differentiating between weak and strong hairpin polyamide−DNA binding through an initial restriction endonuclease protection assay (REPA) using defined templates that contained either a consensus hairpin polyamide binding site or one containing a single-site mutation (*3*). The second step of REPSA is DNA cleavage with a type IIS restriction enzyme (IISRE), which, unlike the regular type II enzymes, binds to a specific sequence of DNA but cleaves DNA some fixed distance from the binding site without any sequence specificity. The ST-2 DNA contains IISRE-binding sites flanking the random 14 bp region. The IISREs bind to these sites and cleave the DNA near the center of the 14 bp region. The samples from the first step were subjected to a 10 min restriction digestion with 0.3 unit of the IISRE *Fok*I. The reaction was stopped by transferring the samples onto dry ice. In the third step of REPSA, the products of the restriction digestion were directly amplified in a 50 μL volume PCR reaction using 100 ng of ^{32}P end-labeled 63AL and 100 ng of unlabeled 63AR primers, dNTPs (0.2 mM each), and 2.5 units of Taq DNA polymerase (Sigma, St. Louis, MO) in a PCR buffer containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 1 mM $MgCl_2$. The amplification profile was 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 2 min, for a total of six cycles. The amplified DNA was purified by phenol extraction

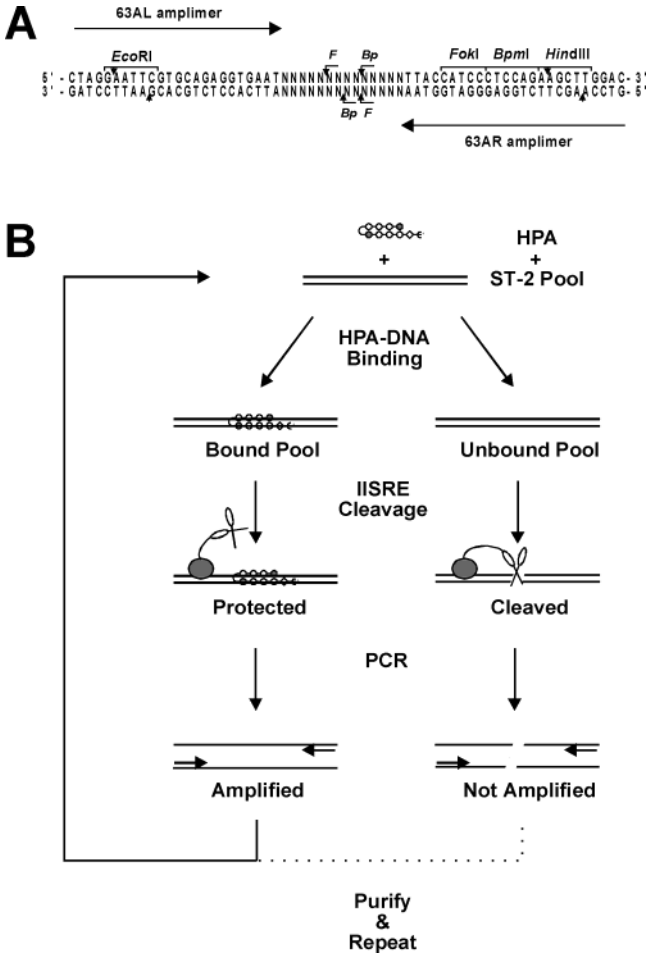

FIGURE 2: (A) Schematic representation of the REPSA selection template ST-2. Restriction endonuclease binding sites are indicated by brackets, and their cleavage sites are indicated by arrows: *F* for *Fok*I and *Bp* for *Bpm*I. (B) Flowchart for the combinatorial method REPSA. Hairpin polyamides are represented by a folded set of beads, and a type IIS restriction endonuclease is represented by scissors (DNA-cleaving domain) tethered to a gray oval (DNA-binding domain).

followed by spin filtration through Ultrafree-MC 10 000 NMWL spin filters (Millipore Corp.) at 15000*g* for 30 min. The filters were washed for 10 min in TE buffer [10 mM Tris-HCl (pH 7.5) and 1 mM ethylenediaminetetraacetic acid (EDTA)]. The DNA retained on the filter was resuspended in 20 μL of TE. Two nanograms of this DNA served as the starting template for the next round of REPSA. *Fok*I was the primary IISRE of choice because its cleavage efficiency is much higher than that of other characterized IISREs (*27*). After each round of REPSA, the purified DNA was analyzed by PAGE and autoradiography to detect the emergence of a cleavage-resistant population. In the study presented here, five rounds of REPSA with *Fok*I were sufficient to obtain a significant cleavage-resistant population (~25%) among the selected DNA. A sixth round of REPSA was carried out using the IISRE *Bpm*I to reduce the selection of DNAs containing *Fok*I-binding sites in the randomized region.

Sequence Determination. After the final round of REPSA with *Bpm*I, the emergent, cleavage-resistant population of ST-2 DNA was digested with *Eco*RI and *Hin*dIII and ligated into a similarly cut pUC19 plasmid. The plasmids were transformed into *Escherichia coli* strain XL1-blue cells, and the cells were cultured on LB/agar plates containing 100 μg/mL ampicillin. The clones were screened for inserts by blue−white color selection, and plasmid minipreps were made from the insert-containing clones. The DNA inserts were manually sequenced by the Sanger dideoxynucleotide sequencing method using the MSU primer.

REPA. For each HPA, 45 REPSA-selected clones were screened by a cleavage protection assay (REPA) to select for those that contained high-affinity hairpin polyamide binding sites. Radiolabeled fragments containing the DNA inserts were generated by PCR amplification using ^{32}P end-labeled FPP-L and unlabeled FPP-R primers. The amplified DNA was purified as described in the third step of REPSA. As in the first and second steps of REPSA, the radiolabeled DNA (10 000 cpm, 5 nM) was incubated with the hairpin polyamide for 1 h, and the DNA was digested with 0.3 unit of *Fok*I for 10 min. The reaction was stopped with 2 μL of a 5% sodium dodecyl sulfate (SDS) solution. The samples were resolved on a 10% nondenaturing PAGE gel. The DNA bands were visualized by autoradiography, and DNA cleavage protection was quantitated by densitometry.

DNase I Footprinting. Quantitative DNase I footprinting was carried out using radiolabeled DNA fragments generated by PCR amplification of plasmid DNA clones with ^{32}P-labeled FPP-L and unlabeled FPP-R primers. Ten thousand counts per minute (2.5 nM) of the labeled DNA probe was incubated with increasing concentrations of the hairpin polyamides for 1 h at 37 °C in a buffer containing 10 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, and 90 ng of poly(dG-dC) (carrier DNA) in a total reaction volume of 20 μL. DNase I [1 μL of a 5 units/L solution containing 20 mM Tris-HCl (pH 7.8), 50 mM NaCl, 30% glycerol, 2 mM $CaCl_2$, 2 mM dithiothreitol, and 0.1 mg/mL bovine serum albumin (BSA)] was added, and the samples were incubated at room temperature for 30 s. The reaction was stopped with 5 μL of a stop solution containing 3 M ammonium acetate and 0.5 M EDTA. The samples were purified by phenol extraction and ethanol precipitation. They were then dried and resuspended in a loading dye containing 80% formamide, bromophenol blue, and xylene cyanol in 1× Tris-borate EDTA buffer [89 mM Tris-borate and 2 mM EDTA (pH 8.0)]. The DNA in the samples was denatured by heating at 100 °C for 5 min and was immediately chilled on ice before being loaded on a denaturing 8% polyacrylamide gel. Products from adenine- or guanine-specific chemical cleavage reactions were used as markers. The samples were then electrophoresed on a prerun 8% urea−PAGE gel at 40 W, allowing the bromophenol blue to migrate to the end of the gel. The gel was dried and autoradiographed. The footprints were quantified densitometrically, and following normalization of sample loading, the apparent binding affinities (C_{50} is the HPA concentration required to provide 50% DNase I protection) were determined from a nonlinear regression analysis.

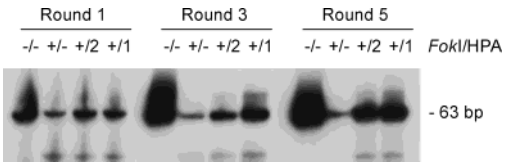

FIGURE 3: Emergence of an HPA-dependent, *Fok*I cleavage-resistant oligonucleotide population. Pools of oligonucleotides (2 ng) obtained after one or more rounds of REPSA selection with HPA-1 or HPA-2 were either left untreated (−/−) or *Fok*I-digested in the absence (+/−) or presence of HPA-1 (+/1) or HPA-2 (+/2), as indicated. An autoradiogram of their PAGE-resolved products is shown. Control reactions shown (−/− and +/−) were carried out with HPA-2 REPSA-selected oligonucleotides, though HPA-1 REPSA-selected oligonucleotides gave qualitatively equivalent results (data not shown).

RESULTS

REPSA Selects HPA-Dependent, FokI Cleavage-Resistant DNAs. The ST-2 DNA template used for REPSA was a 63 bp DNA with a 14 bp random unit flanked on either side by defined segments that contained nested IISRE sites. Type IIS restriction endonucleases, unlike other type II enzymes, bind to a particular sequence of DNA but cleave the DNA some fixed distance from the binding site, irrespective of the sequence being cleaved (Figure 2A). In ST-2 DNA, the IISREs bind in the flanks of the 14 bp random region and cleave the DNA within the 14 bp region. This randomized region potentially offers more than 134 million ($4^{14}/2$) possible sequences. The amount of ST-2 DNA present in a REPSA reaction mixture is typically 2 ng (48 fmol), which is equivalent to ~2.9×10^{10} DNA molecules. Since the expected binding site for the hairpin polyamides under study was 6 bp, a REPSA reaction mixture containing 2 ng of ST-2 would most likely contain all possible 6 bp sequence combinations with which the hairpin polyamides could interact. In addition, the 14 bp random region allowed the selection of larger binding sites resulting from cooperative binding between hairpin polyamide molecules. Such phenomena were previously observed with the related DNA-binding polyamide distamycin A (*28*).

Five rounds of REPSA selection were carried out with the IISRE *Fok*I and either HPA-1 or HPA-2. Aliquots (2 ng) were removed after rounds 1, 3, and 5 and were either left untreated or cleaved with *Fok*I under standard selection conditions in the absence or presence of the appropriate HPA. As shown in Figure 3, after the first round of selection, very little cleavage protection was observed when 20 nM HPA was present. However, after five rounds of REPSA, both

the HPA-2- and the HPA-1-selected DNAs demonstrated 10-fold greater cleavage resistance in the presence of their respective HPAs. Such resistance indicates that a substantial fraction (~25%) of the selection templates present in these populations contained high-affinity HPA-binding sites. This enrichment for ligand-binding sequences was greater than previously observed at the conclusion of REPSA selections with other small molecules, especially considering the smaller number of REPSA rounds required in the selection presented here (*28*, *29*). This most likely reflects the more rigorous IISRE cleavage conditions that were used, which allows a greater degree of selection per each round. A sixth round of REPSA was carried out using a different IISRE (*Bpm*I) to reduce the likelihood of selecting *Fok*I binding sites in the random region.

REPSA-Selected DNAs Contain Consensus HPA−DNA Binding Sites. After subcloning and purification of the pUC19 plasmid clones containing the hairpin polyamide-selected ST-2 templates had been carried out, the insert DNAs were sequenced by the Sanger dideoxynucleotide sequencing method. The sequences of 45 randomly selected clones from each hairpin polyamide selection are shown in Tables 1 and 2. Analysis of the nucleotide ratios in these sequences revealed no significant differences between those of the ST-2 starting material (27% A, 18% C, 16% G, and 39% T) and those of either the HPA-1- (27% A, 15% C, 17% G, and 41% T) or HPA-2-selected (25% A, 17% C, 16% G, and 41% T) DNAs (*27*). Such similarities indicated that under the binding conditions used in our REPSA selection, the hairpin polyamides did not merely bind to generally A/T-rich sequences, which are usually favored by individual *N*-methylpyrrole polyamides such as netropsin and distamycin (*30*). Rather, these findings suggested a greater degree of binding specificity, indicative of correct side-by-side amino acid stacking within the hairpin polyamides. In fact, using the expected recognition sequences for the two hairpin polyamides (5′-WGWWWW-3′/5′-WWWWCW-3′ for HPA-1 and 5′-WGWWCW-3′ for HPA-2) as defined by the Dervan pairing rules (*6*), we found that 22 of 45 HPA-1-selected DNAs and 19 of 45 HPA-2-selected DNAs contained one or more of these expected binding sites within the 14 bp originally randomized region. These numbers increased to 32 of 45 HPA-1-selected DNAs and 24 of 45 HPA-2-selected DNAs when the potential binding sites that partially overlapped with sequences in the defined flanks were included. Interestingly, when these flanking sequences were counted, 26 of 45 HPA-1-selected DNAs contained the sequences 5′-WCWWWW-3′/5′-WWWWGW-3′, while 5 of 45 HPA-2-selected DNAs contained the sequence 5′-WCWWGW-3′. Such sequences are expected when hairpin polyamides bind DNA in a reverse orientation (*31*). Finally, only two of the 90 DNA inserts that were sequenced (clones HPA-2-27 and HPA-2-34) contained *Fok*I recognition sequences (5′-GGATG-3′/5′-CATCC-3′). This frequency was comparable to that expected for any random sequence and indicated either that *Fok*I binding sites were not selected for under our REPSA conditions or that the final REPSA selection with *Bpm*I significantly diminished the prevalence of these sites.

REPSA-Selected DNAs Contain High-Affinity HPA−DNA Binding Sites. To identify the REPSA-selected clones containing high-affinity hairpin polyamide binding sites, radiolabeled fragments carrying the inserts were generated from the above clones by PCR amplification. These DNA fragments were then screened for *Fok*I cleavage protection in the presence of 20 or 40 nM hairpin polyamide. The cleavage-protected DNA was quantified by densitometric scanning of DNA bands after their gel electrophoretic separation and autoradiography. Figure 4 shows a representative *Fok*I cleavage protection assay involving the hairpin polyamide HPA-1 on four of its selected clones. In this example, each of the probes demonstrated more than 70% cleavage protection when 40 nM HPA-1 was present. The extent of cleavage protection conferred by each of the two ligands on their respective REPSA-selected clones is shown in Tables 1 and 2.

With both hairpin polyamides, approximately 25% of the clones demonstrated substantial *Fok*I cleavage protection (more than 70% cleavage inhibition), while an additional 25% demonstrated significant cleavage protection (51−70% inhibition). Most of these well-protected clones contained one or more of the expected recognition sequences for the two hairpin polyamides (5′-WGWWWW-3′/5′-WWWWCW-3′ for HPA-1 and 5′-WGWWCW-3′ for HPA-2). The only exceptions were clones that contained a single reverse-orientation hairpin polyamide binding sequence (HPA-1-34 and HPA-2-46). The poorly protected clones (those with less than 30% cleavage protection), particularly the HPA-2 REPSA-selected clones, typically did not contain the expected recognition sequences or their reverse. The relative abundance of these "nonspecific" DNAs may reflect limitations in a combinatorial selection method like REPSA. However, they may also represent an appreciable level of nonspecific DNA binding by polyamides, since similar proportions of nonspecific DNAs were not previously observed with REPSA selections containing other ligands (*26*, *27*, *29*).

Statistical Analysis of REPSA-Selected, High-Affinity HPA−DNA Binding Sites. We statistically analyzed REPSA-selected clones to identify possible consensus sequences for HPA−DNA recognition. Only sequences present within the original 14 bp randomized cassette were considered, to avoid skewing by the defined sequences in the flanking regions. The 22 clones identified as containing high-affinity HPA binding sites (+++ and ++) obtained in both HPA-1 and HPA-2 selections contained a total of 198 possible 6 bp binding sites. For HPA-1, 19 clones contained one or more 5′-WGWWWW-3′/5′-WWWWCW-3′ sequences, for a total of 24. Using Fisher's exact test, we found a two-sided P value of 0.0186 for the prevalence of these degenerate sequences among all possible sites, which is considered statistically significant. More impressive, the observation of one or more consensus sequences in 19 of these 22 high-affinity clones, when only one was expected by random chance, gave a P of <0.0001, which is considered extremely significant. For HPA-2, 16 clones contained one 5′-WGW-WCW-3′ sequence. This prevalence gave a two-sided P value of 0.0002 for the appearance of this consensus among all the possible sites and a P of <0.0001 for its appearance in most of the high-affinity clones. Both P values are considered extremely significant. No other sequences were represented as abundantly as these two degenerate sequences for HPA-1 and HPA-2. Thus, our analyses suggested that the HPA binding sites defined by the Dervan pairing rules were the

Table 1: REPSA-Selected Sequences Containing HPA-1 Binding Sites

Clone	Sequence (5'→3')[a]	REPA Protection[b]
HPA-1-01	gaa**t GTATTGTTAACA**TC ttac	+++
HPA-1-02	gaat CC**TGTTTT**TATAGC ttac	+++
HPA-1-03	gaa*t ATGTAAAGTGTTTA* ttac	+++
HPA-1-17	gaat ATAGGTACCA**TGTT tt**ac	+++
HPA-1-21	gaa*t TTGTTTT*TTGCA*CA* *tt*ac	+++
HPA-1-22	gaat TAACC**TGTTAAAAC t**tac	+++
HPA-1-25	gaat TAA*TTGTTTTAGTA* **tt**ac	+++
HPA-1-26	gaat GGCT**TTAACA***TATA* ttac	+++
HPA-1-30	gaat A*TAT**TGT**TAT**TCAT* *tt*ac	+++
HPA-1-31	gaa*t TGT*TAC**AGTATAGT** ttac	+++
HPA-1-43	gaa*t TGT***AGTATA**GGGAA ttac	+++
HPA-1-18	gaat TGCAATGG**TTTACT** *tt*ac	++
HPA-1-20	gaat CCTATCC**AGTAAA**T ttac	++
HPA-1-32	gaat A**TTATCAAAATTGA tta**c	++
HPA-1-34	gaat GT*TCATAAGCTTGG* ttac	++
HPA-1-35	g**aat TCA**CGAT**GTAAAGA tta**c	++
HPA-1-40	gaa*t TGTCTCGTTTTATC t*tac	++
HPA-1-45	gaat AA*TTTGATTTTTTT* ttac	++
HPA-1-50	gaa*t TGTTTATTCTAATT* ttac	++
HPA-1-51	gaat C*TTAACTAGCCTTG* ttac	++
HPA-1-53	gaat G**AGTTTATCTTAGT tta**c	++
HPA-1-60	gaat **ATTACATTACTAGC** ttac	++
HPA-1-24	gaat GATGTGGAAAATTT ttac	+
HPA-1-33	gaat CCTAGCTATTTTTT ttac	+
HPA-1-38	gaat **TTTTCT**TGACTATC ttac	+
HPA-1-39	gaat TTTATTTTTAA**AGA tta**c	+
HPA-1-41	gaat CCCA*TCTTTAACAA* ttac	+
HPA-1-42	gaa**t TAACTAC AGTTTTT** ttac	+
HPA-1-44	gaat ATTTTTTGG**AGATT t**tac	+
HPA-1-46	gaat CTTTGT*TCTTTAAC t*tac	+
HPA-1-47	gaat TGCCGGTAAA**AGTT tt**ac	+
HPA-1-52	gaat ATTTAAATAGGA*CA ttac*	+
HPA-1-54	gaat TCGTGCAGA**TGAAT t**tac	+
HPA-1-04	gaat GGCTGAA*CATTTCG* ttac	-
HPA-1-16	gaat GCTAATTGC**TAATC t**tac	-
HPA-1-23	gaat ACGTCGACTAAGTC ttac	-
HPA-1-27	gaat AGGCTTGTTGCA*CA ttac*	-
HPA-1-28	ga**at AACA**CGATTCA*CTA ttac*	-
HPA-1-29	gaat GCCTTTTACCTTAA ttac	-
HPA-1-36	gaat TAGTGATAGCTAAG ttac	-
HPA-1-37	gaa**t TTAC***T**TTG*AACTGC ttac	-
HPA-1-49	gaa*t TGAGTTGTTGTGCT* ttac	-
HPA-1-55	gaat TACTGTCCGTTTTT ttac	-
HPA-1-56	gaat TAAATTGGGCAAGC ttac	-
HPA-1-58	gaat GAATTTCCGTAGAC ttac	-

[a] Sequences present in the 14 bp randomized region are in uppercase letters, while sequences present in the flanking regions are in lowercase letters. Bold type indicates HPA-1 binding sites predicted by side-by-side amino acid pairing rules, while italic type indicates reverse orientation binding sites. Overlap between two forward orientation HPA binding sites is indicated by a single underline, while overlap between two reverse orientation HPA binding sites is indicated by a dotted underline. A wavy underline indicates an overlap between forward and reverse orientation HPA binding sites. A double underline indicates an overlap between three or more forward and reverse orientation HPA binding sites. [b] Values correspond to the percentage of *Fok*I cleavage inhibition observed when 40 nM HPA-1 was present: ≤30% cleavage inhibition (−), 31−50% cleavage inhibition (+), 51−70% cleavage inhibition (++), and >70% cleavage inhibition (+++).

Table 2: REPSA-Selected Sequences Containing HPA-2 Binding Sites

Clone	Sequence (5'→3')[a]	REPA Protection[b]
HPA-2-01	gaat AGG*TCA***AGTTCA**AT ttac	+++
HPA-2-02	gaat ATTT**TGTTCT**TCAA ttac	+++
HPA-2-04	gaat GGTTG**AGTACT**CAG ttac	+++
HPA-2-17	gaat TAAACTC**AGTTCA**A ttac	+++
HPA-2-18	gaat GATCTTA**AGTACA**T ttac	+++
HPA-2-23	gaat ACCG**AGTACT**CTGA ttac	+++
HPA-2-25	gaat CGTCAAT**TGTACT**A ttac	+++
HPA-2-39	gaat TTCATT**TGTACT**TT ttac	+++
HPA-2-40	gaat CACTCATT**TGTTCT** ttac	+++
HPA-2-43	gaat C**TGTTCT**CCTTAGC ttac	+++
HPA-2-52	gaat AGTT*TCT***AGTTCA**T ttac	+++
HPA-2-06	gaat TATCCTA**TGTTCA**T ttac	++
HPA-2-12	gaat GAAATT**TGTTCA**TT ttac	++
HPA-2-27	gaat TGGTCTGGA**TGTAC t**tac	++
HPA-2-31	gaat GAACATTGT**TGTAC t**tac	++
HPA-2-38	gaat TGTGC**TGAACT**CAC ttac	++
HPA-2-46	gaat TTTTGTAA*ACTTGA* ttac	++
HPA-2-47	gaat TTTATCTA**TGTACT** ttac	++
HPA-2-48	gaat TATTAATTT**AGTTC t**tac	++
HPA-2-49	gaa**t GTACT**TACACCAAA ttac	++
HPA-2-56	gaat AGT**TGTTCA**TTTTA ttac	++
HPA-2-58	gaat TGAGGTAGA**AGTTC t**tac	++
HPA-2-08	gaat GATCGATTTTCTTA ttac	+
HPA-2-13	gaat G*TCTAGT*ATATCTA ttac	+
HPA-2-19	gaat TACAG**TGAACA**ACT ttac	+
HPA-2-37	gaat ATTC*TCAAGT*AACT ttac	+
HPA-2-53	gaat **TGAACT**TAATCTGT ttac	+
HPA-2-57	gaat **TTGAACA**AGGTGTG ttac	+
HPA-2-05	gaat TTTTCATTCTCTTT ttac	-
HPA-2-07	gaat TATCCTGTTATACA ttac	-
HPA-2-09	gaat TAGCATGACAGAAT ttac	-
HPA-2-10	gaat TACTATGTATCGAT ttac	-
HPA-2-11	gaat TTAGTTTTCAGTAC ttac	-
HPA-2-15	gaat TTTAGCACTGCTTT ttac	-
HPA-2-16	gaat TTGTGATCCTGAGA ttac	-
HPA-2-21	gaat TCAAGGTCTGAATA ttac	-
HPA-2-26	gaat TCTTTGGACTATCT ttac	-
HPA-2-28	gaat GTTCATGGCTACAA ttac	-
HPA-2-29	gaat AAGTAGTTTTCATC ttac	-
HPA-2-30	gaat ATACTGGTACAATC ttac	-
HPA-2-32	gaat TTAAGTGCACTTCA ttac	-
HPA-2-33	gaat TACGCTTATTTCTT ttac	-
HPA-2-34	gaat TGCTGCCTGGATGA ttac	-
HPA-2-36	gaat ATGGTTATGCGGTT ttac	-
HPA-2-41	gaat TTTTATTTCATAGT ttac	-

[a] Sequences present in the 14 bp randomized region are in uppercase letters, while sequences present in the flanking regions are in lowercase letters. Bold type indicates HPA-2 binding sites predicted by side-by-side amino acid pairing rules, while italic type indicates reverse orientation binding sites. Overlap between forward and reverse orientation HPA binding sites is indicated by a wavy underline. [b] Values correspond to the percentage of *Fok*I cleavage inhibition observed when 40 nM HPA-2 was present: ≤30% cleavage inhibition (−), 31−50% cleavage inhibition (+), 51−70% cleavage inhibition (++), and >70% cleavage inhibition (+++).

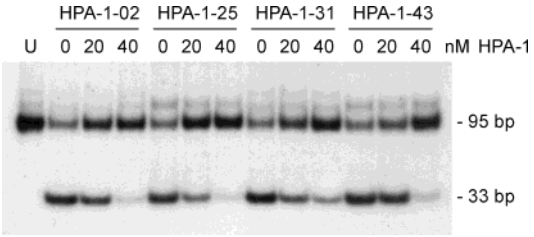

FIGURE 4: Characterization of REPSA-selected clones by a restriction endonuclease protection assay (REPA). Labeled DNA probes (95 bp) generated by PCR from different HPA-1 REPSA-selected clones were incubated with the indicated concentrations of HPA-1 before digestion with *Fok*I. An autoradiogram of their PAGE-resolved products is shown. U represents the uncleaved probe control.

consensus sequences for HPA-1 and HPA-2 for DNAs exhibiting high-affinity binding sites after REPSA selection.

The expected recognition sequences for the hairpin polyamides HPA-1 and HPA-2 are quite degenerate, with 64 and 16 individual sequences possible for each, respectively, including their complements. These are listed in the Supporting Information (Tables A and B). Thus, a question of whether a subset of these degenerate consensus sequences are actually the preferred binding sites for these molecules exists. The sequences 5′-TGTACT-3′, 5′-TGTTCA-3′, and 5′-TGTTCT-3′ were among the most prevalent of the high-affinity HPA-2-selected clones, though this could have reflected in part the abundance of thymine in our ST-2 starting material. Supporting evidence for this contention could be found in the relative absence of their respective complementary sequences (5′-AGTACA-3′, 5′-TGAACA-3′, and 5′-AGAACA-3′) which, although A-rich, provided equivalent binding sites for the duplex DNA-binding ligand HPA-2. Thus, to minimize this weighting, we investigated the summation of prevalences for sequences and their complements in our statistical analyses. The most prevalent potential HPA-2 binding sites that were found were 5′-AGTACA-3′/5′-TGTACT-3′ and 5′-AGTTCA-3′/5′-TGAACT-3′, with four each. Fisher's exact test for the observed prevalence of these sequences compared to their expected prevalence out of 198 possible 6 bp sites did not show statistical significance ($P = 0.1231$). Likewise, the prevalence of each of these pairs of sequences appearing among the 16 WGWWCW sequences present in the high-affinity clones was also not significant ($P = 0.6539$), compared with expectations. Similar results were found for HPA-1-selected sequences (Table A of the Supporting Information). Thus, while it is possible that subsets of preferred sequences exist for HPA-1 and HPA-2, it was not possible to identify such sequences statistically from the limited number of clones that were investigated.

Identification of High-Affinity, HPA−DNA Binding Sites by DNase I Footprinting. To identify the exact binding sites for the hairpin polyamides, DNase I cleavage protection assays (footprinting) were carried out on five clones from each hairpin polyamide selection that showed the highest levels of cleavage protection by REPA. Radiolabeled DNA fragments of the clones were generated by PCR using the [^{32}P]FPP-L and FPP-R primers. These were PAGE purified, incubated with increasing concentrations of the hairpin polyamides (10, 20, 40, 60, and 80 nM) for 1 h at 37 °C, and then subjected to DNase I cleavage for 30 s at room temperature. The cleavage products were subjected to high-resolution denaturing gel electrophoresis and visualized by autoradiography. Cleavage products from guanine-specific chemical sequencing reactions were run in parallel as sequence markers. Representative footprints are shown in Figure 5. As seen in these examples, DNase I cleavage protection was restricted to the expected hairpin polyamide consensus sites present within the 14 bp region, with an offset toward the 3′-side as expected for right-handed helical B-form DNA (*32*). The extent of the binding affinity was quantified from a densitometric analysis of the footprints, after normalizing sample loading using a DNase I cleavage located in the flanking region, and the apparent binding constants (C_{50}) were determined for 10 consensus sites having the best binding affinity for HPA-1 and HPA-2 (Table 3). Under our reaction conditions, we found that all of the high-affinity HPA-binding sites exhibited similar binding affinities, within the C_{50} range of 28−43 nM for HPA-1 and the C_{50} range of 26−44 nM for HPA-2. In each case, these binding constants were considerably greater than the concentration of labeled DNA probe present (2.5 nM), which is necessary if these values are to be considered valid. No additional footprints were observed at the highest hairpin polyamide concentration that was investigated (80 nM), confirming the sequence specificity of these ligands under these binding conditions.

Characterization of Overlapping HPA−DNA Binding Sites by DNase I Footprinting. Most of the high-affinity clones isolated in the REPSA selection with HPA-1 contained multiple consensus HPA-1 binding sequences, in both the expected forward (indicated in bold) and reverse (indicated in italic) orientations (see Table 1). Many of these binding sites overlapped one another, with reverse−forward overlaps (indicated by a wavy underline) more prevalent than forward−forward (bold underline) and reverse−reverse (dotted underline) overlaps. The preponderance of these overlap sites after REPSA selection suggested that they could provide greater binding affinity for HPA-1 than isolated sites. Such a preponderance has been observed before in our REPSA selections with the small molecule DNA binders distamycin A and actinomycin D (*28*, *29*). To better understand the affinity of the ligands for overlapping sites, we used PCR to create two radiolabeled DNA probes (ND-1 and ND-2) containing multiple overlapping sites from the synthetic oligonucleotides NFP1 and NFP2, respectively, using the FPP1 and FPP2 primers. The ND-1 probe contained the complex overlapped sites present in clone HPA-1-01, including its flanking regions (Figure 6A). The ND-2 probe contained a different series of HPA-1 consensus binding sites but with orientations similar to those in ND-1. Both ND-1 and ND-2 contained isolated consensus binding sites, which allowed a direct comparison of binding affinity for overlapped and isolated consensus sites on the same DNA probe. DNase I footprinting analysis of HPA-1 binding to the ND-1 probe revealed a complicated pattern of footprints encompassing the series of overlapping binding sites from F2 to R3 (Figure 6B). An initial footprint appeared at 10 nM HPA-1 in the region of F5(R3), which expanded over the adjacent (R2)F4 region by 20 nM HPA-1 and over the entire region by 60 nM HPA-1. In comparison, the isolated HPA-1 binding sites F6 and F7 exhibited substantial cleavage protection only at 60 nM HPA-1, which was comparable to

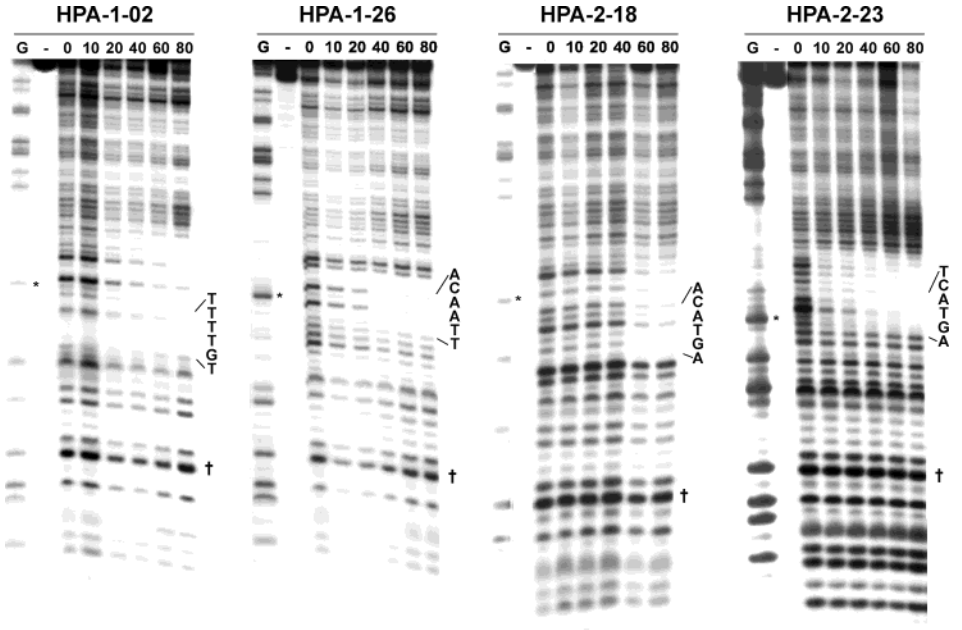

FIGURE 5: Identification of HPA binding sites and determination of their apparent binding affinities by quantitative DNase I cleavage protection and high-resolution denaturing gel electrophoreses. Shown are autoradiograms of the resolved DNA fragments obtained after incubation of singly 5′ end-labeled PCR probes from the indicated clones with increasing concentrations of HPA and limiting DNase I cleavage. G-specific chemical sequencing reaction products (G) served as electrophoretic DNA length markers. Lanes corresponding to the purified probe are indicated with a hyphen (-). The locations and sequences of the expected HPA binding sites are indicated at the right. Asterisks denote PCR artifacts present in the unpurified probe. The dagger denotes the DNase I cleavage product that served as a loading control.

Table 3: Apparent Binding Constants for High-Affinity HPA Binding Sites

clone[a]	binding site[b]	C_{50} (nM)[c]
HPA-1-02	TGTTTT/AAAACA	28
HPA-1-26	TTAACA/TGTTAA	35
HPA-1-31	AGTATA/TATACT	43
ND-1 (HPA-1)[d]	AGTATT/AATACT	30
ND-2 (HPA-1)	TGTATT/AATACA	31
HPA-2-01	AGTTCA/TGAACT	37
HPA-2-02	TGTTCT/AGAACA	40
HPA-2-04	AGTACT/AGTACT	26
HPA-2-18	AGTACA/TGTACT	44
HPA-2-25	TGTACT/AGTACA	27

[a] The binding affinity was determined with the HPA used in the REPSA selection process. [b] Binding site sequences are shown from 5′ to 3′ in the following order: top strand of the 14 bp region/its complement. [c] Apparent binding constant (C_{50}) values were determined from a one-site binding analysis of DNase I cleavage protection through a titration of HPA concentrations. [d] Investigated on the synthetic DNA probe ND-1 with the hairpin polyamide HPA-1.

those values observed for most individual HPA-1 consensus binding sites (Table 3). Similar results were also found with probe ND-2, with footprinting initially appearing at (R2)F4 at 20 nM HPA-1 and expanding over the entire F2−R3 region at 60 nM HPA-1. Taken together, these data indicate that overlapping consensus sequences provided higher-affinity HPA-1 binding sites than did isolated, individual consensus sequences.

DISCUSSION

Manipulating the expression of specific endogenous genes is a wide-ranging research goal in medicine and in experimental and applied biology. To accomplish this goal, a number of pharmacological approaches have been investigated targeting the transcriptional and posttranscriptional levels of gene expression. Approaches targeting either of these levels can be applied successfully in specific circumstances, but because transcription is the first genetic switch that directs the functioning of cellular machinery, transcriptional regulation is being more intensely investigated. Artificial transcription regulation using synthetic DNA-binding ligands has interested chemists and biologists alike for the past two decades and has led to the development of numerous sequence-specific DNA-binding ligands. Examples include the triplex-forming oligonucleotides (*33*), peptide nucleic acids (*34*), oligosaccharides (*35*), and small molecule ligands (*36*−*38*). The oligonucleotides and peptide nucleic acids show highly specific DNA binding and transcription interference; however, they suffer from poor cellular uptake, a hindrance to therapeutic applications. The oligosaccharides, on the other hand, are cell permeable, but they do not recognize a wide range of DNA sequences. Small molecule ligands, such as the lexitropsins, are a more rational pharmacological alternative for artificial transcription regulation. These DNA minor groove-binding molecules based on distamycin have been gaining importance because they show good cellular uptake and DNA sequence-directed binding. The best known molecules of this class are the hairpin polyamides, developed by the group headed by P. Dervan at the California Institute of Technology (Pasadena, CA) (reviewed in ref *1*).

Numerous studies on the DNA binding characteristics of hairpin polyamides have shown that they possess nanomolar to picomolar levels of binding affinity for preferred sequences and weaker affinities for sites with a single base mutation. Specificity for consensus sequence recognition (determined from the amino acid pairing rules for DNA base pair recognition) was also quite high. The hairpin polyamides, however, exhibited certain nonspecific interactions with

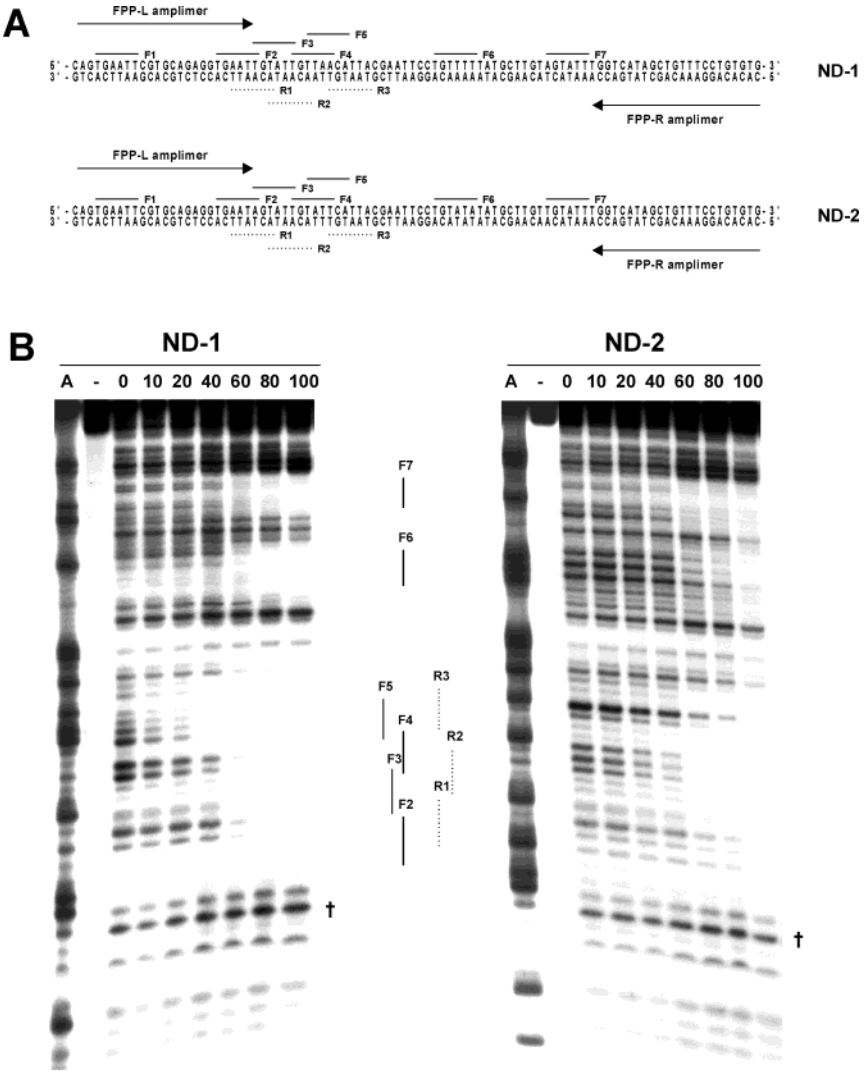

FIGURE 6: Characterization of HPA-1 binding to overlapping sites. (A) Sequences of the synthetic probes ND-1 (top) and ND-2 (bottom). Consensus HPA-1 binding sequences having a forward (F) or reverse (R) orientation are indicated by solid and dotted lines, respectively. (B) Autoradiograms of the resolved DNA fragments obtained after incubation of singly 5′ end-labeled ND-1 (left) or ND-2 (right) with increasing concentrations of HPA-1 and limiting DNase I cleavage. A-specific chemical sequencing reaction products (A) served as electrophoretic DNA length markers. Lanes corresponding to the purified probe are indicated with a hyphen (-). The locations of the consensus HPA-1 binding sequences are indicated at the right. The dagger denotes the DNase I cleavage product that served as a loading control.

much lower affinity. Previous nuclear magnetic resonance (NMR) studies had revealed that this binding was largely due to an inverted or reverse binding orientation of the hairpin polyamides (*31*, *39*). Typically, the binding affinity and specificity analysis for the hairpin polyamides has been determined by quantitative DNA footprinting, which is an extremely powerful technique for determining binding affinity (*40*). But a broad-ranging specificity may not be appropriately determined using such a method because of the limited sequence variability available on a given length of DNA. To understand true binding specificity, the ligand should have the choice of all possible nucleotide base combinations from which to select. Hypothetically, considering the four-base code and the complementarity of duplex DNA, a ligand with a 6 bp binding site would have 2048 ($4^6/2$) possible sequences for binding. This should be considered a minimal value since it does not take into account the effects of flanking sequences on ligand−DNA binding, nor the possibility of cooperative binding between adjacent sites. Screening such a large selection of sequences by conventional methods such as footprinting would require an enormous effort from investigators. A combinatorial method like REPSA offers a similar selection of sequences in a single reaction tube from which a ligand can select the highest-affinity binding sequences. Also, unlike other combinatorial methods that require physical separation of ligand-bound DNA from free DNA, which limits the resolution and identification of all the sequences bound by a ligand, REPSA is a PCR-based enzymatic process that selectively amplifies ligand-bound sequences and eliminates unbound sequences. The sequence information of the emergent population can then be determined and the best binding sequences identified by a cleavage protection assay such as REPA.

In this study, ~50% of the REPSA-selected DNA population exhibited high-affinity binding characteristics. Most all of these contained at least one consensus binding site for the hairpin polyamides, as defined by the Dervan pairing rules. In the case of HPA-1, many of the selected DNAs contained multiple consensus binding sites, in both the expected forward and less favorable reverse orientations.

Also, many of the HPA-1-selected sequences contained overlapping binding sites, especially between reverse and forward orientation sites. However, neither multiple nor overlapping binding sites were prevalent with the high-affinity HPA-2-selected sequences, which instead primarily contained only single, forward orientation sites. These data indicate that different hairpin polyamides, while generally following the established pairing rules, nevertheless preferentially bind DNA differently, with different propensities for complex and simple sites. This would also suggest that binding rules established for one hairpin polyamide may not be entirely applicable to other hairpin polyamides, and that each may need to be independently investigated to obtain a full understanding of its own binding specificity.

Since both HPA-1 and HPA-2 contain multiple Py/Py amino acid pairings, which cannot discriminate between A·T and T·A DNA base pairs, their expected consensus recognition sequence would be rather degenerate, encompassing 32 different sequences for HPA-1 and 10 different sequences for HPA-2. Surprisingly, no individual sequence or subset of these degenerate consensus sequences was preferentially selected by REPSA. Previously, both X-ray crystallographic and NMR studies on the distamycin−DNA interaction indicated a binding preference for stretches of contiguous A (or T) bases by the 1:1 complex, whereas a 2:1 complex, which is structurally most similar to that expected for hairpin polyamide−DNA complexes, preferred sequences with more A-T alterations, which allows for a wider minor groove (*41*−*44*). This preference for alternating A/T bases was not observed in the HPA-1 REPSA-selected sequences, with more than half of the identified sites containing one or no A/T alterations. Likewise, more than half of the HPA-2 REPSA-selected sequences had no A/T alterations in the central WW base pairs. Thus, if minor groove width is an important determinant of hairpin polyamide binding, these sequences containing nonalternating bases must be achieving increased minor groove width through other means. Additionally, this lack of absolute sequence specificity for Py/Py-containing hairpin polyamides may limit their utility for recognizing individual sequences. However, their apparently uniform specificity for a degenerate consensus is fairly unique among small molecule DNA-binding ligands and would make hairpin polyamides superior agents for degenerate A/T-rich targets (e.g., TATA box elements in class II gene promoters).

One hypothesis we sought to test with our REPSA selections was whether sequences flanking the expected 6 bp binding site had a significant effect on hairpin polyamide binding. A statistical analysis of the sequences selected by HPA-1 and HPA-2 found no preferred flanking sequences for either molecule. Thus, formally we found no evidence for specific flanking sequence effects on these hairpin polyamide consensus binding sites. However, the propensity of high-affinity HPA-1-selected sequences to contain multiple, sometimes overlapping consensus binding sites, in both forward and reverse orientations, does strongly suggest that features beyond those of the minimal 6 bp binding sites do influence HPA-1 binding to these best sites.

For the sequences selected by REPSA with HPA-1, and to a far lesser extent with HPA-2, several contained sites showing a reverse binding orientation for the consensus, in which the polyamide strands (N → C) would bind the minor groove of DNA in the less favorable 3′ → 5′ direction. Most of these occurred in the context of an overlap with a forward orientation HPA binding site, though a few (e.g., HPA-1-34 and HPA-2-46) contained autonomous reverse orientation binding sites. Affinity cleavage and NMR studies have found that this reverse binding orientation only occurs for a very small proportion of HPA molecules at any one time (*31*). In addition, this phenomenon was minimized with hairpin polyamides, like those used in this study, that lack a N-terminal acetyl group and have a *β*-Dp instead of glycine-Dp C-terminal modification (*39*). Ignoring for a moment those reverse sites that overlap forward orientation sites, we find that the relative paucity of independent reverse binding sites observed with REPSA selections for both HPA-1 and HPA-2 is consistent with prior studies and would indicate that reverse HPA binding is either unfavorable or of sufficiently low affinity as to not significantly interfere with IISRE cleavage. The cause of the unexpected abundance of reverse−forward orientation overlapped sites, however, is not explained by these previous studies, but may reflect more characteristics of overlapping sites rather than those of reverse orientation binding.

Many of the sequences selected by REPSA with HPA-1 contained overlapping binding sites. Most of these contained reverse and forward orientation binding sites with either three or four base pair overlaps. Examples include HPA-1-26 (**TTA***ACA**TAT*, where the forward orientation site is indicated in bold, the reverse orientation site is indicated in italics, and the overlap is underlined), HPA-1-30 (*TAT**TGT***-**TAT**), and HPA-1-31 (**AG***TATA**GT*). Some overlaps were the consequence of A/T-rich flanking sequences immediately adjacent to the randomized region in the ST-2 template and should not be counted in a proper statistical analysis of their prevalence. However, among the 22 highest-affinity HPA-1 binding sites, 11 had reverse−forward overlapping sites located wholly within the randomized region. One reason for the apparent abundance of these overlaps may be the general A/T-rich nature of the ST-2 starting material, which makes sequences such as **WWW***WCW**WWW* and *WWW*-***WGW*WWW** more likely. This contention is supported by statistical analyses that find the appearance of any particular overlap sequence in the high-affinity HPA-1 REPSA-selected clones as not being statistically significant. However, the overall abundance of a variety of different overlapping sites does suggest that such complex HPA-1 binding sites do provide some advantage during REPSA selection.

DNase I footprinting was used to identify the exact hairpin polyamide binding sites and to determine apparent binding constants. Footprinting is a superior means of determining true binding affinities on native templates, since the relatively nonspecific nuclease DNase I is not appreciably affected by the location of the ligand-binding site within the 14 bp region, whereas the location-specific type IIS restriction endonuclease is. Analysis of high-affinity clones containing individual binding sites yielded footprints centered on the expected consensus sequences and binding affinities in the range of 25−45 nM. This tight clustering of binding affinities was not surprising, since Py/Py pairs are not able to distinguish between A·T and T·A base pairs and show minimal affinity differences for a variety of A/T-rich sequences (*5*, *46*). In addition, this tight clustering of binding affinities helps explain why no subset of the HPA-1 or

HPA-2 degenerate recognition sequences dominated after REPSA selection. Notably, the apparent binding affinities determined in our experiments differed substantially from those previously reported in the literature (e.g., ~25 pM for HPA-1 binding to 5′-ttAGTACTtg-3′ and ~0.3 nM for HPA-2 binding to 5′-ttAGTATTtg-3′) (*3*). These absolute affinity differences may reflect differences in the binding conditions used in the two experiments, including use of different ionic strength buffers, carrier DNA, and different incubation times. Of these, the most significant differences were in the times of incubation (1 h vs 15 h), suggesting that equilibrium was less likely to be achieved under our binding conditions. Thus, the sequences identified in our REPSA selection are those kinetically favored by hairpin polyamides. Interestingly though, the finding of identical preferred binding sites under such different conditions suggests that these consensus sequences are likely to be preferred even under nonideal circumstances, and would more closely reflect their recognition in a more realistic setting (e.g., postreplicative chromosomal DNA prior to nucleosomal assembly *in vivo*).

DNase I footprinting was also used to better understand the interaction of HPA-1 with complex binding sites, which were numerous in its REPSA selection. Synthetic probe ND-1 was derived from the sequences present in clone HPA-1-01, whereas ND-2 was similar in complexity and organization, but utilized different consensus binding sites. With both probes, HPA-1 initially protected at one forward orientation consensus binding site at low HPA-1 concentrations and expanded its protected region to cover adjacent forward binding sites at increasing HPA-1 concentrations. The site of HPA-1 nucleation differed in ND-1 and ND-2, suggesting that sequence and not the arrangement of binding sites is most important. More significant, binding at these initial protection sites occurred at HPA-1 concentrations far lower than that typically necessary to observe protection at isolated consensus binding sites. These data are consistent with a model in which several HPA-1 molecules are simultaneously bound to these complex sites, and that their binding occurs with some degree of cooperativity. How exactly this cooperativity is achieved is difficult to determine from our footprinting analysis. NMR studies have found that two hairpin polyamides can stably occupy two adjacent forward orientation consensus binding sites, but only if they are separated by one intervening base pair (*47*). Likewise, stable occupancy of two reverse orientation sites required a separation of two intervening base pairs. In both cases, any closer association between hairpin polyamides would be unfavorable given the expected steric interference and like-charge repulsion between the C-terminal β-Dp tails of these molecules. Ultimately, it will be necessary to use affinity cleavage methods (*46*) to obtain meaningful occupancy and orientation information for these complex sites as a function of ImPyPyPy-γ-PyPyPyPy-β-Dp-EDTA concentration.

ACKNOWLEDGMENT

We thank GeneSoft Inc. for providing the hairpin polyamides used in this study and Dakshesh Mehta for help in data analysis.

SUPPORTING INFORMATION AVAILABLE

Expected recognition sequences for the hairpin polyamides HPA-1 and HPA-2, with 64 and 16 individual sequences possible for each, respectively, including their complements. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Dervan, P. B. (2001) *Bioorg. Med. Chem. 9*, 2215−2235.
2. Pelton, J. G., and Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5723−5727.
3. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) *Nature 382*, 559−561.
4. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc. 119*, 6953−6961.
5. Wade, W. S., Mrksich, M., and Dervan, P. B. (1992) *J. Am. Chem. Soc. 114*, 8783−8794.
6. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature 391*, 468−471.
7. Bremer, R. E., Szewczyk, J. W., Baird, E. E., and Dervan, P. B. (2000) *Bioorg. Med. Chem. 8*, 1947−1955.
8. Zhan, Z. Y., and Dervan, P. B. (2000) *Bioorg. Med. Chem. 8*, 2467−2474.
9. Nguyen, D. H., Szewczyk, J. W., Baird, E. E., and Dervan, P. B. (2001) *Bioorg. Med. Chem. 9*, 7−17.
10. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E., and Dervan, P. B. (1997) *Nature 387*, 202−205.
11. McBryant, S. J., Baird, E. E., Trauger, J. W., Dervan, P. B., and Gottesfeld, J. M. (1999) *J. Mol. Biol. 286*, 973−981.
12. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 12890−12895.
13. Dickinson, L. A., Trauger, J. W., Baird, E. E., Dervan, P. B., Graves, B. J., and Gottesfeld, J. M. (1999) *J. Biol. Chem. 274*, 12765−12773.
14. Lenzmeier, B. A., Baird, E. E., Dervan, P. B., and Nyborg, J. K. (1999) *J. Mol. Biol. 291*, 731−744.
15. Chiang, S., Burli, R. W., Benz, C. C., Gawron, L., Scott, G. K., Dervan, P. B., and Beerman, T. A. (2000) *J. Biol. Chem. 275*, 24246−24254.
16. Ehley, J. A., Melander, C., Herman, D., Baird, E. E., Ferguson, H. A., Goodrich, J. A., Dervan, P. B., and Gottesfeld, J. M. (2002) *Mol. Cell. Biol. 22*, 1723−1733.
17. Dickinson, L. A., Trauger, J. W., Baird, E. E., Ghazal, P., Dervan, P. B., and Gottesfeld, J. M. (1999) *Biochemistry 38*, 10801−10807.
18. Mapp, A. K., Ansari, A. Z., Ptashne, M., and Dervan, P. B. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 3930−3935.
19. Ansari, A. Z., Mapp, A. K., Nguyen, D. H., Dervan, P. B., and Ptashne, M. (2001) *Chem. Biol. 8*, 583−592.
20. Wang, C. C., and Dervan, P. B. (2001) *J. Am. Chem. Soc. 123*, 8657−8661.
21. Patikoglou, G., and Burley, S. K. (1997) *Annu. Rev. Biophys. Biomol. Struct. 27*, 289−325.
22. Pilch, D. S., Poklar, N., Gelfand, C. A., Law, S. M., and Breslauer, K. J. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 8306−8311.
23. Pilch, D. S., Poklar, N., Baird, E. E., Dervan, P. B., and Breslauer, K. J. (1999) *Biochemistry 38*, 2143−2151.
24. Winkle, S. A., and Krugh, T. R. (1981) *Nucleic Acids Res. 9*, 3175−3186.
25. Bailey, S. A., Graves, D. E., Rill, R., and Marsch, G. (1993) *Biochemistry 32*, 5881−5887.
26. Hardenbol, P., and Van Dyke, M. W. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2811−2816.
27. Hardenbol, P., Wang, J. C., and Van Dyke, M. W. (1997) *Nucleic Acids Res. 25*, 3339−3344.
28. Hardenbol, P., Wang, J. C., and Van Dyke, M. W. (1997) *Bioconjugate Chem. 8*, 617−620.
29. Shen, J., Wang, J. C., and Van Dyke, M. W. (2001) *Bioorg. Med. Chem. 9*, 2285−2293.
30. Abu-Daya, A., Brown, P. M., and Fox, K. R. (1995) *Nucleic Acids Res. 23*, 3385−3392.
31. White, S., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc. 119*, 8756−8765.

32. Van Dyke, M. W., and Dervan, P. B. (1983) *Nucleic Acids Res. 11*, 5555−5567.
33. Giovannangeli, C., and Hélène, C. (2000) *Curr. Opin. Mol. Ther. 3*, 288−296.
34. Pooga, M., Land, T., Bartfai, T., and Langel, U. (2001) *Biomol. Eng. 17*, 183−192.
35. Liu, C., Smith, B. M., Ajito, K., Komatsu, H., Gomez-Paloma, L., Li, T., Theodorakis, E. A., Nicolaou, K. C., and Vogt, P. K. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 940−944.
36. Wemmer, D. E. (2000) *Annu. Rev. Biophys. Biomol. Struct. 29*, 439−461.
37. Gottesfeld, J. M., Turner, J. M., and Dervan, P. B. (2000) *Gene Expression 9*, 77−91.
38. Goodsell, D. S. (2001) *Curr. Med. Chem. 8*, 509−516.
39. Hawkins, C. A., de Clairac, R. P., Dominey, R. N., Baird, E. E., White, S., Dervan, P. B., and Wemmer, D. E. (2000) *J. Am. Chem. Soc. 122*, 5235−5243.
40. Trauger, J. W., and Dervan, P. B. (2001) *Methods Enzymol. 340*, 450−466.
41. Coll, M., Frederick, C. A., Wang, A. H.-J., and Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385−8389.
42. Yoon, C., Privé, G. G., Goodsell, D. S., and Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332−6336.
43. Pelton, J. G., and Wemmer, D. E. (1988) *Biochemistry 27*, 8088−8096.
44. Fagan, P. A., and Wemmer, D. E. (1992) *J. Am. Chem. Soc. 114*, 1080−1081.
45. Herman, D. M., Baird, E. E., and Dervan, P. B. (1999) *Chem. Eur. J. 5*, 975−983.
46. White, S., Baird, E. E., and Dervan, P. B. (1996) *Biochemistry 35*, 12532−12537.
47. Hawkins, C. A., Baird, E. E., Dervan, P. B., and Wemmer, D. E. (2002) *J. Am. Chem. Soc. 124*, 12689−12696.

BI027373S