

Chapter 1

Introduction

Sequence-Specific Recognition of DNA

DNA is the hereditary molecule upon which all cellular life is based. The transcription of DNA into RNA followed by translation into proteins is one of the most fundamental processes in biology. This process is regulated by the sequence-specific binding to DNA of a class of proteins called transcription factors (1).

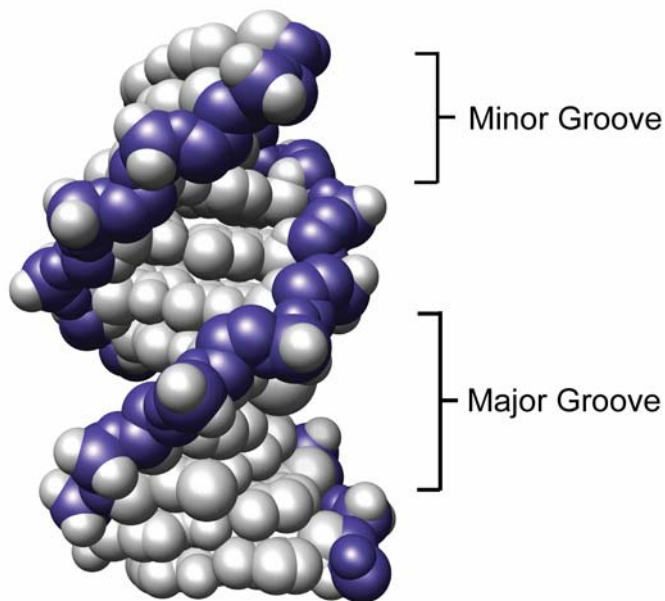


Figure 1.1. Crystal structure of a DNA double helix

In its native form, DNA consists of two antiparallel polydeoxyribonucleotide strands wrapped together in a double helix (Figure 1.1) (2,3). Between the two strands, hydrogen bonding of the four heterocyclic bases pairs adenine (A) to thymine (T) and guanine (G) to cytosine (C). The helix forms two grooves: the major groove, which is wide and shallow, and the minor groove, which is narrow and deep. The edges of the base pairs present unique molecular surfaces at the major and minor groove floors, providing for the sequence-specific recognition of DNA by transcription factors (Figure 1.2).

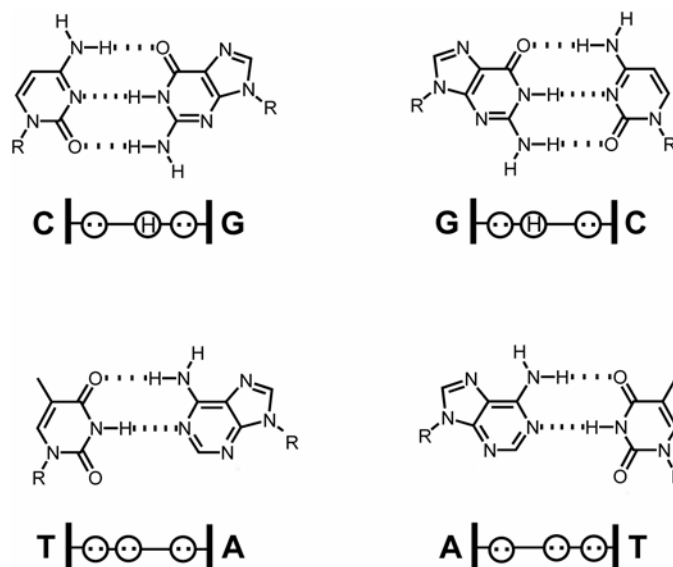


Figure 1.2. Structures of the DNA base pairs. Differences in the hydrogen bonding patterns of the base pairs in the minor groove are illustrated. Circles with two dots represent hydrogen bond acceptors. Circles containing an H represent a hydrogen bond donor. R represents the DNA sugar-phosphate backbone. For each base pair, the major groove is oriented above and the minor groove is oriented below.

Besides proteins, some naturally-occurring small molecules are also capable of binding DNA in a sequence-specific manner. Among these are distamycin A and netropsin, which are crescent-shaped natural products that preferentially bind in the DNA minor groove at A•T sequences (Figure 1.3) (4).

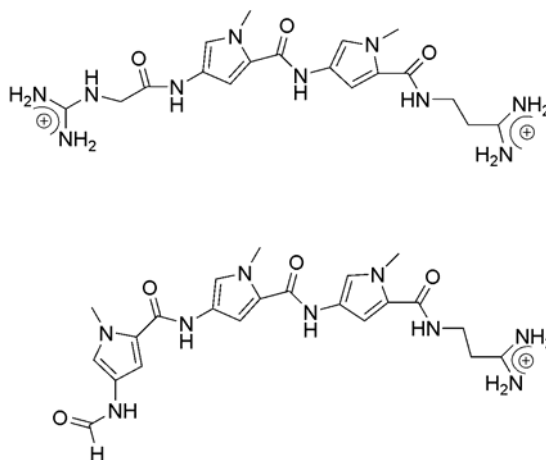


Figure 1.3. Structures of netropsin (top) and distamycin A (bottom)

Pyrrole-Imidazole Polyamides

Aberrant gene expression is the basis for many human diseases, such as cancer. The ability to specifically regulate the expression of proteins associated with disease states would clearly be beneficial. Toward this end, the development of programmable synthetic molecules capable of recognizing predetermined DNA sequences would be valuable, with potential applications in human medicine. In addition, such molecules could serve as chemical tools for understanding biological systems. These molecules would ideally have DNA binding affinities and sequence specificities rivaling those of endogenous proteins, as well as the ability to localize to the nuclei of living cells and access their DNA. Pyrrole-imidazole polyamides were developed for this purpose (5).

Polyamides were based on distamycin A due to the structural simplicity and modularity of its repeating *N*-methylpyrrole (Py) units. A significant advancement came when it was found that the replacement of a Py ring with *N*-methylimidazole (Im) allowed for the recognition of a G•C base pair by accommodating the exocyclic amine of guanine with the nitrogen of Im (6,7). Binding as side-by-side, antiparallel dimers in the minor groove, the heterocyclic rings of two polyamide strands can recognize DNA sequence-specifically by pairing rules: a Py opposite an Im (i.e., Py/Im) targets C•G, Im/Py recognizes G•C, and Py/Py is partially degenerate, targeting both A•T and T•A (Figures 1.4–1.5). Additional studies have since expanded the heterocycle repertoire (8). In particular, *N*-methyl-3-hydroxypyrrole (Hp) paired with Py is specific for T•A, and Py/Hp targets A•T (9).

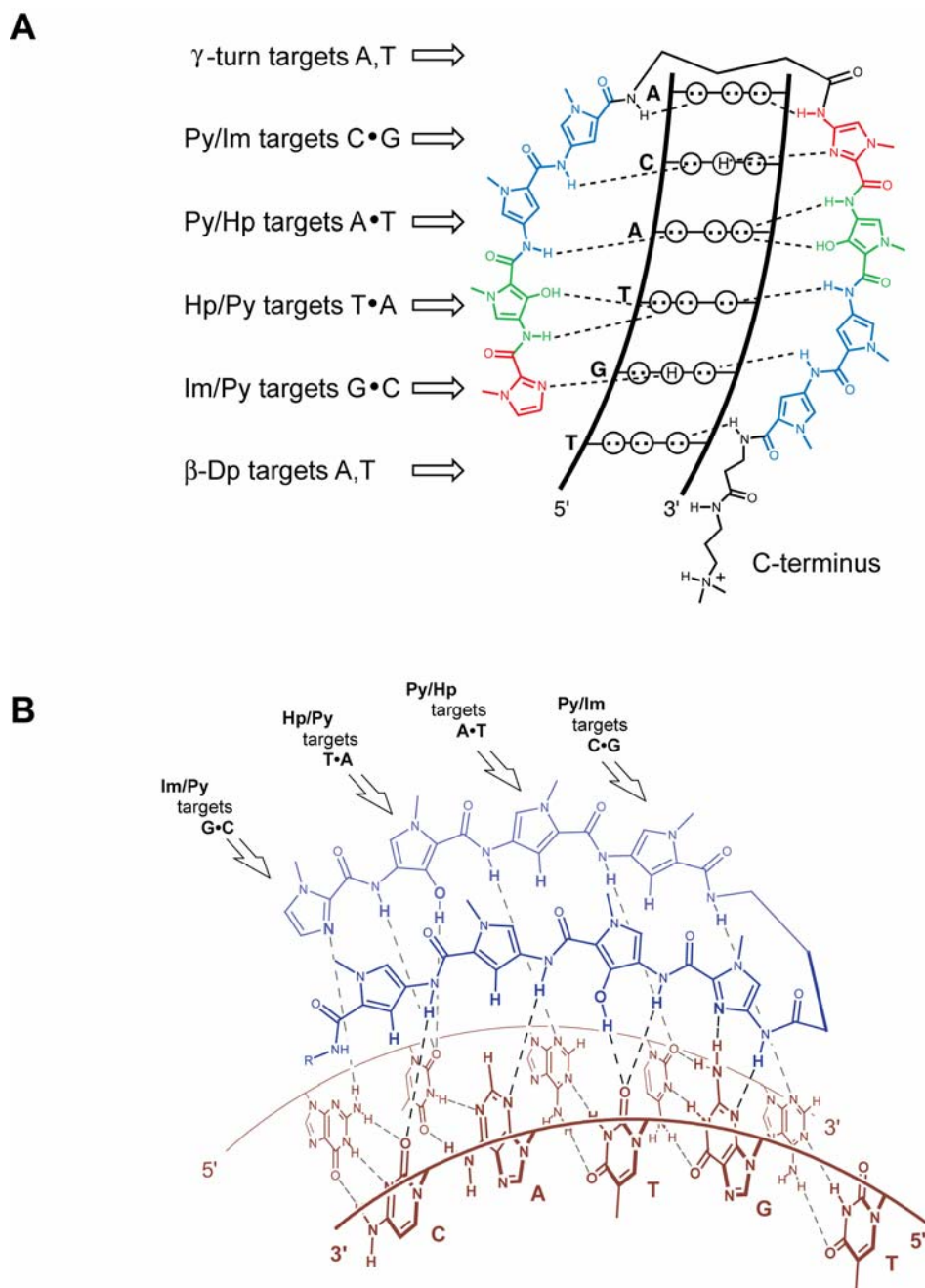


Figure 1.4. (A) Model for a hairpin polyamide binding a 5'-TGTACA-3' DNA sequence. Putative hydrogen bonds are shown as dashed lines. A circle with two dots represents a hydrogen bond acceptor. A circle with an H represents a hydrogen bond donor. (B) An alternative view of the polyamide-DNA complex illustrating hydrogen-bonding patterns and shape complementarity. The minor groove curvature has been lessened and the twist between the base pairs has been removed. The sugar-phosphate backbone is indicated by a solid curved line (8).

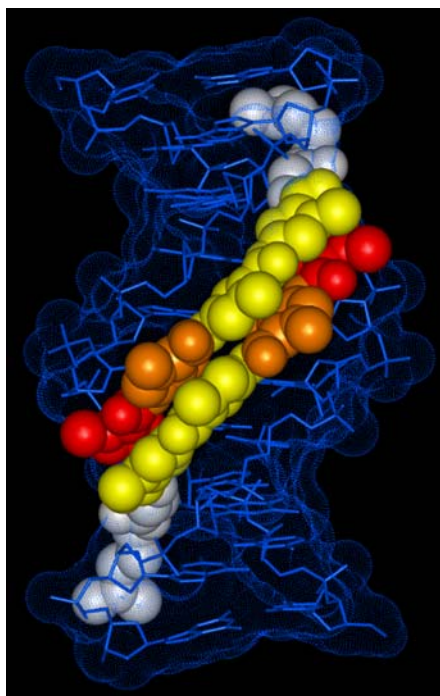


Figure 1.5. Crystal structure of a polyamide homodimer bound to DNA (10)

By covalently linking the two polyamide strands via an alkyl turn moiety, polyamides form a ‘hairpin’ structure in which side-by-side pairings are enforced, preventing slipped binding modes. Early studies exploring a variety of linker lengths for the turn unit found that use of γ -aminobutyric acid as the turn moiety was optimal and enhanced binding affinity ~ 100 -fold (11). Furthermore, linkage with the chiral (*R*)-2,4-diaminobutyric acid enhanced binding affinity an additional ~ 10 -fold (12). Both turn units display sequence specificity for A•T and T•A (i.e., W) over G•C and C•G. Utilizing these advancements, polyamides are capable of binding DNA with affinities comparable to those of natural proteins (13).

Since longer DNA sequences occur less frequently in the genome, maximizing the binding site size of polyamides is an important issue. Beyond five contiguous rings,

polyamide binding affinity begins to decrease (14). Crystal structures of DNA bound to polyamide indicate that this is due to overcurvature of the polyamide in relation to the curvature of the DNA minor groove (15). The use of β -alanine (β) residues, which can pair with themselves or Py to give specificity for W, appears to relax ligand curvature and allow the targeting of longer sites at high affinity by hairpin polyamides (16,17).

The nuclear uptake of polyamides by live cells has been directly established with visualization of polyamide-fluorophore conjugates, as well as indirectly through quantitative real-time RT-PCR of targeted genes (18–20). By targeting the promoter or enhancer regions of genes, polyamides have demonstrated the ability to bind specific DNA sequences and modulate gene transcription in a variety of systems *in vitro* and *in vivo* (5,20–24). For example, polyamides have been shown to bind the hypoxia response element (HRE), disrupting the binding of hypoxia-inducible factor (HIF) *in vitro* (20,21). In cultured cells, the polyamides inhibited expression of vascular endothelial growth factor, which is regulated by the binding of HIF to the HRE in the promoter region of its gene.

Polyamide-Alkylator Conjugates

Polyamides targeted to the coding region of genes do not appear to inhibit gene expression. Based on the hypothesis that covalent modification of DNA in gene coding regions would inhibit transcription, polyamide-alkylator conjugates have been created (5,25).

DNA alkylators generally display limited sequence selectivity. Conjugation of a polyamide to an alkylating agent creates a bifunctional molecule capable of the sequence-

specific alkylation of DNA. Several classes of alkylating agents have been conjugated to polyamides, including 1-chloromethyl-5-hydroxy-1,2-dihydro-3*H*-benz[*e*]indole (*seco*-CBI), cyclopropylpyrroloindole (CPI), and chlorambucil (Chl) (Figure 1.6). Polyamide-alkylator conjugates from each of these classes have demonstrated the ability to alkylate predetermined DNA sequences (25–27). For example, a polyamide-chlorambucil conjugate has demonstrated the ability to alkylate target sites in the HIV-1 promoter sequence *in vitro* and in live CEM T-cells containing a stably integrated copy of the HIV-1 long terminal repeat, which encompasses the HIV-1 promoter (27). Alkylator conjugates have also demonstrated the ability to induce biological responses in cellular systems, such as gene silencing and the induction of cell cycle arrest (28–30).

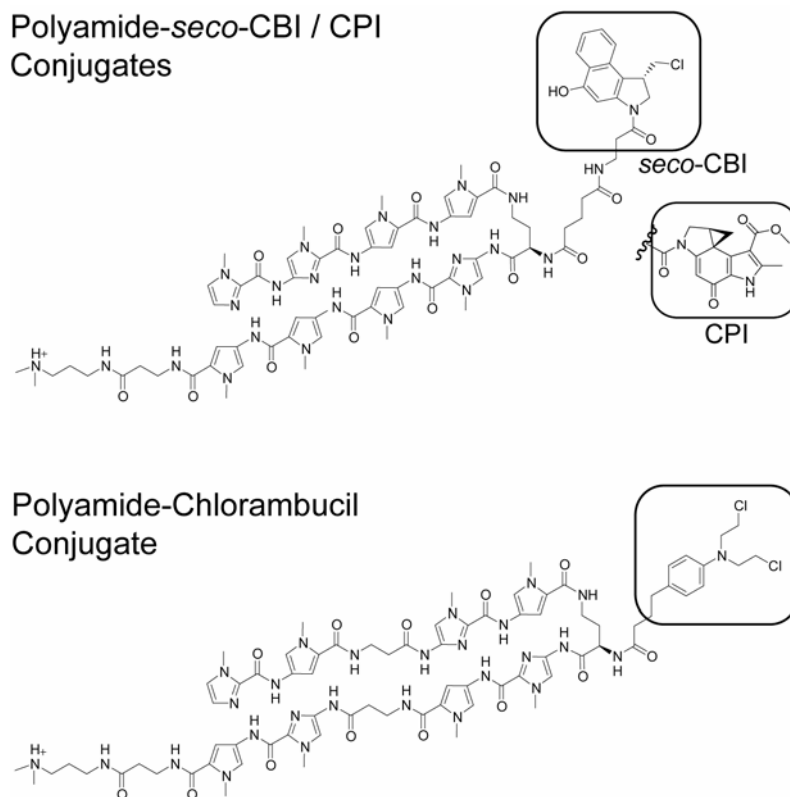


Figure 1.6. Examples of several classes of polyamide-alkylator conjugates. The alkylating moiety is boxed (5).

Scope of This Work

This thesis describes work directed at using synthetic chemistry to understand biological systems. Chapters 2 and 3 describe the identification and characterization of polyamides linked by the novel turn unit, α -diaminobutyric acid (α -DABA). Impetus for this characterization was provided by the discovery that an α -DABA-linked polyamide-chlorambucil conjugate arrested tumor growth in immunodeficient mice while the standard γ -DABA-linked analog was toxic. The DNA binding affinities and sequence specificities of a small library of α -DABA-linked polyamides were determined, and the kinetics and specificities of their chlorambucil conjugates were measured. The binding orientation of α -DABA-linked polyamides was also investigated. In addition, a variety of cellular assays were conducted with the polyamides and their conjugates.

Chapter 4 describes the development of the convergent synthesis of a class of sulfated oligosaccharides known as chondroitin sulfate (CS) glycosaminoglycans. These molecules play important roles in a variety of biological functions, but understanding of their function is limited due to their structural complexity and heterogeneity. Synthetic CS molecules of defined structures will serve as critical tools for understanding their biological roles.

References

1. Kadonaga, J.T. (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, **116**, 247–257.
2. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
3. Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.M., Fratini, A.V., and Kopka, M.L. (1982) The anatomy of A-, B-, and Z-DNA. *Science*, **216**, 475–485.
4. Arcamone, F., Penco, S., Orezzi, P., Nicoletta, V., and Pirelli, A. (1964) Structure and Synthesis of Distamycin A. *Nature*, **203**, 1064–1065.
5. Dervan, P.B., Poulin-Kerstien, A.T., Fechter, E.J., and Edelson, B.S. (2005) Regulation of gene expression by synthetic DNA-binding ligands. *Top. Curr. Chem.*, **253**, 1–31.
6. Wade, W.S., Mrksich, M., and Dervan, P.B. (1992) Design of Peptides That Bind in the Minor Groove of DNA at 5'-(a,T)G(a,T)C(a,T)-3' Sequences by a Dimeric Side-by-Side Motif. *J. Am. Chem. Soc.*, **114**, 8783–8794.
7. Mrksich, M., Wade, W.S., Dwyer, T.J., Geierstanger, B.H., Wemmer, D.E., and Dervan, P.B. (1992) Antiparallel Side-by-Side Dimeric Motif for Sequence-Specific Recognition in the Minor Groove of DNA by the Designed Peptide 1-Methylimidazole-2-Carboxamide Netropsin. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 7586–7590.
8. Dervan, P.B. and Edelson, B.S. (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.*, **13**, 284–299.
9. White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., and Dervan, P.B. (1998) Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*, **391**, 468–471.
10. Kielkopf, C.L., White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., Dervan, P.B., and Rees, D.C. (1998) A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science*, **282**, 111–115.
11. Mrksich, M., Parks, M.E., and Dervan, P.B. (1994) Hairpin Peptide Motif — A New Class Of Oligopeptides For Sequence-Specific Recognition In The Minor-Groove Of Double-Helical DNA. *J. Am. Chem. Soc.*, **116**, 7983–7988.
12. Herman, D.M., Baird, E.E., and Dervan, P.B. (1998) Stereochemical control of the DNA binding affinity, sequence specificity, and orientation preference of chiral hairpin polyamides in the minor groove. *J. Am. Chem. Soc.*, **120**, 1382–1391.

13. Hsu, C.F., Phillips, J.W., Trauger, J.W., Farkas, M.E., Belitsky, J.M., Heckel, A., Olenyuk, B., Puckett, J.W., Wang, C.C., and Dervan, P.B. (2007) Completion of a programmable DNA-binding small molecule library. *Tetrahedron*, **63**, 6146–6151.
14. Kelly, J.J., Baird, E.E., and Dervan, P.B. (1996) Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 6981–6985.
15. Kielkopf, C.L., Baird, E.E., Dervan, P.D., and Rees, D.C. (1998) Structural basis for G•C recognition in the DNA minor groove. *Nat. Struct. Biol.*, **5**, 104–109.
16. Turner, J.M., Swalley, S.E., Baird, E.E., and Dervan, P.B. (1998) Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.*, **120**, 6219–6226.
17. Wang, C.C.C., Ellervik, U., and Dervan, P.B. (2001) Expanding the recognition of the minor groove of DNA by incorporation of beta-alanine in hairpin polyamides. *Bioorg. Med. Chem.*, **9**, 653–657.
18. Best, T.P., Edelson, B.S., Nickols, N.G., and Dervan, P.B. (2003) Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 12063–12068.
19. Edelson, B.S., Best, T.P., Olenyuk, B., Nickols, N.G., Doss, R.M., Foister, S., Heckel, A., and Dervan, P.B. (2004) Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res.*, **32**, 2802–2818.
20. Nickols, N.G., Jacobs, C.S., Farkas, M.E., and Dervan, P.B. (2007) Improved nuclear localization of DNA-binding polyamides. *Nucleic Acids Res.*, **35**, 363–370.
21. Olenyuk, B.Z., Zhang, G.J., Klco, J.M., Nickols, N.G., Kaelin, W.G., and Dervan, P.B. (2004) Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 16768–16773.
22. Burnett, R., Melander, C., Puckett, J.W., Son, L.S., Wells, R.D., Dervan, P.B., and Gottesfeld, J.M. (2006) DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA center dot TTC repeats in Friedreich's ataxia. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 11497–11502.
23. Kageyama, Y., Sugiyama, H., Ayame, H., Iwai, A., Fujii, Y., Huang, L.E., Kizaka-Kondoh, S., Hiraoka, M., and Kihara, K. (2006) Suppression of VEGF transcription in renal cell carcinoma cells by pyrrole-imidazole hairpin polyamides targeting the hypoxia responsive element. *Acta Oncol.*, **45**, 317–324.

24. Nickols, N.G. and Dervan, P.B. (2007) Suppression of Androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 10418–10423.
25. Bando, T. and Sugiyama, H. (2006) Synthesis and biological properties of sequence-specific DNA-alkylating pyrrole-imidazole polyamides. *Accounts Chem. Res.*, **39**, 935–944.
26. Chang, A.Y. and Dervan, P.B. (2000) Strand selective cleavage of DNA by diastereomers of hairpin polyamide-seco-CBI conjugates. *J. Am. Chem. Soc.*, **122**, 4856–4864.
27. Wurtz, N.R. and Dervan, P.B. (2000) Sequence specific alkylation of DNA by hairpin pyrrole-imidazole polyamide conjugates. *Chem. Biol.*, **7**, 153–161.
28. Wang, Y.D., Dziegielewski, J., Chang, A.Y., Dervan, P.B., and Beerman, T.A. (2002) Cell-free and cellular activities of a DNA sequence selective hairpin polyamide-CBI conjugate. *J. Biol. Chem.*, **277**, 42431–42437.
29. Wang, Y.D., Dziegielewski, J., Wurtz, N.R., Dziegielewska, B., Dervan, P.B., and Beerman, T.A. (2003) DNA crosslinking and biological activity of a hairpin polyamide-chlorambucil conjugate. *Nucleic Acids Res.*, **31**, 1208–1215.
30. Shinohara, K., Sasaki, S., Minoshima, M., Bando, T., and Sugiyama, H. (2006) Alkylation of template strand of coding region causes effective gene silencing. *Nucleic Acids Res.*, **34**, 1189–1195.