Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 3151–3156
Bioorganic & Medicinal Chemistry Letters

Sequence specific recognition of DNA by tailor-made hairpin conjugates of achiral *seco*-cyclopropaneindoline-2-benzofurancarboxamide and pyrrole–imidazole polyamides

Carly A. Price,[a] Brian M. Lingerfelt,[a] Heather L. Handl,[a] Konstantinos Kiakos,[b] John A. Hartley[b] and Moses Lee[a,*]

[a]*Department of Chemistry, Furman University, Greenville, SC 29613, United States*
[b]*Cancer Research UK Drug–DNA Interactions Research Group, Department of Oncology, University College London, 91 Riding House Street, London W1W 7BS, UK*

Received 24 February 2005; revised 1 April 2005; accepted 5 April 2005
Available online 3 May 2005

Abstract—Hairpin conjugates of achiral *seco*-cyclopropaneindoline-2-benzofurancarboxamide (achiral *seco*-CI-Bf) and three diamides (ImPy **1**, PyIm **2**, and PyPy **3**, where Py is pyrrole, and Im is imidazole), linked by a γ-aminobutyrate group, were synthesized. The sequence-specific covalent alkylation of the achiral CI moiety with adenine-N3 in the minor groove was ascertained by thermally induced DNA cleavage experiments. The results provide evidence that hairpin conjugates of achiral *seco*-CI-Bf-γ-polyamides could be tailored to target specific DNA sequences according to a set of general rules: the achiral CI moiety selectively reacts with adenine-N3, a stacked pair of imidazole/benzofuran prefers a G/C base pair, and a pyrrole/benzofuran prefers an A/T or T/A base pair. Models for the binding of hairpin conjugates **1**–**3** with sequences 5′-TCA(888)G-3′, 5′-CAA(857)C-3′, and 5′-TTA(843)C-3′ are proposed.

Small molecules that can be designed to specifically recognize predetermined DNA sequences are important in the discovery of novel gene-based therapeutic agents,[1] DNA-based diagnostic and biosensor devices,[2] as well as tools for probing the structure and function of nucleic acids.[3] The DNA minor groove is an attractive target for the design and development of sequence-specific ligands.[4] Non-covalent minor groove binding agents include a wide range of structural types. For example, Hoechst 33258 contains benzimidazole units,[5] DAPI contains an indole moiety,[6] furamidine contains a central furan unit,[7] and polyamides, which are pyrrole and imidazole-containing analogs of the natural product distamycin A.[4,8] Examples of covalent minor groove binding agents include CC-1065 and the duocarmycins,[9] the mitomycins,[10] and the pyrrolobenzodiazepines.[11]

Of the minor groove binding agents, the pyrrole–imidazole polyamides are the most extensively studied.[4,8] It is now well established that linear polyamides, as well as their hairpin analogs, readily form 'stacked' dimers.[12] The agents are able to recognize specific DNA sequences through a combination of hydrogen bonding, van der Waals, and electrostatic forces between the ligands and groups on the floor of the minor groove.[4,8] A set of rules for their sequence specific recognition has been established.[4,8] A stacked Im/Py pair selects for a G/C base pair, a Py/Im pair recognizes a C/G base pair, and a Py/Py pair could target an A/T or T/A site.[4,8] Furthermore, the γ-aminobutyrate (or γ) linker demonstrates some selectivity for an A/T or T/A base pair,[13] and the dimethylaminopropyl group at the C-terminus selectively binds an A/T or T/A base pair.[4,8,14] Our group has recently reported that the language of DNA recognition can be expanded to include 'word' of two base pairs, instead of the existing paradigm of recognizing one 'letter' or base pair at a time.[15] We discovered that an -ImPy- central pairing, that interacts with its cognate sequence -GC-, provides significant enhancement to the binding affinity over -PyPy- (for -AT- cognate), followed

Keywords: Duocarmycins; Polyamides; Conjugates; DNA; Sequence specificity.
* Corresponding author. Tel.: +1 864 294 3368; fax: +1 864 294 3559; e-mail: moses.lee@furman.edu

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.04.006

by -PyIm- and -ImIm-, which codes for -AG- and -GG-, respectively.[15] With these rules, one could envision the possibility of designing a polyamide that could target any DNA sequence.

Even though significant progress has been made toward understanding the molecular recognition of non-covalent minor groove binding agents such as polyamides, our ability to design covalently reactive minor groove binding agents is less sophisticated. Even though numerous analogs of CC-1065 and the duocarmycins have been synthesized by others[9] and us,[16] the analogs have largely displayed similar sequence preference for AT-rich sequences as the parent molecules. There is an exception, stacked heterodimers of duocarmycin A and polyamides have been found to recognize and to covalently react with sequences different from duocarmycin alone.[17] Even though distamycin A and duocarmycin A bind tightly to AT-rich sequences, the heterodimer alkylates a GN3 position in GC-rich sequences, such as 5′-AGGT*G*-3′.[17] More recently, we have demonstrated that imidazole-containing triamides, such as compound **8**, dramatically modulated the sequence specificity of duocarmycin A, by forcing the heterodimer to alkylate guanine-N3 atoms exclusively at GC sequences.[18] These findings indicate that hairpin conjugates of polyamides and duocarmycin A may provide an attractive strategy for the discovery of a new class of minor groove and sequence specific alkylating agents. To date, several conjugates of the duocarmycin alkylating pharmacophore and its analogs with polyamides, including hairpin polyamides, have been reported.[19] However, the sequence-specific alkylation exhibited by these conjugates were directed by the imidazole/pyrrole pairings and side-by-side binding motif of the polyamides.[19] Conjugates of CC-1065 analogs with oligodeoxynucleotides have also been reported.[20] Our group has recently reported a series of racemic *seco*-cyclopropaneindoline-2-benzofurancarboxamide (CI-Bf) conjugates with four different polyamides (ImIm **4**, PyIm **5**, ImPy **6**, and PyPy **7**).[16a] The Im-containing conjugates were found to accept G residues, but they exhibited a significant 'memory' for the AAAAA site of the duocarmycins. Using a model of sequence recognition for the hairpin racemic CI-Bf-γ-polyamide, along with our interest to investigate the influence of the chiral center on DNA recognition, we have designed and synthesized three achiral analogs, **1**–**3**, whose structures are given in Figure 1.

The synthetic strategy for preparing the target hairpin molecules **1**–**3** is depicted in Scheme 1. Following published methods,[21] reduction of the nitro group of chloride **9** by catalytic hydrogenation with Adam's catalyst gave an amine, which was coupled with 5-nitrobenzofuran-2-carboxylic acid in the presence of PyBOP and DIPEA. The desired amide **10** was isolated in 22% yield. Selective reduction of the nitro moiety in compound **10** was again achieved using Adams catalyst and hydrogenation at a pressure of 55 PSI. Subsequent reaction of the amine **11** with carboxylic acids **12**–**14** in the presence of EDCI gave the protected hairpin compounds **15**–**17** in reasonable yields (51–58%). Removal of the benzyl protecting group of compounds **15**–**17** by hydrogenation over 10% palladium on carbon, followed by treatment of the products with 3 M hydrochloric acid in dry ethyl acetate gave the desired off-white precipitates of products **1**–**3** in modest yields (36–46%). All the compounds prepared in this study were characterized by 500 MHz ^{1}H NMR, FT-IR, FAB mass spectrometry, and accurate mass measurements.

The covalent sequence specificity of compounds **1**–**3** was assessed by thermally induced DNA strand cleavage, which is commonly used to probe sequence-specific covalent interactions with purine-N3 in the minor

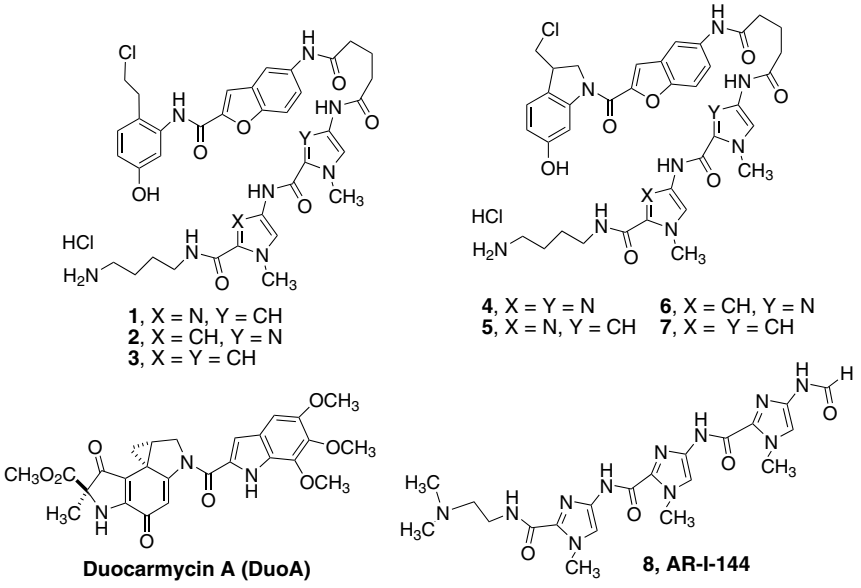
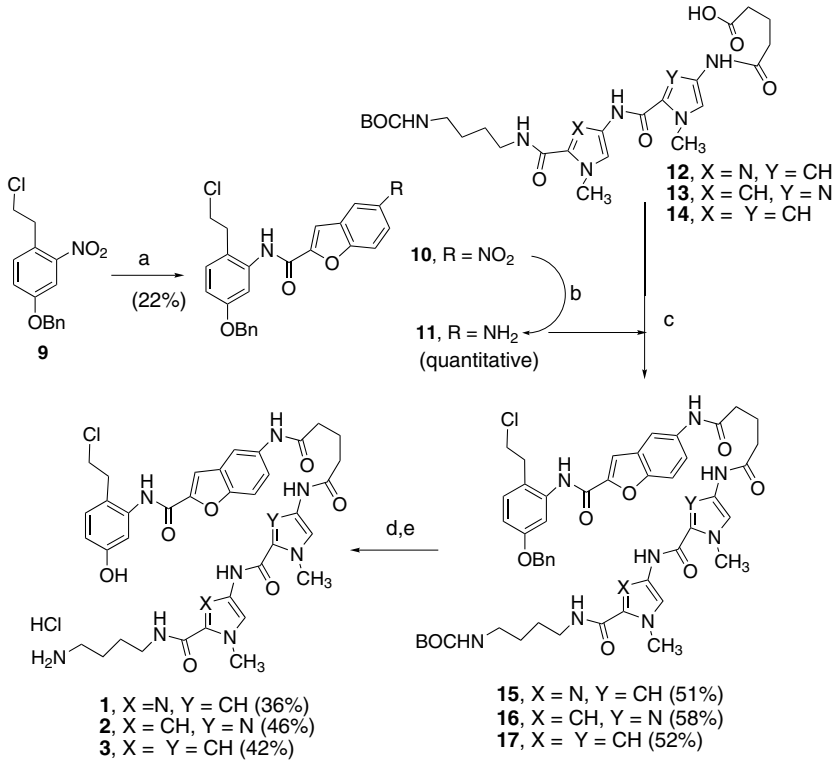


Figure 1. Structures of the target compounds achiral *seco*-CI-polyamide conjugates **1**–**3**, the previously reported racemic *seco*-CI-polyamides **4**–**7**, duocarmycin A (DuoA), and AR-1-144 (**8**).



Scheme 1. Synthesis of the target compounds **1**–**3**. Reagents and conditions: (a) (i) hydrogen, PtO_2, THF, 55 PSI, 45 min; (ii) 5-nitrobenzofuran-2-carboxylic acid, PyBOP, DIPEA, CH_2Cl_2, N_2, 2 days. (b) Hydrogen, PtO_2, THF, 55 PSI, 45 min. (c) Polyamide acid **12**–**14**, EDCI, HOBt, DMF, N_2, rt, 3 days. (d) Hydrogen, 10% Pd/C, THF, atmospheric pressure, 1 day. (e) 3 M HCl, EtOAc, ice bath, 3–4 h.

groove.[9,16a,19,22] The DNA fragment used in these studies was obtained from PCR amplification of base pairs 749–940 of the pUC18 plasmid that was linearized with *Hin*d III. A 5′-^{32}P labeled 5′-CTGTCGGGTTT-3′ primer was used as the forward primer so that each final probe copy was singly end-labeled. Results from the thermally induced DNA strand break experiment, as depicted in Figure 2, indicate that unlike the previously reported racemic hairpins,[16a] compounds **1**–**3** did not retain a memory for alkylating adenine-N3 within an AAAAA(865) sequence, a site that is preferred by CC-1065, adozelesin, and *seco*-CI-trimethoxyindole (or *seco*-CI-TMI) (data not shown). Interestingly, all three hairpin conjugates produced alkylation bands at 5′-TTT**A**(843)CGGTT-3′ (alkylation at the purine-N3 of the underlined base). However, hairpin **1** produced unique alkylation bands at 5′-CG**TCA**(888)GGG-3′, which was not recognized by hairpin **2** or **3**. Conversely, hairpin **2** reacted at 5′-G**CAA**(857)CGCG-3′, which was not recognized by hairpin **1**. Based on our previously published model for the binding of the racemic hairpin conjugates (**4**–**7**), the results suggest that hairpins **1**–**3** would reside on the bolded sequence, with alkylation at the adenine-N3 and the remaining part of the molecule wrapped toward the 5′ side of the sequence. Examination of these sequences reveals a recognition pattern related to the presence and relative position of the imidazole moiety within the polyamide portion of the molecules. Conjugate **1**, which contains an imidazole paired with a benzofuran, recognizes the sequence 5′-TC**A**-3′ indicating that the stacked Im/Bf pair behaves much like the Im/Py pair of the polyamide and selects for a G/C

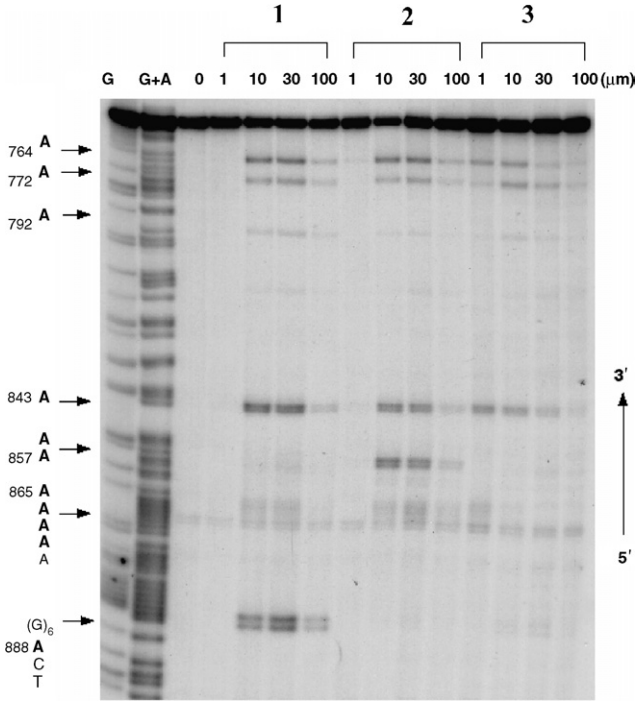


Figure 2. Thermally induced and sequence selective DNA strand break study on compounds **1**–**3**, using a pUC18 DNA fragment.

base pair. Likewise, the sequence preference of hairpin **2** for 5′-CA**A**-3′ is consistent with having its imidazole stacked with the linker and the pyrrole is stacked with the benzofuran. The Py/Bf pair, thus, behaves like the

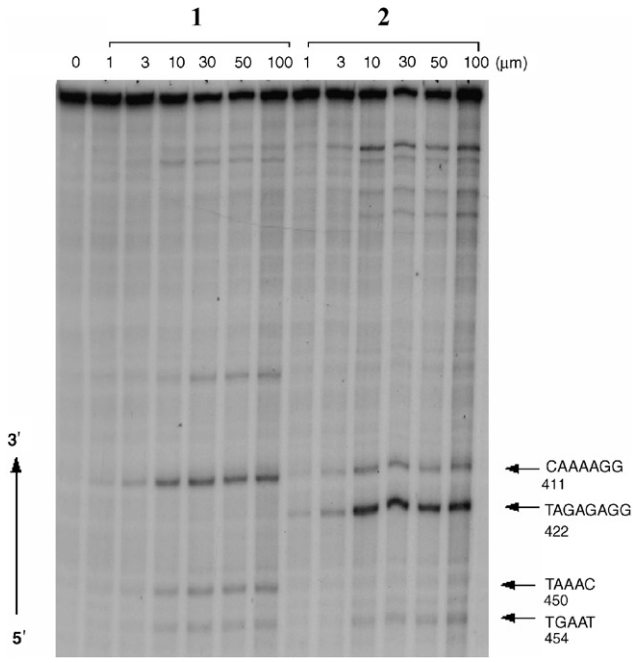


Figure 3. Thermally induced and sequence selective DNA strand break study on compounds **1**–**3**, using a topoisomerase IIα promoter fragment.

Py/Py pair of the polyamides. Conjugate **3** did not produce any alkylation at the A(857) and A(888) site presumably due to the pyrrole's lack of tolerance for G/C or C/G sites due to steric hindrance between the G-2-amino group and the pyrrole-H3 atom. Consequently, by default and like a Py/Py pair of the polyamide, hairpin **3** binds to the 5′-TT<u>A</u>(843)-3′. This sequence must be a conformationally preferred site for this class of hairpins because hairpins **1** and **2**, which do not possess structurally discriminating pyrrole-H3 atoms, were also accommodated in the A(843) site.

To confirm the sequence recognition properties of the achiral hairpins, the covalent sequence specificity of conjugates **1** and **2** was further tested on a 5′-^{32}P radio-labeled probe of 479 bp corresponding to sites −489 through −10 relative to the transcriptional start site of the human topoisomerase IIα promoter. The DNA fragment was generated using a radio-labeled primer, 5′-^{32}P labeled 5′-GTCGGTTAGGAGAGCTCCAC-TTG-3′, in a polymerase chain reaction previously described.[23]

The thermally induced DNA cleavage gel depicted in Figure 3 shows that the hairpins have unique sequence selectivity. Hairpin **1** mainly gave an alkylation band at 5′-TTC<u>A</u>(411)AAA-3′, whereas conjugate **2** gave a strong band at 5′-ACT<u>A</u>(422)GA-3′ and a weak band

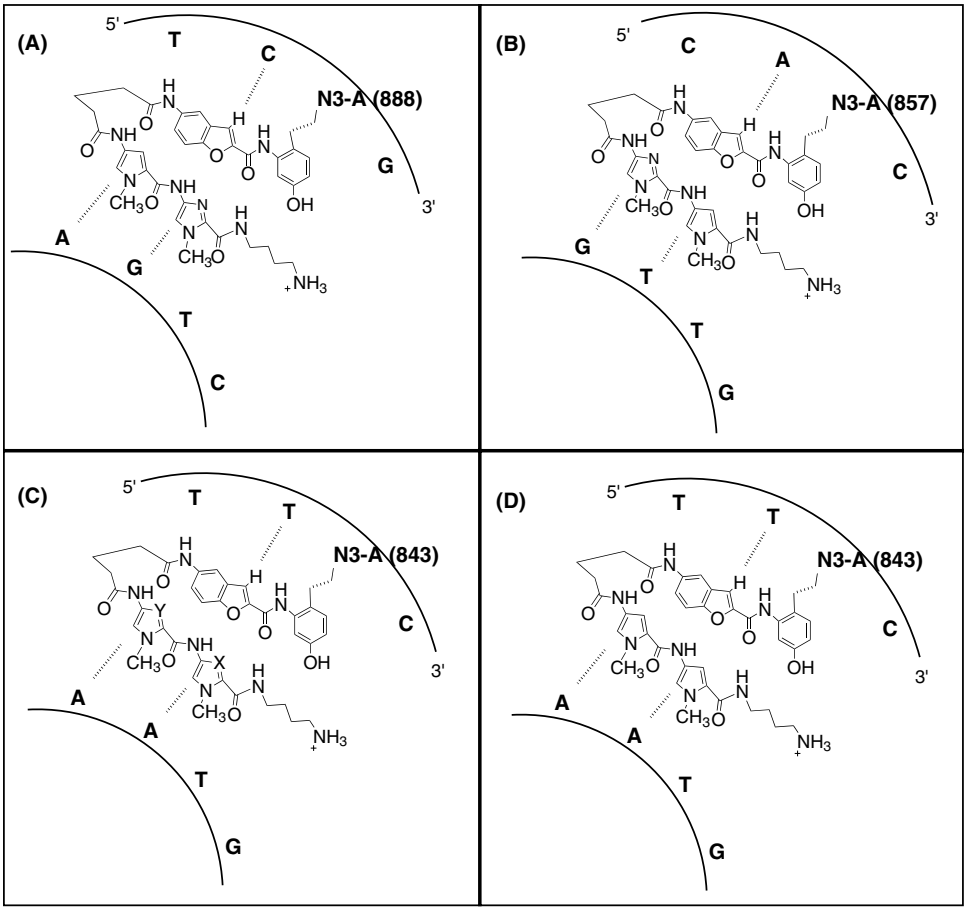


Figure 4. Model of molecular interactions between conjugate **1** with 5′-TC<u>A</u>(888)G-3′ (Part A), conjugate **2** with 5′-CA<u>A</u>(857)C-3′ (Part B), conjugates **1** and **2** with 5′-TT<u>A</u>(843)C-3′ (Part C), and conjugate **3** with 5′-TT<u>A</u>(843)C-3′ (Part D). The adenine-N3 is covalently bonded to the underlined A residue in each of the sequences in the pUC18 plasmid given in Figure 1.

was seen at A(411). The sequence preference of hairpin **1** for 5′-TCA (411)-3′ on the topoisomerase IIα fragment is consistent with the 5′-TCA(888)-3′ preference observed from the pUC18 experiment. Likewise, the preference of hairpin **2** for 5′-CTA(422)-3′ agrees with the 5′-CAA(857)-3′, in which the Py/Bf could bind either an A/T or T/A base pair.

Models that depict the molecular interactions of hairpin conjugates **1**–**3** with the 5′-TCA(888)G-3′, 5′-CAA(857)C-3′, and 5′-TTA(843)C-3′ sequences, found in the pUC18 plasmid are shown in Figure 4. In these models, the compounds are oriented toward the 5′ direction from the alkylation site. This orientation is consistent with the previously reported model of a complex of compound (S)-**4** with the 3′-GGGA(888)**CT**GCTC-5′ sequence, in which the imidazole would stack with the benzofuran to form a complex similar to that of the pyrrole/imidazole pairing that is capable of recognizing a C/G base pair.[16a] It is not likely that compounds **1**–**3** would bind to the three sequences in an extended conformation, because compounds **1** and **2** would require the benzofuran group to bind unfavorably to a C/G base pair. We have found that like *seco*-CI-TMI, the achiral *seco*-CI-Bf also bound to the A(865) cluster and not to GC-containing sites (data not shown).[16a] Further support of the hairpin conformation of compounds **1** and **2** came from a report that an aliphatic dicarboxamide of polyamides, such as a bis-linked netropsin with pimelic acid, was capable of forming a hairpin structure.[24]

In summary, the studies described in this communication show that conjugates of duocarmycins and polyamides that can fold into a hairpin conformation are capable of recognizing specific sequences of DNA. More importantly, removal of the chiral center in the duocarmycin-polyamide conjugates released the compounds for a strong memory for the AAAAA(865) sequence, thereby permitting the achiral hairpin molecules for further structural refinement in order to attain DNA sequence specificity. Current studies on compounds **1**–**3** are focused on ascertaining the nature by which they covalently react with DNA, including isolation and characterization of purine-N3 adducts.[9,19,21,22,25]

Acknowledgements

Financial support from Research Corporation, NSF-REU (to M.L.), and Cancer Research UK (to J.A.H.) is gratefully acknowledged. The authors also thank Dr. Robert Kelly at Pharmacia & Upjohn for a generous gift of CC-1065.

References and notes

1. Uil, T. G.; Haisma, H. J.; Rots, M. G. *Nucleic Acids Res.* **2003**, *31*, 6064.
2. (a) Janata, J.; Josowicz, M.; Vanysek, P.; DeVaney, D. M. *Anal. Chem.* **1998**, *70*, 179R; (b) Colella, G.; Marchini, S.; D'Incalci, M.; Broggini, M. *Br. J. Cancer* **1999**, *80*, 338; (c) Wang, D. G.; Lander, E. S. *Science* **1998**, *280*, 1077; (d) Evans, W. E.; Relling, M. V. *Science* **1999**, *286*, 487; (e) Bunk, S. *The Scientist* **2002**, *16*, 13.
3. (a) Neidle, S. *DNA Structure and Recognition*; IRL: Oxford, 1994; (b) *Molecular Aspects of Anticancer Drug–DNA Interaction*; Neidle, S., Waring, M., Eds.; CRC: Boca Raton, 1993; Vol. 1; (c) *Molecular Aspects of Anticancer Drug–DNA Interaction*; Neidle, S., Waring, M., Eds.; CRC: Boca Raton, 1994; Vol. 2; (d) *Nucleic Acid Targeted Drug Design*; Propst, C. L., Perun, T. J., Eds.; Marcel Dekker: New York, 1992; (e) *Advances in DNA Sequence-Specific Agents*; Hurley, L. H., Ed.; JAI: Greenwich, 1992; Vol. 1; (f) *Advances in DNA Sequence-Specific Agents*; Jones, G. B., Palumbo, M., Eds.; JAI: Greenwich, 1998; Vol. 3; (g) Neidle, S.; Thurston, D. E. In *New Molecular Targets for Cancer Chemotherapy*; Kerr, D. J., Workman, P., Eds.; CRC: Boca Raton, 1994, pp 159–175.
4. (a) Dervan, P. B. *Science* **1986**, *232*, 464; (b) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284; (c) Krowicki, K.; Lee, M.; Hartley, J. A.; Ward, B.; Kissinger, K.; Skorobogaty, A.; Dabrowiak, J. C.; Lown, J. W. *Struct. Express.* **1988**, *2*, 251; (d) Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E. *Structure* **1997**, *5*, 1033; (e) *DNA and RNA binders*; Demeunynck, M., Bailly, C., Wilson, W. D., Eds.; Wiley-VCH: Germany, 2003.
5. James, P. L.; Le Strat, L.; Ellervik, U.; Bratwall, C.; Norden, B.; Brown, T.; Fox, K. R. *Biophys. Chem.* **2004**, *111*, 205.
6. Spackova, N.; Cheatham, T. E., III; Ryjacek, F.; Lankas, F.; Van Meervelt, L.; Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2003**, *125*, 1759.
7. Nguyen, B.; Hamelberg, D.; Bailly, C.; Colson, P.; Stanek, J.; Brun, R.; Neidle, S.; Wilson, W. D. *Biophys. J.* **2004**, *86*, 1028.
8. (a) Dervan, P. B.; Burlii, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688; (b) Wemmer, D. E.; Dervan, P. B. *Curr. Opin. Struct. Biol.* **1997**, *7*, 355; (c) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215.
9. (a) Boger, D. L.; Johnson, D. S. *Angew. Chem., Int. Ed.* **1996**, *35*, 1438; (b) Boger, D. L.; Johnson, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3642; (c) Boger, D. L.; Boyce, C. W.; Garbaccio, R. M.; Goldberg, J. A. *Chem. Rev.* **1997**, *97*, 787.
10. Tomasz, M. *Chem. Biol.* **1995**, *9*, 575.
11. (a) Gregson, S. J.; Howard, P. W.; Gullick, D. R.; Hamaguchi, A.; Corcoran, K. E.; Brooks, N. A.; Hartley, J. A.; Jenkins, T. C.; Patel, S.; Guille, M. J.; Thurston, D. E. *J. Med. Chem.* **2004**, *47*, 1161; (b) Cooper, N.; Hagan, D. R.; Tiberghien, A.; Ademefun, T.; Matthews, C. S.; Howard, P. W.; Thurston, D. E. *Chem. Commun.* **2002**, *21*, 1764.
12. (a) Wemmer, D. E. *Biopolymers* **2001**, *52*, 197; (b) Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 439.
13. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147.
14. Lee, M.; Rhodes, A. L.; Wyatt, M. D.; Forrow, S.; Hartley, J. A. *Biochemistry* **1993**, *32*, 4237.
15. Buchmueller, K. L.; Staples, A. M.; Howard, C. M.; Horick, S. M.; Uthe, P. B.; Le, N. M.; Cox, K. K.; Nguyen, B.; Pacheco, K. A. O.; Wilson, W. D.; Lee, M. *J. Am. Chem. Soc.* **2005**, *127*, 742.
16. (a) Toth, J. L.; Price, C. A.; Madsen, E. C.; Handl, H. L.; Hudson, S. J.; Hubbard, R. B., III; Bowen, P. J.; Kiakos, K.; Hartley, J. A.; Lee, M. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2245; (b) Howard, T. T.; Lingerfelt, B. M.; Purnell, B.

L.; Scott, A. E.; Price, C. A.; Townes, H. M.; McNulty, L.; Handl, H. L.; Summerville, K.; Hudson, S. J.; Bowen, P. J.; Kiakos, K.; Hartley, J. A.; Lee, M. *Bioorg. Med. Chem.* **2002**, *10*, 2941.
17. Sugiyama, H.; Lian, C.; Isomura, M.; Saito, S.; Wang, A. H.-J. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14405.
18. Fujiwara, T.; Tao, Z.-H.; Ozeki, Y.; Saito, I.; Wang, A. H.-J.; Lee, M.; Sugiyama, H. *J. Am. Chem. Soc.* **1999**, *121*, 7706.
19. (a) Chang, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 4856; (b) Tao, Z.-F.; Fujiwara, T.; Saito, I. *J. Am. Chem. Soc.* **1999**, *121*, 4961; (c) Tao, Z.-F.; Saito, I.; Sugiyama, H. *J. Am. Chem. Soc.* **2000**, *122*, 1602; (d) Tao, Z.-F.; Fujiwara, T.; Saito, I.; Sugiyama, H. *Angew. Chem., Int. Ed.* **1999**, *38*, 650.
20. Kutyavin, I. V.; Afonina, I. A.; Mills, A.; Gorn, V. V.; Lukhtanov, E. A.; Belousov, E. S.; Singer, M. J.; Walburge, D. K.; Lokhov, S. G.; Gall, A. A.; Dempcy, R.; Reed, M. W.; Meyer, R. B.; Hedgpeth, J. *Nucl. Acids Res.* **2000**, *28*, 655.
21. Kupchinsky, S.; Centioni, S.; Howard, T.; Trzupek, J.; Roller, S.; Carnahan, V.; Townes, H.; Purnell, B.; Price, C.; Handl, H.; Summerville, K.; Johnson, K.; Toth, J.; Hudson, S.; Kiakos, K.; Hartley, J. A.; Lee, M. *Bioorg. Med. Chem.* **2004**, *12*, 6221.
22. Brooks, N.; Hartley, J. A.; Simpson, J. E., Jr.; Wright, S. R.; Woo, S.; Centioni, S.; Fontaine, M. D.; McIntyre, T. E.; Lee, M. *Bioorg. Med. Chem.* **1997**, *5*, 1497.
23. Tolner, B.; Hartley, J. A.; Hochhauser, D. *Mol. Pharmacol.* **2001**, *59*, 699.
24. Surovaya, A. N.; Burckhardt, G.; Birch-Hirschfeld, E. B.; Nikitin, A. M.; Frizsche, H.; Zimmer, C.; Gursky, G. V. *J. Biomol. Struct. Dyn.* **2001**, *18*, 689.
25. Barry, C. G.; Day, C. S.; Bierbach, U. *J. Am. Chem. Soc.* **2005**, *127*, 1160.