

## Head-to-head *bis*-Hairpin Polyamide Minor Groove Binders and Their Conjugates with Triplex-forming Oligonucleotides: Studies of Interaction with Target Double-stranded DNA

<http://www.jbsdonline.com>

### Abstract

Two hairpin hexa(*N*-methylpyrrole)carboxamide DNA minor groove binders (MGB) were linked together *via* their N-termini in head-to-head orientation. Complex formation between these *bis*-MGB conjugates and target DNA has been studied using DNase I footprinting, circular dichroism, thermal dissociation, and molecular modeling. DNase I footprint revealed binding of these conjugates to all the sites of 492 b.p. DNA fragment containing (A/T)<sub>n</sub>X<sub>m</sub>(A/T)<sub>p</sub> sequences, where n>3, p>3; m=1,2; X = A,T,G, or C. Binding affinity depended on the sequence context of the target. CD experiments and molecular modeling showed that oligo(*N*-methylpyrrole)carboxamide moieties in the complex form two short antiparallel hairpins rather than a long parallel head-to-head hairpin. Binding of *bis*-MGB also stabilized a target duplex thermodynamically. Sequence specificity of *bis*-MGB/DNA binding was validated using *bis*-conjugates of sequence-specific hairpin (*N*-methylpyrrole)/(*N*-methylimidazole) carboxamides.

In order to increase the size of recognition sequence, the conjugates of *bis*-MGB with triplex-forming oligonucleotides (TFO) were synthesized and compared to TFO conjugated with single MGB hairpin unit. *Bis*-MGB-oligonucleotide conjugates also bind to two blocks of three and more A·T/T·A pairs similarly to *bis*-MGB alone, independently of the oligonucleotide moiety, but with lower affinity. However, the role of TFO in DNA recognition was demonstrated for *mono*-MGB-TFO conjugate where the binding was detected mainly in the area of the target sequence consisting of both MGB and TFO recognition sites.

Basing on the molecular modeling, three-dimensional models of both target DNA/*bis*-MGB and target DNA/TFO-*bis*-MGB complexes were built, where *bis*-MGB forms two antiparallel hairpins. According to the second model, one MGB hairpin is in the minor groove of 5'-adjacent A/T sequence next to the triplex-forming region, whereas the other one occupies the minor groove of the TFO binding polypurine tract. All these data together give a key information for the construction of MGB-MGB and MGB-oligonucleotide conjugates possessing high specificity and affinity for the target double-stranded DNA.

Key words: Oligonucleotide; Minor groove binder; Conjugate; Circular dichroism; DNase I footprint; Molecular modeling; Binding affinity; Sequence specificity.

### Introduction

Molecules specifically interacting with double-stranded DNA are intensively studied as biologically active compounds potentially capable of acting on genomic DNA in living cells. Besides natural specific peptides and zinc fingers, two classes of synthetic molecules are known that recognize and bind specifically to double-stranded DNA sequences: triple helix-forming oligonucleotides (TFO) (1-3) and hairpin oligo(*N*-methylpyrrole)/(*N*-methylimidazole) carboxamide minor groove binders (MGB) (4-9).

Ludovic Halby<sup>1</sup>  
Vladimir A. Ryabinin<sup>2</sup>  
Alexandre N. Sinyakov<sup>2</sup>  
Darya S. Novopashina<sup>2</sup>  
Alya G. Venyaminova<sup>2</sup>  
Serguei L. Grokhovsky<sup>3</sup>  
Anna N. Surovaya<sup>3</sup>  
Georgy V. Gursky<sup>3</sup>  
Alexandre S. Boutorine<sup>1,\*</sup>

<sup>1</sup>Muséum National d'Histoire Naturelle  
RDDM, USM 0503  
57 rue Cuvier, B.P. 26  
Paris Cedex 05, F-75231 France;  
INSERM, U565  
Paris, CNRS, UMR 5153  
Paris, F-75231 France.

<sup>2</sup>Institute of Chemical Biology and  
Fundamental Medicine  
Siberian Division of Russian  
Academy of Sciences  
9, Lavrentyev Prosp.  
630090, Novosibirsk, Russia.

<sup>3</sup>Engelhardt Institute of Molecular Biology  
Russian Academy of Sciences  
ul. Vavilova 32, Moscow, 117984 Russia

\*Phone: +33 1 40 79 36 96  
Fax: +33 1 40 79 37 05  
Email: alexandre.boutorine@mnhn.fr

In our previous work, we constructed head-to-head *bis*-hairpin minor groove binders that demonstrated remarkable dsDNA affinity (apparent  $K_d = 4\text{--}5$  nM) and DNA recognition sequence specificity of 8-10 base pairs (10). However, several questions remained still unclear. First, the sequence context preferences of *bis*-conjugates within relatively long DNA possessing several different binding sites had to be studied. Second, despite some indirect indications, a binding mode of two polyamides in the DNA minor groove was not elucidated. Since the DNA minor groove cannot contain simultaneously more than two polyamide strands (11), they can bind to DNA either in linear parallel orientation (as a long parallel hairpin) or in the form of two short hairpins that occupy the minor groove in antiparallel directions. These questions are very important for design of future reagents directed to specific genomic DNA sequences, because the chemical structure of the reagent will be different depending on the presumed binding mode.

In order to increase the specificity of the ligands, covalent linkage of TFO to MGB has been proposed (12-15). In our previous publication we described conjugates of triplex-forming oligonucleotides with one (*mono*-conjugates) and two (*bis*-conjugates) hairpin oligocarboxamide minor groove binders. They are able to bind simultaneously to the major and the minor grooves of the target DNA in a sequence-specific manner. Synergistic contributions of both components in the DNA recognition have been demonstrated (15). At low pH ( $\leq 6$ ) and temperatures ( $< 24$  °C) both TFO and MGB components formed a sequence-specific complex with the target DNA. However, a drastic increase in the DNA binding affinity was demonstrated by MGB part of the *bis*-conjugate: a triple complex was still stable even after dissociation of oligonucleotide part from the DNA major groove (at high pH and temperatures). *Bis*-MGB moiety even appeared to increase the melting temperature of the target duplex (15). However, the detailed structure of the triple complex, the conformation of *bis*-MGB in the binding site, as well as the binding site preferences of conjugates on relatively long DNA fragments with several different potential binding sites were not studied.

In the present work, we use DNase I footprint, circular dichroism, temperature denaturation, and molecular modeling methods to study these questions. We approve our proposal about two-hairpin antiparallel binding mode of *bis*-MGB ligands in DNA minor groove (10). To validate our conclusions, we synthesized sequence-specific *bis*-MGB ligands containing *N*-methylimidazole/*N*-methylpyrrole (Im/Py) pairs. They are able to recognize their cognate DNA targets containing all four base pairs (according to Dervan rules) only in a two-hairpin configuration. Their high affinity to the designed target sequences directly demonstrated that this configuration is preferable. Molecular models of DNA/*bis*-MGB and DNA/*bis*-MGB-TFO conjugate complexes were built on the base of these data.

## Materials and Methods

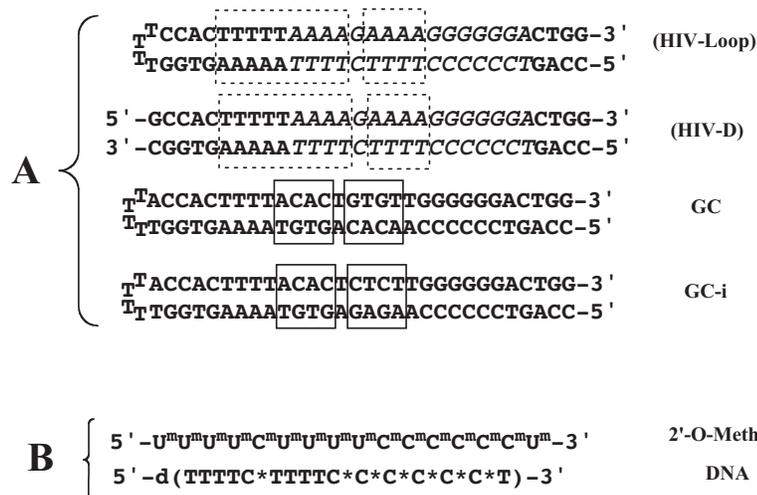
### Reagents

Reagents, sorbents, and solvents were purchased from Sigma-Aldrich-Fluka, VWR, Acros Organics and Reachim (Russia). They were used without additional purification. Oligodeoxyribonucleotides and their 5-methylcytosine analogues were purchased from Eurogentec (Belgium). Synthesis of oligo(2'-O-methylribonucleotides) was described in a previous publication (15). Sequences of TFO and target dsDNA are presented on Figure 1. Restriction nucleases and other gene engineering enzymes were from Promega, DNase I – from Sigma – Aldrich, radioactive nucleoside triphosphates – from Amersham or Izotop, RAN, Russia. Poly(dA)·poly(dT) and poly[d(AT)]·poly[d(AT)] duplexes were purchased from Sigma-Aldrich.

### General Experimental Procedures

<sup>1</sup>H-NMR analysis of synthesized compounds was carried out in C<sub>2</sub>D<sub>6</sub>SO on the

## Head-to-head bis-Hairpin Polyamide Minor Groove Binders



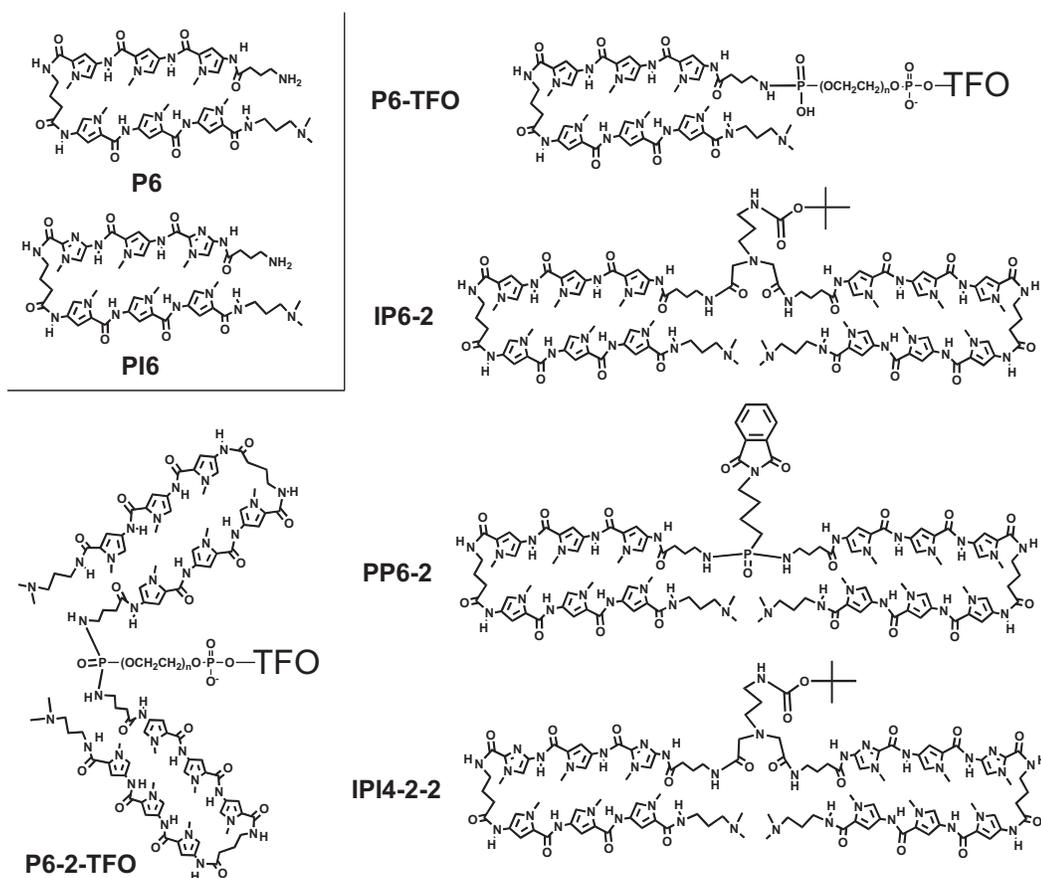
**Figure 1:** Sequences of target double-stranded DNA (A) and triplex-forming oligonucleotides (B). (A) Potential binding sites for oligopyrrolecarboxamides are marked by dashed rectangles, binding sites for oligo(pyrrole/imidazole)carboxamides – by solid rectangles. Triple helix oligonucleotide binding site is marked in *italics* by larger letters. (B) index m means 2'-O-methylribose, d-deoxyribose, C\*-5-methylcytosine.

Brüker AM400 NMR spectrometer. Analytical and preparative HPLC of MGB and MGB conjugates was performed using 1100 chromatography system from Agilent Technologies (USA). The absorption spectra of the products were recorded on a Kontron 923 Uvikon instrument (Bio-Tek, USA). Electrospray Q-TOF mass spectrometry (ES Q-TOF MS) was performed on a Q-Star instrument from Applied Biosystems (USA). The fluorescent and radioactive gels were visualized with an Amersham Typhoon or Molecular Dynamics PhosphorImager instruments and treated with ImageQuant 5.2 software (Molecular Dynamics) (15).

### MGB and MGB/TFO Conjugate Synthesis

Synthesis, purification, NMR, and mass-spectrometry analysis of amino-functionalized minor groove binders was carried out as it was described previously (16, 17).

**Figure 2:** Hexacarboxamides **P6**, **PI6**, and their conjugates. Sequences of triple-helix forming oligonucleotides (TFO) – see Figure 1, B; n = 3 or 6.



All the oligocarboxamides contained an amino group at the N-terminus to provide its coupling to the carboxyl or phosphate groups. Synthesis of oligonucleotide – MGB conjugates, their purification and analysis has also been described previously (15). Chemical formulas of ligands and their conjugates are presented on Figure 2.

Synthesis, purification, and characterization of *bis*-netropsin minor groove binders connected by different linkers have been described earlier (18).

#### *Synthesis of Head-to-Head bis-MGB Conjugates*

The synthesis of four different types of head-to head *bis*-oligo(N-methylpyrrole) carboxamides has been described in our recent publication (10). Here we give details of the synthetic procedures for two *bis*-MGB used in this study

**di-Tert-butyl-4-(N-phtalimido)butylphosphonate:** 1 ml of 15% potassium bis(trimethylsilyl) amide (KHMDs, 712  $\mu$ moles) and 1 ml of di-Tert-butylphosphite (138 mg; 712  $\mu$ mol) in toluene were cooled at -10 °C and mixed. After 30 min of incubation at -10 °C, N-(4-bromobutyl)phtalimide (100 mg; 356  $\mu$ mol) in 1 ml of toluene was added, and the reaction mixture was refluxed for eight hours. The organic phase was extracted twice with 0.1M HCl and once with water saturated with NaCl. The toluene was removed by evaporation. The residue was purified by column chromatography on silica gel (CH<sub>2</sub>Cl<sub>2</sub>/ethylacetate 1: 1). *Analytical data:* Yield: 66 mg (47%). ES-MS found: (M+H) 396.15; C<sub>20</sub>H<sub>30</sub>NO<sub>5</sub>P; calculated: 395.19. <sup>1</sup>H NMR (300MHz, C<sub>2</sub>D<sub>6</sub>SO): 1.3 (s, 18H); 1.4 (m, 4H); 1.7 (m, 2H); 3.5 (t, 2H, J = 6.8Hz); 7.8-7.9 (m, 4H).

**4-(N-phtalimido)butylphosphonic Acid:** di-Tert-butyl-4-(N-phtalimido)-butylphosphonate (50 mg; 126  $\mu$ moles) was dissolved in 5 ml of 1:1 mixture of methanol/2M aqueous HCl. The reaction mixture was stirred for four hours at room temperature. The aqueous phase was extracted twice with ethyl acetate and then saturated with NaCl. The solution was extracted five times with ethyl acetate. The combined organic phases were concentrated by rotary evaporation and purified by reverse phase HPLC (XTerra C18 column from Waters, 7  $\mu$ m, 7.8  $\times$  300 mm) using a linear acetonitrile gradient (0-90% CH<sub>3</sub>CN in 0.1 % TFA). *Analytical data:* Yield 32 mg (87%). ES-MS: found: (M-H) 283.22; C<sub>12</sub>H<sub>14</sub>NO<sub>5</sub>P; calculated: 283.06. <sup>1</sup>H NMR (300 MHz, C<sub>2</sub>D<sub>6</sub>SO): 1.4 (m, 4H); 1.7 (m, 2H); 3.6 (t, 2H, J = 6.8Hz); 7.8-7.9 (m, 4H).

**PP6-2:** Freshly distilled pyridine solutions of 4-(N-phtalimido)butylphosphonic acid (0.2 mg in 50  $\mu$ l, 710  $\mu$ mol), 4-dimethylaminopyridine (5 mg in 25  $\mu$ l, 41  $\mu$ mol), dipyridyl disulfide (6.6 mg in 25  $\mu$ l, 30  $\mu$ mol), and triphenylphosphine (7.9 mg in 50  $\mu$ l, 30  $\mu$ mol) were mixed and incubated for 15 min at room temperature. The ligand P6 (1.7 mg, 1.7  $\mu$ mol) was dissolved in this mixture and incubated at room temperature for 48 h. The conjugate was precipitated by a dropwise addition of the reaction mixture to 2 ml of ethyl ether and purified by reversed phase HPLC (XTerra C18 column from Waters, 7  $\mu$ m, 7.8  $\times$  300 mm) using a linear acetonitrile gradient (0-80% CH<sub>3</sub>CN 0.02 M NH<sub>4</sub>OOCCH<sub>3</sub>). *Analytical data:* Yield 490  $\mu$ g (31%). ES-MS: Found: 2256.60; C<sub>110</sub>H<sub>138</sub>N<sub>33</sub>O<sub>19</sub>P calculated: 2256.06.

#### *Bis-MGB ligands IP6-2 and IPI4-2-2*

**Linker Synthesis:** Iodoacetic acid (240 mg, 1.1 mmol) was mixed with N-Boc-1,3-diaminopropane (87 mg, 0.5 mmol) in anhydrous acetonitrile (10 ml) and incubated for 48 h at 50 °C in the presence of K<sub>2</sub>CO<sub>3</sub> (210 mg, 1.5 mmol.). The solvent was removed by evaporation and the reaction mixture was extracted with methanol/chloroform 9:1 (3  $\times$  5 ml). Combined organic phase was evaporated *in vacuo*. The crude product was washed with ethyl ether (3  $\times$  5 ml) to afford a white solid. *Analytical data:* 3-N-(BOC-aminopropyl)aminodiacetic acid. Yield: 78%

(178 mg). ES-MS: Found: (M-H) 290.12;  $C_{12}H_{22}N_2O_6$ , calculated: 290.31.  $^1H$  NMR (300 MHz,  $C_2D_6SO$ ): 1.3 (m, 9H); 1.4-1.5 (q, 2H;  $J = 7.1$ Hz); 2.6-2.7 (t, 2H,  $J = 7.3$ Hz); 2.9-3.0 (m, 2H); 3.2 (s, 4H); 6.7 (m, 1H).

**Coupling of Oligocarboxamides:** O-7-azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium-hexafluorophosphate (HATU, 3mg, 7.9  $\mu$ mol), dimethylamino-pyridine (DMAP, 1mg; 8.2  $\mu$ mol) in dimethylformamide (DMF, 30  $\mu$ l), and 1  $\mu$ l of N,N-diisopropylethylamine (DIPEA) were added to a solution of 3-N-(BOC-aminopropyl)aminodiacetic acid (0.5 mg; 1.7  $\mu$ mol) in DMF (50  $\mu$ L). The reaction was incubated under agitation for five minutes and the minor groove binder (3.5 mg; 3.5  $\mu$ mol) in 20  $\mu$ l of DMF was then added. After 4 h at room temperature, the mixture was placed drop by drop into 1-2 ml of diethyl ether. Yellow precipitate of product was formed. The crude products were purified by HPLC on a C-18 X-terra column (Waters, 7  $\mu$ m, 7.8  $\times$  300 mm) in a linear 0-90% water/acetonitrile gradient with 0.1% trifluoroacetic acid. The flow rate was 2 ml/min. *Analytical data:* IP6-2. Yield 59% (2.3 mg), ES-MS: Found: 2264.21;  $C_{110}H_{146}N_{34}O_{20}$ , calculated: 2264.55. IPI4-2-2. Yield 42% (1.6 mg). ES-MS: Found: 2268.14;  $C_{106}H_{142}N_{38}O_{20}$ , calculated: 2268.50.

#### *Studies of Complex Formation Between Conjugate and Target dsDNA*

The DNA/MGB complex formation was followed by non-denaturing gel electrophoresis with fluorescent or  $^{32}P$ -labeled DNA target fragment (HIV-Loop, GC, or GCi) and various concentrations of conjugates (10, 15). Samples containing 1 nM DNA fragment and different concentrations of conjugate (between 0.01 and 1000 nM) in 0.05 M HEPES buffer, pH 7.3, were loaded on 20% non-denaturing polyacrylamide gel and electrophoresis was carried out at 5 W during 6 h. Calculation of dissociation constants, as well as triple helix gel retardation studies, have been described in our previous publication (15).

#### *Thermal Denaturation Assay*

The UVIKON XL (Secomam, a Nova Analytic Company) instrument with thermostated mobile cell holder for twelve cells and the LifePower™ software (DuS-oTec GmbH) for managing the thermal denaturation program were used. Thermal denaturation assay was carried out in 0.01 M cacodylate buffer, pH 6.0, 0.1 M NaCl, 5 mM  $MgCl_2$  in Hellma quartz Suprasil cells QS114B with a 10 mm light path and Teflon hermetic caps. The concentrations of the strands in each sample were 1.3  $\mu$ M for target duplex HIV-D and 5  $\mu$ M for the MGB. MGB was dissolved in methanol and added to the sample before experiment. The sample temperature was changed at 0.1  $^{\circ}C/min$  and the absorption was recorded every 200 s at 260, 320, and 580 nm. The treatment of the melting curves was carried out using KaleidaGraph and Microsoft Excel softwares.

#### *DNA Fragments*

492 b.p. DNA fragments were prepared by digestion of a modified plasmid pGEM7(f+) containing an insert of synthetic oligonucleotides within the polylinker (19-21) with restriction endonucleases NcoI and ApaI (fragment 1552, see Supplementary Materials). The radioactive label was introduced into 3'-end of the fragment using [ $\alpha$ - $^{32}P$ ]dATP or [ $\alpha$ - $^{33}P$ ]dATP, unlabeled dNTPs, and Klenow fragment of *Escherichia coli* DNA polymerase I (22). DNA fragments were isolated by electrophoresis in 5% denaturing polyacrylamide gel (PAG).

#### *Footprinting with DNase I*

To prepare the complex, a solution of a labeled fragment (10  $\mu$ l, about  $10^4$  Bq, DNA concentration was 20  $\mu$ M in base pairs) in 10 mM Tris-HCl (pH 6.0) containing

0.1 M NaCl was mixed with a ligand solution (10  $\mu$ l, concentrations are given in figure captions) in the same buffer and cooled to 0 °C. The mixture was incubated at 0 °C at least one hour to provide complex formation (15). A solution (20  $\mu$ l) of DNase I (Sigma, about 0.1  $\mu$ g/ml) in 10 mM Tris-HCl (pH 6.0) containing 0.1 M NaCl and 5 mM MnCl<sub>2</sub> was added, and the mixture was kept for 3 min at 0 °C. The reaction was quenched by a solution (85  $\mu$ l) containing 0.15 M NaCl, 50 mM Tris-HCl (pH 7.5), 10 mM EDTA, and 10  $\mu$ g/ml tRNA. The mixture was extracted with phenol, the DNA was precipitated with ethanol, washed with 70% ethanol, dried, and dissolved in 95% formamide (1  $\mu$ l) containing 15 mM EDTA (pH 8.0), 0.05% bromophenol blue, and 0.05% xylene cyanol FF. The mixture was heated for 1 min at 90 °C, rapidly chilled, and loaded on a 6% denaturing polyacrylamide gel (40 cm long) with a gradient thickness of 0.15-0.45 mm (19, 20). Gel was performed for 55 min at 100 W (2.3 kV), fixed in 10% acetic acid, and dried on a glass plate preliminarily treated with  $\gamma$ -methacrylpropyloxysilane (LKB, Sweden). The results were analyzed on a PhosphorImager instrument (Molecular Dynamics).

#### Circular Dichroism Measurements

The CD spectra were recorded on a Jasco Model 715 instrument using 0.1, 0.2, and 1.0 cm pathlength cells. Unless specified, all measurements were carried out at 20 °C in 1 mM sodium cacodylate buffer (pH 7.0) in the presence of 0.1 M NaCl in the wavelength intervals between 220 and 400 nm.

Poly[d(AT)]·poly[d(AT)], poly(dA)·poly(dT) or oligonucleotide duplex HIV-D (Fig. 1) were dissolved in the above mentioned buffer, and ligand solution in the same buffer in case of oligonucleotide-MGB conjugates or in methanol in case of *bis*-MGB conjugates was added. The concentration of target DNA and ligand/DNA concentration ratio are indicated for each individual case in the legends to figures. Difference CD spectra were obtained by subtraction of the CD spectrum for target DNA alone from the CD spectra of ligand-DNA complex.

#### Determination of Thermodynamic Parameters for Interaction of *bis*-hairpin Minor Groove Binder with the Oligonucleotide Duplex HIV-D

The thermodynamic parameters for MGB-DNA interaction were calculated from both footprinting and CD data. Analysis of footprinting diagrams is carried out as described in previous publications (19, 23). For CD data, the following calculation scheme was used. It was assumed that the ligand could form both mono- and bidentate complexes. In bidentate mode, it can bind to two AT-clusters in the duplex, which are eventually separated by a single G-C-pair. The HIV-D DNA oligomer contains two overlapped interaction sites for bidentate binding of IP6-2 and has three potential sites for monodentate binding, which can be occupied simultaneously by three IP6-2 molecules. The occupancy of the oligonucleotide duplex HIV-D by the ligand using one or two hairpin polyamide modules for binding to the duplex can be calculated from the following equations:

$$R_1 = (K_{11} + K_{12})m/Z \quad [1]$$

$$R_2 = (7K_2m + 18K_2^2m^2 + 9K_2^3m^3)/Z \quad [2]$$

$$R_{12} = (K_{11} + K_{12})K_2m^2/Z \quad [3]$$

$$Z = 1 + (K_{11} + K_{12})m + 7K_2m + 9K_2^2m^2 + (K_{11} + K_{12})K_2m^2 + 3K_2^3m^3 \quad [4]$$

$$m = C - O(R_1 + R_2 + 2R_{12}) \quad [5]$$

$$\Delta D/O = \Delta\varepsilon_1R_1 + \Delta\varepsilon_2R_2 + (\Delta\varepsilon_1 + \Delta\varepsilon_2)R_{12} \quad [6]$$

Here  $O$  is the molar concentration of the DNA oligomer,  $Z$  is the grand partition function for the system under study. Each term of the grand partition function refers to the statistical weighting factor for a particular state of the thermodynamic system.  $R_1$  and  $R_2$  refer to the occupancies of the DNA oligomer by the ligand for bidentate and monodentate binding modes, respectively.  $R_{12}$  refers to the occupancy of the HIV-D oligomer by two ligands using bidentate and monodentate binding modes simultaneously.  $K_{11}$  and  $K_{12}$  are the equilibrium affinity constants for bidentate binding of the ligand to two target sites 5'-TAAAAGAAA-3' and 5'-TTTTTAAA-3', respectively [previous data (10) and footprint studies demonstrated that they are different].  $\Delta D/O$  is the measured CD amplitude at 330 nm expressed per mole of DNA oligomer and 1 cm pathlength cell.  $\Delta\epsilon_1$  and  $\Delta\epsilon_2$  are the molar dichroism values for bidentate and monodentate modes of binding, respectively ( $\Delta\epsilon_1 = 370 \pm 10$ ,  $\Delta\epsilon_2 = 185 \pm 10$ ). Thermodynamic parameters were determined from CD titration curves at 330 nm by iterative least square fitting of the experimental plots of  $\Delta D/O$  versus  $C/O$  to Eqs. [1-6]. Average data from several independent experiments were calculated.

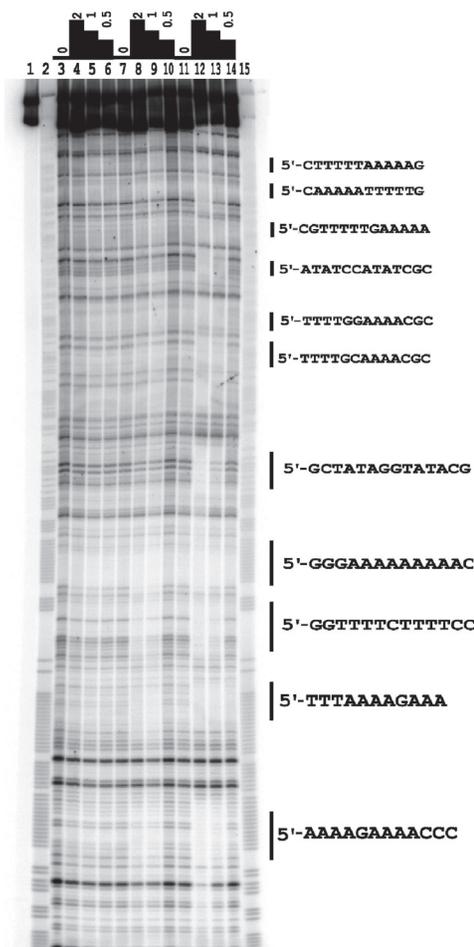
## Results and Discussion

### bis-Hairpin Minor Groove Binders

In our previous studies, we looked for an efficient ligand able to recognize and to bind with a high affinity to a specific sequence of HIV proviral DNA. We used HIV polypurine tract (PPT) as a main model target for DNA-binding ligands and 16-mer triplex-forming oligonucleotide TTTTCTTTTCCCCCT with natural or modified backbone as a model TFO (15, 24). However, the triple complex was not stable enough to permit its practical applications, especially for *in vivo* experiments (dissociation temperature 25° and slightly acid pH for natural DNA or 2'-O-methylribooligonucleotides). Only modified oligonucleotides such as phosphoroamidates or LNA demonstrated sequence-specific effects in the cell cultures (25, 26). The second class of dsDNA binding molecules was also used in our studies: oligo(*N*-methylpyrrole/*N*-methylimidazole) carboxamides developed mainly by P. Dervan's group (8) (for simplicity, the terms "pyrrole" and "imidazole" will be used). In particular, we have synthesized several versions of head-to-head *bis*-conjugates of standard hairpin hexapyrroles that demonstrated remarkably high affinity for the target sequence ( $K_d = 4.5 \cdot 10^{-9}$  M) and recognition sequence length of 8-10 base pairs (10). But the experiments have been done only on a short target HIV PPT fragment; no information was obtained concerning binding specificity and sequence preferences on a long DNA molecule that can possess several different (A/T)-rich sequences.

### DNase I Footprinting Studies

In order to compare the binding affinities of sequence-specific ligands to different (A/T)-rich sites, we have used cloned synthetic DNA fragments as substrates for footprinting studies. The 492 bp DNA sequence applied for these studies is shown in Supplementary Materials accessible on-line. The oligonucleotide inserts involved: (i) HIV polypurine tract sequence; (ii) a contiguous track of nine successive A·T/T·A pairs; and (iii) oligonucleotide fragments with pseudosymmetrical sequences containing tetramers and pentamers with different combinations of A·T/T·A pairs that are separated by zero, one, two, and three G·C/C·G pairs. Typical DNase I footprint patterns for complexes of *bis*-hairpin minor groove binder IP6-2 are presented on Figure 3. As it is seen from the figure, IP6-2 at less than micromolar concentrations demonstrate dose-dependent protection of target sequences in the A/T-rich regions containing two tetramer or pentamer A·T/T·A pairs separated by zero, one, and two G·C/C·G pairs. Each protected zone contains 12-13 base pairs. The strongest affinity site for IP6-2 has the sequence 5'-TTTAAAGAAA-3' that corresponds to HIV polypurine tract. Full protection at this region is observed at a concentration level as low as 0.5  $\mu$ M. Clear footprints are produced at sequences 5'-AAAAGAAAACCC-3', 5'-GGTTTCTTTTCC-3', and



**Figure 3:** DNase I footprints produced by the minor groove binder IP6-2 (for chemical structure see Figure 2) (lanes 11-14) and *bis*-netropsins Lys-Gly-(Py)<sub>2</sub>-Gly<sub>3</sub>-(Py)<sub>2</sub>-Dp (lanes 3-6) and Lys-Gly-(Py)<sub>2</sub>-NH-(CH<sub>2</sub>)<sub>5</sub>-NH-(Py)<sub>2</sub>-Gly-Lys (lanes 7-10). Experimental conditions are described in **Materials and Methods** section. Fragment 1552 was used as a target DNA (for sequence, see Supplementary Materials). DNA concentration was 20 μM (base pairs), ligand concentration were 0 μM (lanes 3, 7, 11), 0.5 μM (lanes 6, 10, 14), 1 μM (lanes 5, 9, 13), and 2 μM (lanes 4, 8, 12). Lane 1, non-hydrolyzed DNA; lane 2, 15, A+G sequence markers.

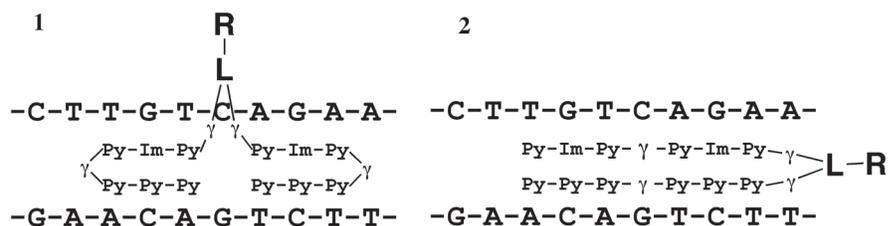
5'-GGAAAAAAAAA-3' at a concentration of about 1 μM. Protection in DNA regions harboring the sequences 5'-TTTTGCAAAA-3', 5'-TTTTGGAAAA-3', 5'-TATAGGTATA-3', 5'-TTTTCCA AAAA-3', 5'-ATATCCATAT-3', 5'-TTTTT-GAAAAA-3', 5'-TTTTTAAAAA-3' requires higher concentrations of IP6-2. Interesting, protection effect is very weak when two blocks of four A/T-base pairs are separated by three G/C/C-G-pairs.

Thus, according to DNase I footprint data, *bis*-MGB conjugate has a high affinity for any sequence that contains two blocks of at least four contiguous (A·T/T·A) base pairs eventually separated by maximum two (G·C/C·G) base pairs. The best target on the native DNA seems to be HIV polypurine tract [as it has been already shown in our previous work (10)] that may have structure favorable for IP6-2 binding. For comparison, the same structure fragment AAAAGAAAA and its complementary version TTTTGTTTT while being targeted by IP6-2 show lower affinity. The difference is that they are flanked by several G/C pairs from both sides. Thus, the affinity of targeted A/T tracts depends on the sequence context, including the order of A/T-base pairs and the flanking sequences.

Looking at the results for *bis*-netropsins in which two monomers are bridged by triglycine or aliphatic diamine residue [Lys-Gly-(Py)<sub>2</sub>-Gly<sub>3</sub>-(Py)<sub>2</sub>-Dp and Lys-Gly-(Py)<sub>2</sub>-NH-(CH<sub>2</sub>)<sub>5</sub>-NH-(Py)<sub>2</sub>-Gly-Lys, respectively] one can conclude that binding of these compounds exhibits far less protection effects as compared to that shown by IP6-2. For *bis*-netropsin where two netropsin analogs are bridged by an aliphatic diamine residue (Figure 3, lane 8), clear footprints are produced at concentration 1 μM (Figure 3, lane 8). As a rule, the protected regions occur near zones protected by IP6-2. However, the diamine-bridged *bis*-netropsin protects adenine residues within the sequence 5'-TTTTGGGAAAA-3', which is poorly protected by IP6-2. A closely related compound PP6-2 (Fig. 2), an analog of IP6-2 containing shorter linker between two polyamide hairpins, binds to DNA less strongly than IP6-2. Footprint studies show that PP6-2 and IP6-2 exhibit similar sequence preferences (data not shown). As in case of IP6-2, the preferred site for binding of PP6-2 on the DNA fragment is also the region of HIV PPT tract: 5'-TTTAAAAGAAAAA-3'.

#### Circular Dichroism Studies

Two orientations of *bis*-ligands in the minor groove are possible: (i) two antiparallel hairpins situating in the minor groove at adjacent DNA sequences or (ii) a long parallel hairpin containing two parallel strands. In our first tests we used only oligopyrrole minor groove binders profiting from the presence of long (A/T) blocks in the targeted region. For them, both orientations are in agreement with Dervan's rules. However, for construction of *bis*-ligands that contain pyrrole and imidazole rings, the ligand-binding mode has to be known. For different orientations, the MGB sequence should be completely different, as it is seen from the example in Figure 4. Several arguments (as tolerance of two GC-base pairs between two clusters of four AT-pairs) approved indirectly antiparallel *bis*-hairpin configuration (10); however, the direct proof had to be obtained.



**Figure 4:** Two possible configurations of head-to-head *bis*-oligocarboxamide ligands in the DNA minor groove: 1, two short antiparallel hairpins; 2, long parallel hairpin. R, protection or functional group; L, linker; γ, residue of γ-aminobutyric acid.

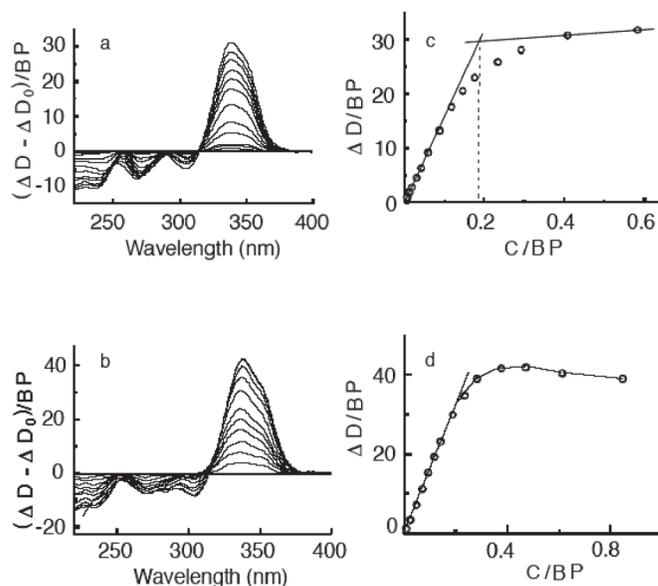
In order to confirm first orientation, circular dichroism studies and molecular modeling of the complex structure were carried out. CD method permits to distinguish between parallel and antiparallel alignment of two oligopyrrolicarboxamide fragments bound in the side-by-side manner to DNA. In our previous publications (27, 28) we used a model of head-to-tail and head-to-head *bis*-netropsin derivatives and demonstrated that these molecules are able to bind to (A/T)-rich sequences in different conformations: extended, parallel double-stranded (for head-to-head), and antiparallel double-stranded (for head-to-tail) forms. The CD spectral profiles reflecting the binding of *bis*-netropsins in the parallel-stranded hairpin form are quite different from those observed for binding in the extended conformation and antiparallel side-by-side hairpin forms. The CD difference spectrum of *bis*-netropsin bound in parallel-stranded hairpin form exhibits two positive peaks at 330 and 290 nm with a negative peak at 310 nm. When two netropsin strands are connected in a head-to-tail orientation and bind to DNA as an antiparallel hairpin, the negative peak at 310 nm disappears, and amplitude of the positive peak at 290 nm decreases transforming it into a shoulder. Finally, when a *bis*-netropsin binds in an extended conformation using two netropsin-like fragments, only one positive peak at 314 nm is observed, with a negative peak near 280 nm (28).

As additional control, in the present studies we used an antiparallel double-stranded MGB **P8** ( $\gamma$ -Py<sub>4</sub>- $\gamma$ -Py<sub>4</sub>-Dp), which represents two blocks of tetrapyrrole carboxamides, connected "head to tail" by  $\gamma$ -aminobutyric acid linker. It is known to interact with DNA in antiparallel hairpin orientation (6). Figure 5 shows CD spectra of **P8** complexes with poly[d(AT)]·poly[d(AT)] (a,c) and poly(dA)·poly(dT) (b,d). A peak at 330 nm that shifts to red by 4 nm upon titration, as well as absence of positive peaks at 290 or 314 nm, clearly indicate an antiparallel double-stranded orientation.

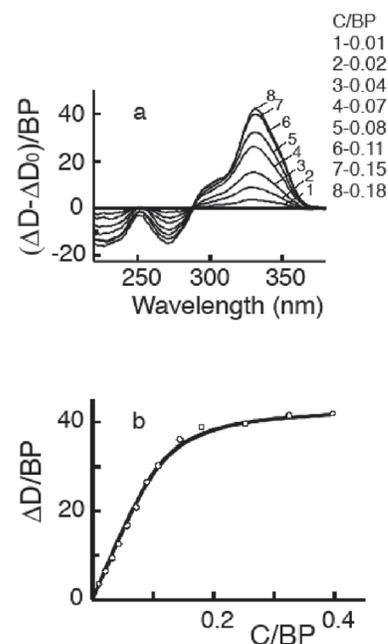
Molar dichroism value calculated from the initial slope of the titration curve is 160-170. It means that both tetrapyrrole fragments, but may be not all the cycles are implicated in the interaction with a polymer. The stoichiometry of the complex appears to be one ligand molecule per six A·T base pairs. This excludes a possibility that **P8** binds to DNA in the extended conformation.

Figure 6 shows the difference CD spectra for complexes of the minor groove binder IP6-2 with poly[d(AT)]·poly[d(AT)]. At low C/BP values (C/BP < 0.07) the CD spectra exhibit an isodichroic point at 290 nm. The shapes of difference CD patterns are characteristic of a polyamide bound to DNA in the hairpin form with antiparallel orientation of two pyrrolicarboxamide fragments. In order to determine thermodynamic parameters from experimental CD titration curves it is convenient to compare the experimental and theoretically calculated plots of  $\Delta D/BP$  versus C/BP (Figure 6b). Here  $\Delta D/BP$  is the CD amplitude of the ligand-DNA complex measured at a given wavelength and calculated per 1 cm pathlength cell and one mole of DNA base pairs. If the ligand can bind to poly[d(AT)]·poly[d(AT)], using either one or two hairpin polyamide modules, then  $\Delta D/BP$  can be calculated from Eq. [7]:

**Figure 6:** (a) Difference CD spectra of IP6-2 complexes with poly[d(AT)]·poly[d(AT)] (126  $\mu$ M in base pairs). (b) CD titrations of poly[d(AT)]·poly[d(AT)] (5.7  $\mu$ M) by IP6-2.  $\Delta D/BP$  refers to the measured CD amplitude at 330 nm calculated per 1 cm pathlength cell and divided by the molar concentration of DNA base pairs. C/BP is the molar ratio of the ligand to DNA base pairs. The calculated CD titration curve exhibiting the best fit with the experimental data (open circles) is shown by a solid line.



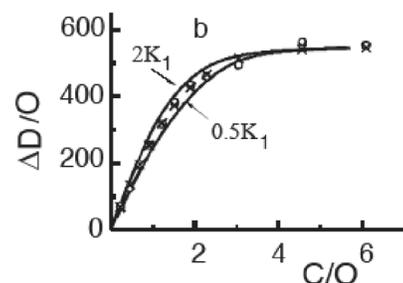
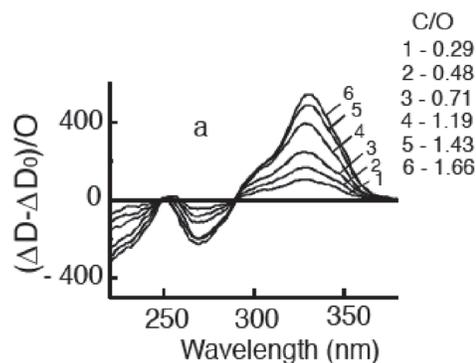
**Figure 5:** Difference CD spectra of complexes of **P8** ( $\gamma$ -Py<sub>4</sub>- $\gamma$ -Py<sub>4</sub>-Dp, where Py is the N-methylpyrrolicarboxamide unity,  $\gamma$  is the aminobutyric acid residue, Dp- is the N,N'-dimethylaminopropylamino residue) with poly(dA)·poly(dT) (a) and poly[d(AT)]·poly[d(AT)] (b). CD titrations of poly(dA)·poly(dT) (c) and poly[d(AT)]·poly[d(AT)] (d) by **P8**.  $\Delta D/BP$  refers to the CD amplitude at 330 nm calculated per 1 cm pathlength cell and divided by molar concentration of DNA base pairs. C/BP is the ligand to DNA base pairs molar ratio. Concentration of the DNA is 20  $\mu$ M (base pairs).



$$\Delta D/BP = r_1 \Delta \epsilon_1 + r_2 \Delta \epsilon_2 \quad [7]$$

where  $\Delta \epsilon_1$  and  $\Delta \epsilon_2$  are the molar dichroism values for bidentate and monodentate modes of binding,  $r_1$  and  $r_2$  refer to the molar ratios of bound ligand to DNA base pairs calculated for bidentate and monodentate modes of binding, respectively.

Let  $K_1$  and  $K_2$  be the equilibrium association constants corresponding to the bidentate and monodentate modes of binding of IP6-2 to an isolated site on poly[d(AT)]·poly[d(AT)]. It is assumed that each bound ligand occupies  $L_1$  and  $L_2$  consecutive base pairs when two or one hairpin polyamide modules of the ligand are implicated in the interaction with DNA, respectively. The procedure used for determination of thermodynamic parameters of binding exhibits some resemblance with that outlined earlier for a different thermodynamic model (27). The theoretical plots of  $\Delta D/BP$  versus  $C/BP$  were calculated from Eqs. [7-10] for different values of  $K_1$  and  $K_2$  and site sizes  $L_1$  equal to 8 or 9 and  $L_2$  equal to 4 or 5, using experimentally determined values of  $\Delta \epsilon_1$  and  $\Delta \epsilon_2$  at 330 nm ( $\Delta \epsilon_1 = 370 \pm 10$ ,  $\Delta \epsilon_2 = 185 \pm 10$ ).



**Figure 7:** (a, b) Difference CD spectra of IP6-2 complexes with the oligonucleotide duplex HIV-D (10.5  $\mu\text{M}$ ). (b) CD titrations of the oligonucleotide duplex HIV-D (0.27  $\mu\text{M}$ ) by IP6-2.  $\Delta D/O$  refer to the measured CD amplitude at 330 nm calculated per 1 cm pathlength cell and divided by the molar concentration of the DNA oligomer.  $C/O$  is the molar ratio of the ligand to the oligonucleotide duplex. The best fit between experimental (open circles) and calculated (crosses)  $\Delta D/O$  values is observed when  $K_1 = 6.62 \times 10^8 \text{ M}^{-1}$  and  $K_2 = 0.75 \times 10^8 \text{ M}^{-1}$ . The calculated CD titration curves that correspond to the binding of IP6-2 to the DNA oligomer with the affinity constants  $2K_1$  and  $0.5K_1$  ( $K_1 = 6.6 \times 10^8 \text{ M}^{-1}$ ) and  $K_2 = 0.75 \times 10^8 \text{ M}^{-1}$  are shown by solid lines.

$$r_1 = K_1 m (1 - r_1 L_1 - r_2 L_2)^{L_1 - 1} / (1 - r_1 L_1 - r_2 L_2 + r_1 + r_2)^{L_1 - 1} \quad [8]$$

$$r_2 = K_2 m (1 - r_1 L_1 - r_2 L_2)^{L_2 - 1} / (1 - r_1 L_1 - r_2 L_2 + r_1 + r_2)^{L_2 - 1} \quad [9]$$

$$m = C - (r_1 + r_2)BP \quad [10]$$

Here  $m$  is the concentration of the ligand in the free solution,  $C$  is the total concentration of the ligand,  $BP$  is the concentration of DNA expressed in moles of base pairs. Eqs. [8] and [9] are derived for description of non-cooperative binding of large ligands that can form two types of the complex with a polynucleotide lattice (29).

Thermodynamic parameters were determined using iterative nonlinear least squares data-fitting procedure in which the minimum of the mean square deviation between the experimental and theoretical values of  $\Delta D/BP$  at different  $C/BP$  values was taken as a criterion of the quality of the fit between the experimental and calculated curves.

The best fit is observed when  $K_1 = 2.0 \times 10^8 \text{ M}^{-1}$ ,  $K_2 = 0.34 \times 10^8 \text{ M}^{-1}$ ,  $L_1 = 8$ , and  $L_2 = 4$  (Figure 6b). The root mean square deviation of the experimental and calculated values of  $\Delta D/BP$  is equal to 0.68. It increases up to 1.83 when  $L_1 = 9$  and  $L_2 = 4$ , provided that equilibrium association constants  $K_1$  and  $K_2$  take the same values, as before. Average data from three independent experiments are  $K_1 = (1.9 \pm 0.3) \times 10^8 \text{ M}^{-1}$  and  $K_2 = (0.21 \pm 0.15) \times 10^8 \text{ M}^{-1}$ .

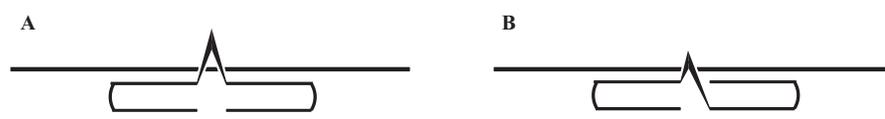
Figure 7a shows the CD difference spectra obtained for complexes of IP6-2 with the oligonucleotide duplex HIV-D. As before, the shapes of the difference CD spectral profiles are characteristic of a polyamide bound to DNA in the hairpin form with antiparallel orientation of two pyrrolecarboxamide fragments. Figure 7b shows typical CD titration curve obtained from binding of IP6-2 to the HIV-D duplex. The CD amplitude at 330 nm was measured as a function of the ligand to DNA oligomer molar ratio ( $C/O$ ).

The CD amplitudes at  $C/O$  values approaching the saturation level of binding and divided by the DNA oligomer concentration are found to be equal approximately to 550, a value which is greater than that expected for bidentate binding of IP6-2 to the duplex with 1:1 stoichiometry. Evidently, complex of the second type corresponding to the monodentate binding of IP6-2 also contributes to the overall CD signal at 330 nm. The site sizes for mono- and bidentate binding (4 and 8 to 9 AT-base pairs, respectively) were determined from the binding of the ligand to poly[d(AT)]·poly[d(AT)] and poly(dA)·poly(dT). The equilibrium association constants for monodentate and bidentate modes of binding of IP6-2 to the duplex can be determined from the CD titration curves by nonlinear least square fitting of the

experimental and calculated plots of  $\Delta D/O$  versus  $C/O$  values. Minimum of the sum of the square deviations between experimental and calculated  $\Delta D/O$  values at different  $C/O$  values was taken as a criterion for the quality of the fit. DNase I footprinting studies reveal that preferred sites for bidentate binding of the ligand correspond to DNA regions with the sequences 5'-TAAAAGAAA-3' and 5'-TTTTTAAAA-3' and that the affinity constant for the binding of IP6-2 to the first sequence is approximately four times greater than that observed for the binding to the second site. We set into Eqs. [1], [3], and [4]  $K_{11} = K_1$  and  $K_{12} = K_1/4$ . The best fit is observed when  $K_1 = 6.62 \times 10^8 \text{ M}^{-1}$  and  $K_2 = 0.75 \times 10^8 \text{ M}^{-1}$ . Here,  $K_1$  and  $K_2$  are the equilibrium association constants for bidentate and monodentate modes of binding of IP6-2 to the HIV-D duplex, respectively. The root mean square deviation of the experimental and calculated  $\Delta D/O$  values is equal to 10.4. The average data from three independent experiments are:  $K_1 = (5.76 \pm 0.96) \times 10^8 \text{ M}^{-1}$  and  $K_2 = (0.65 \pm 0.15) \times 10^8 \text{ M}^{-1}$ .

#### *Design and DNA Binding Studies of Sequence-specific bis-hairpin Minor Groove Binders*

In order to validate our conclusions, we constructed a *bis*-hairpin minor groove binder IPI4-2-2 (Fig. 2) that contains both N-methylpyrrole and N-methylimidazole carboxamide units and that is able to interact with the target DNA only in antiparallel hairpin conformation (in extended parallel conformation non-recognizing imidazole/imidazole pairs will form). We studied the affinity of this MGB to two types of target DNA: "GC" and "GC-I" (Table I). Due to sequence differences in these duplexes, IPI4-2-2 interacts with them in two different configurations that we called "direct" and "inversed". They differ by orientation of two MGB hairpins as it is shown on Figure 8.



**Figure 8:** "Direct" (A) and "inversed" (B) orientations of ligand IPI4-2-2 depending on the target sequence of DNA.

We measured the apparent dissociation constants of both complexes by gel mobility retardation experiments as it has been described earlier (15, 10). The results are shown in Table I.

**Table I**

Apparent dissociation constants of bis-MGB/target DNA complexes at 37 °C in 0.05 M HEPES buffer (pH 7.2) in the presence of 0.05 M NaCl.

Ligand	Target DNA	Target sequence	Kd, nM
IPI4-2-2	HIV-Loop	No target	No gel shift
IPI4-2-2	GC	ACACTGTGT TGTGACACA	3.6±0.5
IPI4-2-2	GC-i	ACACTCTCT TGTGAGAGA	4.1±0.8
IP6-2	HIV-Loop	TTTTTAAAAGAAA AAAAATTTTCTTT	4.8±0.8

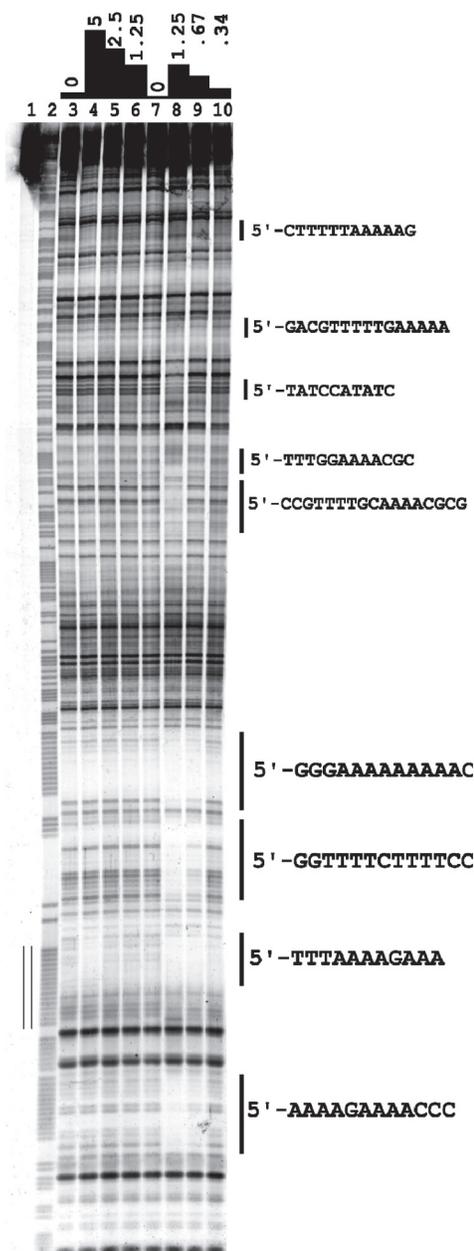
The dissociation constants of both "direct" and "inversed" complexes are almost the same and practically do not differ from that of IP6-2/HIV-Loop complex, the best of those reported previously (10) and based on the recognition of A/T sequences by oligopyrrole carboxamides. Moreover, the interaction of the ligand with both target sequences slightly but reproducibly stabilizes the duplex as it was determined by measuring of thermal dissociation temperatures of complexes (Table II). As it is seen from the table, in the presence of a cognate ligand the melting temperature of the DNA increases by 3-5 °C. Slight increase of  $T_m$  in case of HIV-Loop/IPI4-2-2 pair could be explained by monodentate interactions with GTG/CAC and GAG/CTC sequences present in flanking regions of HIV-Loop.

It must be noted, however, that similar affinity of both "direct" and "inversed" sequences can be a source of ambiguities and has to be taken into account in design of sequence-specific DNA-binding molecules.

**Table II**

Melting points  $T_m$  of DNA duplexes and their complexes with bis-hairpin oligocarboxamides.

DNA	$T_m$ , °C		
	no ligand	IPI4-2-2	IP6-2
GC	76,9	79,6	76,6
GC-i	76,4	79,3	76,4
HIV-Loop	74	76	79



Our results demonstrated that *bis*-hairpin conjugates have high affinity and sequence specificity for target double-stranded DNA. They can be used for targeting specific DNA sequences. However, relatively short length of the recognition sequence, presence of one-two degenerated base pairs in the center of the target sequence and interaction in two conformations (“direct” and “inversed”) do not permit to consider these ligands as absolutely sequence-specific ones, directed to unique DNA genomic sequence. In our previous work, following the experiments of Dervan *et al.* (12, 30), we conjugated triplex-forming oligonucleotide and oligopyrrole hairpin minor groove binder *via* tri- or hexa(ethyleneglycol) linker (long enough to join ligands in two different DNA grooves) (13, 15). When only one hairpin MGB was attached to TFO, conjugate binding to a short duplex target revealed higher stability of the complex and higher sequence specificity, though the effect on the stability was small. Attachment of two minor groove binders to the same terminal phosphate of oligonucleotide in a head-to-head orientation led to very high complex stabilities at high pH (> 8.3) and temperatures (up to 74 °C), but the main role in DNA affinity belonged to *bis*-MGB moiety (13, 15). The question was: can the conjugated triplex-forming oligonucleotide increase the *bis*-MGB specificity and deliver the construction to a unique site of long double-stranded DNA? To answer this question, we attached two P6 ligands to a triplex-forming oligonucleotide TTTTCTTTTCCCCCT interacting with HIV PPT sequence (25) and studied its interaction with long DNA fragment containing several recognition sites for free *bis*-MGB, but only one for MGB-TFO conjugate.

#### DNase Footprint Studies

The interaction of *bis*-MGB-TFO conjugate with plasmid DNA fragment has been studied by DNase I footprint method. Results were compared to DNase footprint of the same fragment in the presence of *mono*-MGB-TFO conjugate.

Figure 9 shows DNase I cleavage patterns for complexes of conjugates P6-2-TFO and P6-TFO with the same DNA fragment. Footprint patterns for complexes with P6-2-TFO look similar to those observed for complexes of the IP6-2, as if there is no oligonucleotide. As revealed from the concentration dependence of the cleavage protection effect, the strongest affinity binding site for this conjugate is located at the region HIV PPT with the sequence 5'-TTTAAAAGAAA-3'. Cleavage protection at this site is detected at a concentration level as low as 0.3  $\mu$ M. Full protection of this region extending over 12 base pairs is produced at the concentration of 0.65  $\mu$ M. P6-2-TFO also interacts strongly with other AT-tracts, with the higher protection effects seen at the sequences 5'-AAAAGAAA-3', 5'-TTTTCTTT-3', and 5'-AAAAAAA-3', and the lower effects at sites 5'-TTTTTAAAA-3' and 5'-TTTTGGAAA-3'.

Conjugate P6-TFO, containing only one polyamide hairpin module, binds mainly to the same HIV PPT region (indicated with double line at the left side on Fig. 9) protecting 5'-TTAAAAGAA-3' sequence. Interestingly, the protected DNA region extends over nine base pairs. One begins to observe clear footprint only at a concentration of 5  $\mu$ M (compare to full protection at 0.65  $\mu$ M in case of conjugate P6-2-TFO). In contrast to P6-2-TFO, DNA sites with sequences 5'-AAAAGAAAACC-3', 5'-AAAAAAA-3', and 5'-TTTTCTTT-3' exhibit no protection at the highest conjugate concentration used (5  $\mu$ M). However, low pro-

**Figure 9:** DNase I footprint in the presence of conjugates TFO-MGB: P6-TFO (lanes 3-6) and P6-2-TFO (lanes 7-10) (for designations see Fig. 2). DNA oligonucleotides with 5-methylated C were used as TFO and linked to MGB by tri(ethyleneglycol) linker. Experimental conditions are described in *Materials and Methods* section. Fragment 1552 was used as target DNA (for sequence, see Supplementary Materials). Triplex-forming HIV PPT sequence is indicated by double line from the left side of the gel. DNA concentration was 20  $\mu$ M (base pairs), ligand concentrations were 0 (lanes 3 and 7), 5 (4), 2.5 (5), 1.25 (6), 1.25 (8), 0.67 (9), and 0.34 (10)  $\mu$ M. 1, non-hydrolyzed DNA; 2, A+G sequence markers.

tection effects are seen at sites with sequences 5'-TTTTTAAAAA-3', 5'-TTTTT-GAAAAA-3', and 5'-TTTTGGAAAA-3'.

An interesting feature is that only a MGB-binding A/T tract and not triplex-forming region is protected in both cases. In the complex, interaction of DNA with the MGB part of the conjugate may decrease binding of the TFO part with the oligopurine tract in the major groove. So, one can suggest that only MGB part in *bis*-conjugate is bound to the cognate recognition site. However, gel retardation experiments show that DNA triplexes are also formed at low pH and temperatures on binding of both conjugates to the cognate duplex (15). It means that DNase I has lower affinity for DNA than MGB moiety and higher affinity than TFO moiety. If partially bound states of TFO are allowed in the complex, DNase I can cleave the target DNA from the side of the minor groove, except for the DNA region protected from cleavage by polyamide modules of the conjugate. DNase cleavage of single-stranded TFO moiety does not seem to be probable, in addition, when we compared footprints of conjugates containing 2'-O-methyl RNA as TFO moieties, no difference was found. The fact that the sizes of the protected regions in complexes with IP6-2 and P6-2-TFO are approximately equal can be attributed to a loosely coupled triplex structure formed by TFO with the target DNA site.

Footprints produced by P6-TFO, P6-2-TFO, and their analogs containing longer hexa(ethyleneglycol) linker between TFO and hairpin polyamide modules were also compared. No difference of their binding specificity and affinity was noted.

In conclusion, *bis*-conjugate P6-2-TFO shows a protection pattern similar to that of *bis*-MGB IP6-2 alone. P6-TFO has lower affinity, but higher sequence specificity, protects less secondary sites and exhibits an opposite order of secondary site protecting activity. These results suggest also that both components of the *mono*-MGB-TFO conjugate play a role in the target recognition.

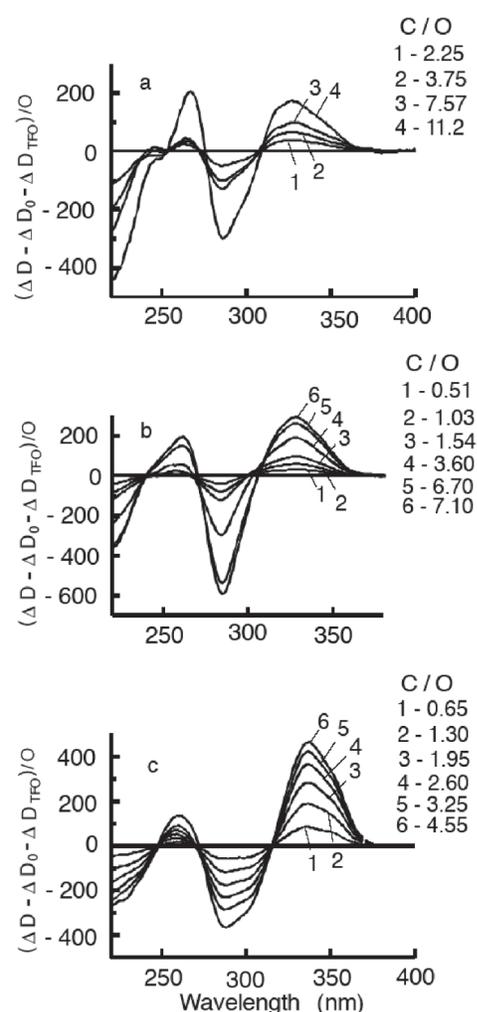
#### Circular Dichroism Studies of *bis*-MGB-TFO Conjugates

For complexes of MGB-TFO conjugates with the oligonucleotide duplex HIV-D the target duplex bound to the MGB and TFO part contribute into the overall CD signal. In order to compare CD spectral profiles characteristic of bound MGB in complexes formed by different MGB-TFO conjugates with HIV-D duplex, we calculated CD difference spectra by subtraction of the CD spectrum of the DNA oligomer alone and the spectrum of TFO alone from CD spectra of complexes formed by MGB-TFO conjugates with the HIV-D oligomer. The calculated difference CD spectra obtained at different C/O values are shown in Figure 10.

All the spectra are quite similar and contain positive CD band near 330 nm and negative band near 280 nm. The CD spectral profiles are characteristic of MGB binding in a hairpin form built on the basis of side-by-side motif with antiparallel orientation of two pyrrolecarboxamide fragments. A general conclusion drawn from CD studies is that in complexes of *bis*-MGB-TFO conjugates with the target duplex a linear parallel side-by-side pyrrolecarboxamide motif is not formed, and the ligands bind to DNA using one or two hairpin modules with antiparallel orientation of two oligopyrrolecarboxamide fragments.

#### Molecular Modeling Studies

We used molecular modeling for obtaining more detailed structural information about interaction of IP6-2 with the DNA duplex harboring AT-tract and oligopurine sequence from HIV genome. The AMBER force field developed by Weiner *et al.* for modeling nucleic acids and widely used in the DNA/RNA field can be also applied to model complexes of DNA with polyamides and conjugates *bis*-MGB-TFO (31). A suitable modification of AMBER force field to model polyamide-DNA



**Figure 10:** Difference CD spectra of complexes of mono-conjugate P6-TFO (a) and *bis*-linked conjugates P6-2-TFO (b) and P8-2-TFO (c) with the oligonucleotide duplex HIV-D. Both contributions of duplex ( $\Delta D_0$ ) and TFO ( $\Delta D_{TFO}$ ) are subtracted. C/O is the molar ratio of the MGB to the duplex (indicated on the spectra). DNA oligonucleotides with 5-methylated C were used as TFO linked to MGB by tri(ethyleneglycol) linker. Duplex concentration was 9  $\mu$ M.

complexes is reported elsewhere (32, 33). Molecular modeling of the complexes of IP6-2 analogue and conjugate P6-2-TFO with the HIV-D oligonucleotide duplex was carried out, using modified Amber 4.1 force field. The analogue is composed of N-methylpyrrole carboxamide units and contains no charged N-dimethylamino groups at C-termini of the two pyrrolecarboxamide modules of the IP6-2 molecule. For the electrostatic component of the empirical energy function the effect of solvent was approximated by a distance-dependent dielectric function and by reducing the net charge on the phosphate group to  $-0.46 e$ . New atom types were created for the nitrogen atom bound to the phosphate atom as well as for nitrogen atom of the cytosine ring at the site of protonation. Force field parameters for new atom types were developed by comparing parameters found in the literature with the structural features of published crystal structures. The length of P-N bond is assumed to be equal to  $1.68 \text{ \AA}$  with the force constant of  $230 \text{ kcal/mol}\cdot\text{A}^2$ . Relevant parameters for torsion angles, NPN bond angle, and bending force constant were also introduced. A set of fractional atomic charges of the protonated 5-methyl cytosine and modified phosphate group containing one or two NH groups instead of oxygen atoms were different from typical AMBER values for cytosine and phosphate group.

Molecular mechanics calculations show that two polyamide modules of IP6-2 can be inserted simultaneously into the minor groove of the DNA oligomer and interact with two runs of four A·T-base pairs separated by a G·C base pair in the oligopurine track of the duplex. In the complex, the guanine 2-amino group is projected into the interface between the two polyamide hairpin modules. The connecting chain between the hairpin modules of IP6-2 extends over two base pairs. In the complex, one of the polyamide hairpins interacts with a tetramer TTAA, whereas another hairpin binds to a site GAAA. This may explain footprinting data showing that IP6-2 binds most tightly to the target site with the sequence TTAAAAGAAAA and exhibits lower affinity to the site GGAAAAGAAAA. The structure of the complex is stabilized by Van-der-Waals contacts with the floor and both sides of the minor groove, electrostatic interactions and hydrogen bonding of the NH groups of two hairpin pyrrolecarboxamide modules to the adenine N3 and thymine O2 atoms, including acceptor sites of thymine and adenine residues at 5'-end of the tetramer TTAA. In order to accommodate hairpin polyamide modules of IP6-2 into the DNA minor groove the width of the minor groove should be increased by approximately  $2 \text{ \AA}$ . In the complex, two hairpin modules are approximately related by two-fold rotation symmetry.

**Figure 11:** Molecular models of binding of *bis*-MGB and *bis*-MGB-TFO conjugate to the target DNA. DNA-binding conformation is suitable for specific interaction of the polyamide hairpin modules with A/T-clusters in the minor groove and TFO binding in the major groove to the oligopurine tract.

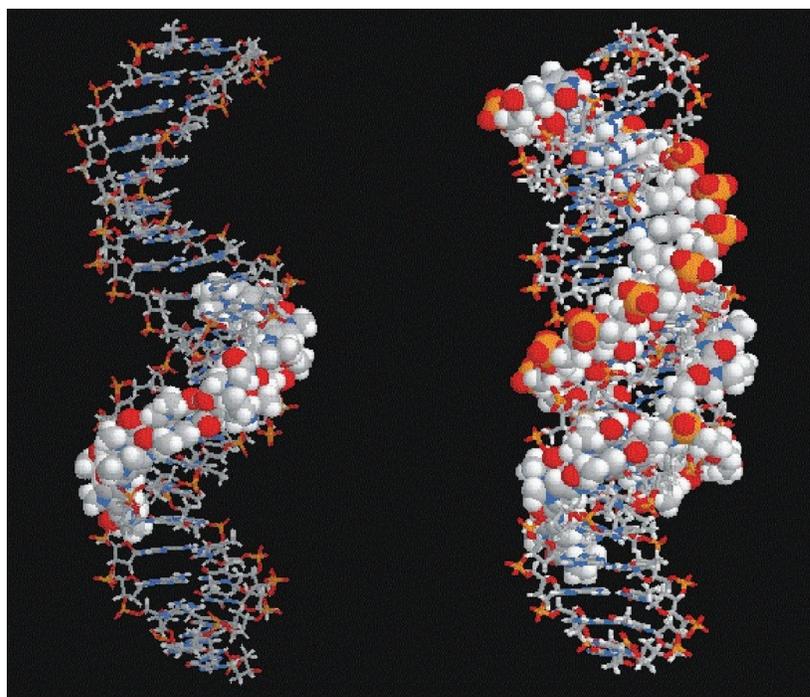


Figure 11 shows the proposed model for binding of IP6-2 and P6-2-TFO conjugate to the HIVD duplex containing the target site for the conjugates. In the complex, two hairpin polyamide modules of MGB occupy spatial positions similar to those in the complex with IP6-2. In addition, the TFO can form Hoogsteen base pairs with bases of the oligopurine tract in the major groove. These results are also confirmed by calculations based on MMFF94 force field using MOE software (34).

Molecular mechanics calculations shows that tri(ethyleneglycol) and  $\gamma$ -aminobutyric acid linkers are suitable to permit simultaneous binding of TFO and *bis*-MGB in the major and minor groove, respectively. Conjugates with longer hexa(ethyleneglycol) linker can be placed on the target DNA more nicely. Optimization of the linker structure is required to increase both the strength and the specificity of binding of a conjugate to the target DNA site. One obvious modification involves introduction into the linker of a chemical group capable of hydrogen bonding to the guanine 2-amino group.

### Conclusions

By using both physical and molecular biology methods we have shown that head-to-head *bis*-conjugates of oligocarboxamide minor groove binders recognize a number of target sequences within the double-stranded DNA according to established recognition rules. However, the binding affinity could be different for the same sequences and depends not only on the target sequence itself, but also on its context: the flanking sequences, the order, and the orientation of the target base pairs. *Bis*-MGB conjugates bind to target DNA using mode of two short antiparallel hairpins, both in free and conjugated to TFO forms. These results are validated by the synthesis of sequence-specific *bis*-MGB that recognizes and binds tightly to a designed target sequence containing all the four bases. The development of *bis*-hairpin minor groove conjugates is a perspective way for construction of highly specific dsDNA-binding agents with high affinity. Combinatorial approach and DNA-templated organic synthesis (35, 36) could be of great utility on this way.

Oligonucleotide-TFO conjugates without serious modifications of components do not provide a great improvement for the affinity and specificity simultaneously. In case of *mono*-MGB-TFO conjugate, the binding is highly sequence-specific and profits from recognition profile of both MGB and TFO. However, the affinity of the complex is not very high, and low temperatures and slightly acid pH are necessary to preserve the complex. *Bis*-MGB-TFO conjugates form very stable complexes with their target DNA. However, the oligonucleotide part plays a minimal role in the recognition and binding in this case. Thus, the recognition site for conjugates is determined by *bis*-MGB binding. It is possible that the use of modified oligonucleotides [PNA, phosphoramidates (24, 37), LNA (25, 26, 38)] or attachment of strong DNA binder (such as intercalators) will increase affinity of conjugates.

### Acknowledgments

This work was supported by INSERM, CNRS, European Community (project INTAS 2001-0638), Russian Academy of Sciences (Program for molecular and cell biology) and Russian Foundation for Basic Research (project 07-04-01031). We express our acknowledgements to Dr. J. P. Brouard, L. Dubost, and A. Marie for mass spectra, Dr. A. Blond for NMR spectra, Dr. A. Nikitin for MOE calculations and permission to use his software for visualizing of molecular structures, and Chemical Group Inc. for permission to use MOE software. The work of L. Boulay and J. Lavaury on conjugate synthesis and thermal denaturation is also acknowledged.

## References and Footnotes

1. Sun, J. S., Garestier, T., Hélène, C. *Curr Opin Struct Biol* 6, 327-333 (1996).
2. Praseuth, D., Guieysse, A. L., Hélène, C. *Biochim Biophys Acta - Gene Struct Express* 1489, 181-206 (1999).
3. Giovannangeli, C., Hélène, C. *Curr Opin Mol Ther* 2, 288-296 (2000).
4. Parks, M. E., Baird, E. E., Dervan, P. B. *J Am Chem Soc* 118, 6153-6159 (1996).
5. Turner, J. M., Baird, E. E., Dervan, P. B. *J Am Chem Soc* 119, 7636-7644 (1997).
6. Dervan, P. B., Bürl, R. W. *Curr Opin Chem Biol* 3, 688-693 (1999).
7. Dervan, P. B. *Bioorgan Med Chem* 9, 2215-2235 (2001).
8. Dervan, P. B., Edelson, B. S. *Curr Opin Struct Biol* 13, 284-299 (2003).
9. Marques, M. A., Doss, R. M., Foister, S., Dervan, P. B. *J Amer Chem Soc* 126, 10339-10349 (2004).
10. Halby, L., Ryabinin, V. A., Sinyakov, A. N., Boutorine, A. S. *Bioorg Medicinal Chem Letter* 15, 3720-3724 (2005).
11. Ryabinin, V. A., Boutorine, A. S., Hélène, C., Pyshnyi, D. V., Sinyakov, A. N. *Nucleos Nucleot Nucleic Acids* 23, 789-803 (2004).
12. Szewczyk, J. W., Baird, E. E., Dervan, P. B. *Angew Chem Int Ed* 35, 1487-1489 (1996).
13. Sinyakov, A. N., Ryabinin, V. A., Grimm, G. N., Boutorine, A. S. *Mol Biol Engl Tr* 35, 251-260 (2001).
14. Zhilina, Z. V., Ziemba, A. J., Trent, J. O., Reed, M. W., Gorn, V., Zhou, Q., Duan, W., Hurley, L., Ebbinghaus, S. W. *Bioconjug Chem* 15, 1182-1192 (2004).
15. Novopashina, D. S., Sinyakov, A. N., Ryabinin, V. A., Venyaminova, A. G., Halby, L., Sun, J. S. Boutorine, A. S. *Chemistry & Biodiversity* 2, 936-952 (2005).
16. Ryabinin, V. A. Sinyakov, A. N. *Bioorg Khim* 24, 601-607 (1998).
17. Sinyakov, A. N., Feshchenko, M. V., and Ryabinin, V. A. *Russ J Bioorg Chem* 29, 402-404. (2003).
18. Grokhovsky, S. L., Gottikh, B. P., and Zhuze, A. L. *Bioorg Khim* 18, 570-583 (1992).
19. Grokhovsky S. L., Surovaya A. N., Burkhardt G., Pismensky V. F., Chernov B. K., Zimmer Ch., Gursky G. V. *FEBS Letters* 439, 346-350 (1998).
20. Sukhanova, A., Grokhovsky S., Zhure, A., Roper, D., Bronstein, I. *Biochem Mol Biol Int* 44, 997-1010 (1998).
21. Kraev, A. S. *Mol Biol* 22, 1164-1197 (1998).
22. Sambrook, J., Fritsch, E. F., Maniatis, T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press (1989).
23. Nechipurenko, Y. D., Jovanović, B., Riabokon, V. F., and Gursky, G. V. *Ann NY Acad Sci* 1048, 206-214 (2005).
24. Giovannangeli, C., Diviacco, S., Labrousse, V., Gryaznov, S., Charneau, P., Hélène, C. *Proc Nat Acad Sci USA* 94, 79-84 (1997).
25. Brunet, E., Corgnali, M., Perrouault, L., Roig, V., Asseline, U., Sorensen, M. D., Babu, B. R., Wengel, J., and Giovannangeli, C. *Nucl Acids Res* 33, 4223-4234 (2005).
26. Brunet, E., Corgnali, M., Cannata, F., Perrouault, L., and Giovannangeli, C. *Nucl Acids Res* 34, 4546-4553 (2006).
27. Surovaya, A. N., Burckhardt, G., Grokhovsky, S. L., Birch-Hirschfeld, E., Gursky, G. V. Zimmer, C. *J Biomol Struct Dyn* 14, 595-606 (1997).
28. Surovaya, A. N., Burckhardt, G., Grokhovsky, S. L., Birch-Hirschfeld, E., Nikitin, A. M., Fritzsche, H., Zimmer, C., Gursky, G. V. *J Biomol Struct Dyn* 18, 689-701 (2001).
29. Gursky, G. V., Zasedatelev, A. S. *Sov Sci Rev D Physicochem Biol* 5, 53-139 (1984).
30. Parks, M. E., Dervan, P. B. *Bioorgan Med Chem* 4, 1045-1050 (1996).
31. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. Jr., Weiner P. *J Am Chem Soc* 106, 765-784 (1984).
32. Cardwell, J., Kollman, P. *Biopolymers* 25, 249-266 (1986).
33. Belikov, S. V., Grokhovsky, S. L., Isagulians, M. V., Surovaya, A. N., Gursky, G. V. *J Biomol Struct Dyn* 23, 193-202 (2005).
34. Halgren, T. A. *J Comp Chem* 17, 490-497 (1996).
35. Li, X. Y., Liu, D. R. *Angew Chem Int Ed* 43, 4848-4870 (2004).
36. Poulin-Kerstien, A. T., Dervan, P. B. *J Amer Chem Soc* 125, 15811-15821 (2003).
37. Giovannangeli, C., Perrouault, L., Escudé, C., Gryaznov, S. M., Hélène, C. *J Mol Biol* 261, 386-398 (1996).
38. Sun, B. W., Babu, B. R., Sorensen, M. D., Zakrzewska, K., Wengel, J., Sun, J. S. *Biochemistry* 43, 4160-4169 (2004).

Date Received: April 30, 2007

Communicated by the Editor Valery Ivanov