The Chemical Evolution of DNA−DNA Interstrand Cross-Linkers That Recognize Defined Mixed AT and GC Sequences

Hyun-Ju Park,[†] Robert C. Kelly,[‡] and Laurence H. Hurley*[,†]

Contribution from the Drug Dynamics Institute, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, and Pharmacia Upjohn Co., 301 Henrietta Street, Kalamazoo, Michigan 49001

Received May 17, 1996[⊗]

Abstract: The monoalkylation and cross-linking reactivities of a group of four structurally related DNA−DNA interstrand cross-linkers have been determined on restriction enzyme fragments and select oligomers. These highly potent cytotoxic DNA−DNA cross-linkers consist of two cyclopropa[*c*]pyrrolo[3,4-3]indol-4(*5H*)-ones indoles [(+)-CPI-I] joined by a urea (Bizelesin) or a bisamido furan, bisamido pyrrole, or bisamido N-methylpyrrole linker. Using a thermal cleavage assay in combination with radio-labeled restriction enzyme fragments, we have shown that these compounds cross-link duplex DNA six or seven base pairs apart on opposite strands, but they differ among themselves for both alkylation reactivity and DNA sequence selectivity. Bizelesin and the $[(+)\text{-CPI-I}]_2$ bisamido furan and $[(+)\text{-CPI-I}]_2$ bisamido N-methyl pyrrole compounds prefer purely AT-rich sequences (e.g., 5′-$\underline{T}(A/T)_{4\text{ or }5}$A*-3′, where $\underline{T}$ represents the cross-strand adenine alkylation and A* represents an adenine alkylation), while the $[(+)\text{-CPI-I}]_2$ bisamido pyrrole requires a centrally positioned GC base pair for high cross-linking reactivity (i.e., 5′-$\underline{T}(A/T)_2G(A/T)_2$A*-3′). By comparison of the cross-linking reactivity of the four compounds in 21-mer duplex oligomers containing strategically placed GC or IC base pairs, the sequence and linker requirements for high reactivity of the six- and seven-base-pair cross-linkers in 5′-$\underline{T}(N)_{4\text{ or }5}$A*-3′ sequences were determined. In the duplex, to attain highest reactivity, a centrally placed GC base pair and the exocyclic 2-amino group were required, while for the linker in the bisamido pyrrole compound, an unsubstituted amine in the pyrrole ring was necessary. On the basis of the known requirements for monoalkylation of duplex DNA by (+)-CPI-derived compounds and the structural consequences of monoalkylation, together with the information gleaned from this study, we are able to provide a rationale for the structural requirements for the specific sequence cross-linked with high reactivity by the pyrrole compound. We propose that, because monoalkylation of the duplex produced a bent DNA duplex that is unsuitable for cross-linking, the duplex has to first undergo a ligand-induced rearrangement involving two hydrogen-bonding donor−acceptor pairs, which reinstates the requirements necessary for the second alkylation reaction.

Introduction

The problem of recognition of DNA sequences by small molecular weight ligands has received considerable attention during the last few years.[1] Historically, the minor groove binders derived from netropsin and distamycin were the first group of molecules to be studied in detail, and much effort has gone into attempts to both understand the molecular basis for their sequence recognition[2] and convert ligands with strict AT specificity to GC or mixed AT−GC specificities.[2c,3] A solution to this problem seemed elusive until a new structural paradigm involving side-by-side dimers was recently discovered.[4] In parallel efforts, the DNA sequence specificity of covalent modification by the cyclopropa[*c*]pyrrolo[3,4-3]indol-4(*5H*)-ones (CPIs),[5] pyrrolo(1,4)benzodiazepines [P(1,4)Bs],[6] mitomycins,[1d,7] and pluramycins[8] has been investigated, and, in each case, considerable insight into the molecular basis for their sequence recognition determined. For the monoalkylation minor groove covalent bonding compounds, DNA−DNA intrahelical cross-linkers have been subsequently designed and synthesized. For the P(1,4)Bs,[9] the sequence specificity of cross-linkers appears to be an additive phenomenon based upon the established sequence specificity of the monoalkylation compounds.[6] However, for Bizelesin, the CPI cross-linker, although the sequence specificity remains largely AT,[10] there are conformational and structural differences in the products of cross-linking that are unique to the cross-linkers,[11] and therefore this is not an additive direct readout phenomenon. In this

* To whom correspondence should be addressed

† The University of Texas at Austin. Tel: (512) 471-4841/Fax: (512) 471-2746. E-mail: dg-dna@mail.utexas.edu.

‡ Pharmacia Upjohn Co.

⊗ Abstract published in *Advance ACS Abstracts,* September 15, 1996.

(1) (a) Dervan, P. B. *Science* **1986**, *232*, 464. (b) *Advances in DNA Sequence Specific Agents*; Hurley, L. H., Ed.; JAI Press Inc.: Greenwich, CT, 1992; Vol. 1. (c) Stubbe, J.; Kozarich, J. W. *Chem. Rev.* **1987**, *87*, 1107. (d) Tomasz, M. *Chem. Biol.* **1995**, *2*, 575. (e) Hurley, L. H. *Trends Pharmacol. Sci.* **1988**, *9*, 402. (f) Hurley, L. H. *J. Med. Chem.* **1989**, *32*, 2027. (g) *Molecular Basis of Specificity in Nucleic Acids−Drug Interactions*; Pullman, B.; Jortner, J., Eds.; Kluwer Academic Publishers, Netherlands, 1990. (h) Hopkins, P. B.; Millard, J. T.; Woo, J.; Werdner, M. F.; Kirchner, J. J.; Sigurdsson, S. T.; Raucher, S. *Tetrahedron* **1991**, *47*, 2475. (i) *Molecular Aspects of Anticancer Drug−DNA Interactions*; Neidle, S.; Waring, M., Eds.; Macmillan, London, 1994, Vol. 2. (j) Mountzouris, J. A.; Hurley, L. H. In *Bioorganic Chemistry: Nucleic Acids*; Hecht, S. M., Ed.; Oxford University Press Inc.: New York, 1996; pp 288−323.

(2) (a) Wartell, R. M.; Larson, J. E.; Wells, R. E. *J. Biol. Chem.* **1974**, *249*, 6719. (b) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5470. (c) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376. (d) Zimmer, Ch.; Wahnert, V. *Prog. Biophys. Mol. Biol.* **1986**, *47*, 31. (e) Kahne, D. *Chem. Biol.* **1995**, *2*, 7, and references therein.

(3) (a) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408. (b) Lown, J. W. *J. Mol. Recog.* **1994**, *7*, 79. (c) Lown, J. W. *Drug Dev. Res.* **1995**, *34*, 145.

(4) (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646. (c) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Nature Struct. Biol.* **1996**, *3*, 321, and references therein.

S0002-7863(96)01678-2 CCC: $12.00

contribution, we have studied a second generation of DNA−DNA interstrand cross-linkers, also based upon the monoalkylating CPIs, that recognize mixed AT−GC oligomers and provided the first example of a (+)-CPI-derived drug that has a strict requirement for a partially mixed AT−GC specificity.

The DNA cross-linking prodrug dimers U-78778 and U-78779 (Pharmacia Upjohn Co.) each consist of two open-ring homologs of the (+)-CC-1065 CPI subunit connected by either a bis-indole-amido furan or pyrrole linker (see Figure 1). For ease of reference, we will refer to these DNA−DNA cross-linkers as the furan and pyrrole compounds, respectively. Structurally and mechanistically, these compounds are expected to be similar to Bizelesin, a demonstrated DNA−DNA interstrand cross-linker (Figure 1).[12] The original design of these compounds as DNA−DNA interstrand cross-linkers is based upon the structure and reactivity of (+)-CC-1065, which is known to alkylate N3 of adenine in a sequence specific manner.[5c] In the (+)-CPI series, two drugs are in clinical trials, the monoalkylation drugs Adozelesin and Carzelesin,[13] and the cross-linker Bizelesin is in preclinical development. Bizelesin has increased potency and efficiency over Adozelesin,[14] and the furan and pyrrole compounds are even more cytotoxic than Bizelesin.[15] The synthesis and biological properties of these second-generation DNA−DNA interstrand cross-linkers are the subject of a separate publication.[15]

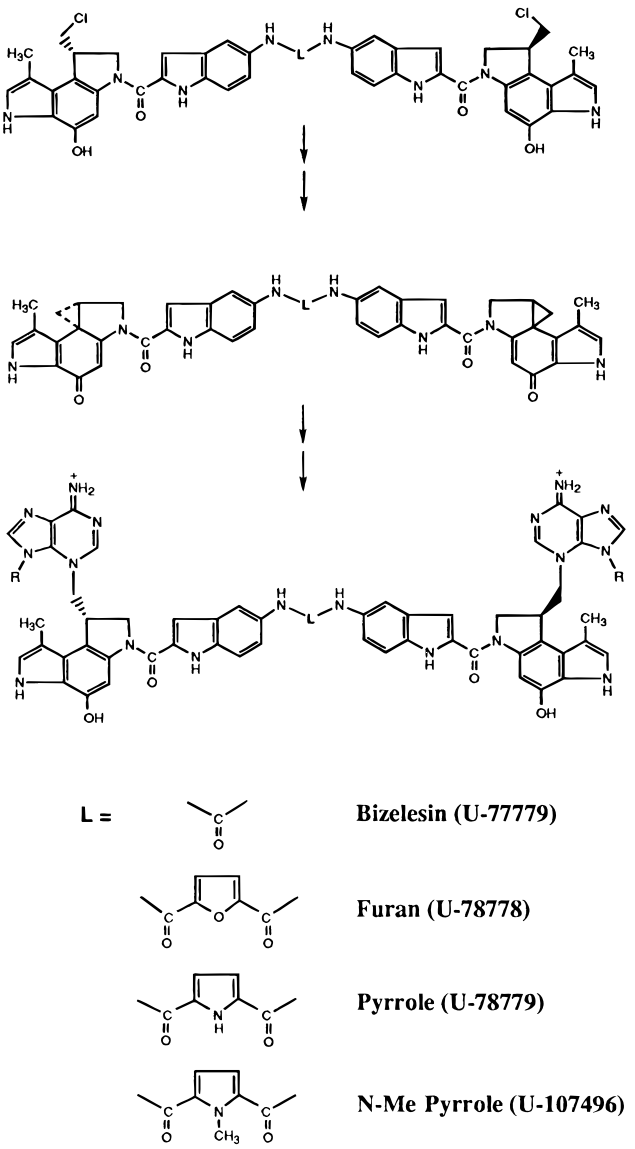


Figure 1. Structures of Bizelesin and its analogs, the furan (U-78778), pyrrole (U-78779), and N-Me pyrrole (U-107496) compounds and reaction of Bizelesin and its analogs to form the DNA-reactive cyclopropyl derivative and their subsequent reaction with adenines of opposite DNA strands to give the cross-linked adducts.

In our previous studies on the sequence specificity of Bizelesin, we have found that there is an over representation of cross-links in proportion to the number of monoadducts.[10a] In addition, although some of the cross-linked sequences, such as 5′-TAATTA* and 5′-TAAAAA* (_ and * show the opposite-strand and same-strand alkylation sites, respectively), were predicted on the basis of monoalkylation sequence specificity [e.g., 5′-TTA* and 5′-$(A)_nA$* tracts],[10,12] upon examination of their solution structures by NMR, it was discovered that they represented unexpected conformations. For example, the central AT step of the 5′-TAATTA* sequence was Hoogsteen base-paired[11] and the TAAAAA* sequence was a straight rather than

(5) (a) Hurley, L. H.; Reynolds, V. L.; Swenson, D. H.; Petzold, G. L; Scahill, T. A. *Science* **1984**, *226*, 843. (b) Warpehoski, M. A.; Hurley, L. H. *Chem. Res. Toxicol.* **1988**, *1*, 315. (c) Hurley, L. H.; Lee, C.-S.; McGovren, J. P.; Warpehoski, M. A.; Mitchell, M. A.; Kelly, R. C.; Aristoff, P. A. *Biochemistry* **1988**, *27*, 3886. (d) Hurley, L. H.; Warpehoski, M. A.; Lee, C.-S.; McGovren, J. P.; Scahill, T. A.; Kelly, R. C.; Mitchell, M. A.; Wicnienski, N. A.; Gebhard, I.; Johnson, P. D.; Bradford, V. S. *J. Am. Chem. Soc.* **1990**, *112*, 4633. (e) Warpehoski, M. In *Advances in DNA Sequence Specific Agents*; Hurley, L. H., Ed.; JAI Press Inc.: Greenwich, CT, 1992; Vol. 1, pp 217−245. (f) Hurley, L. H.; Draves, P. H. In *Molecular Aspects of Anticancer Drug−DNA Interactions*; Neidle, S., Waring, M. J., Eds.; Macmillan, London, 1993; Vol. 2; pp 89−133. (g) Boger, D. L; Ishizaki, T.; Zarrinmayeh, H.; Munk, S. A.; Kitos, P. A.; Suntornwat, O. *J. Am. Chem. Soc.* **1990**, *112*, 8961. (h) Boger, D. L.; Munk, S. A.; Zarrinmayeh, H.; Ishizaki, T.; Haught, J.; Bina, M. *Tetrahedron* **1991**, *47*, 2661. (I) Boger, D. L.; Johnson, D. S.; Palanki, M. S. S.; Kitos, P. A.; Chang, J.; Dowell, P. *Bioorg. Med. Chem.* **1993**, *1*, 27. (j) Boger, D. L.; Johnson, D. S.; Yun, W.; Tarby, C. M. *Bioorg. Med. Chem.* **1994**, *2*, 115. (k) Boger, D. L.; Johnson, D. S.; Yun, W. *J. Am. Chem. Soc.* **1994**, *116*, 1635. (l) Boger, D. L.; Johnson, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3642. (m) Boger, D. L.; Mésini, P. *J. Am. Chem. Soc.* **1995**, *117*, 11647.

(6) (a) Hertzberg, R. P.; Hecht, S. M.; Reynolds, V. L.; Molineux, I. J.; Hurley, L. H. *Biochemistry* **1986**, *25*, 1249. (b) Tang, M. S.; Nazimiec, M. E.; Doisy, R. P.; Pierce, J. R.; Hurley, L. H.; Alderete, B. E. *J. Mol. Biol.* **1991**, *220*, 855. (c) Mountzouris, J. A.; Hurley, L. H. In *Advances in DNA Sequence Specific Agents*; Hurley, L. H., Ed.; JAI Press Inc.: Greenwich, CT, 1992; Vol. 1, pp 263−292. (d) Kizu, R.; Draves, P. H.; Hurley, L. H. *Biochemistry* **1993**, *32*, 8712.

(7) (a) Borowy-Borowski, H.; Lipman, R.; Tomasz, M. *Biochemistry* **1990**, *29*, 2999. (b) Kumar, S.; Lipman, R.; Tomasz, M. *Biochemistry* **1992**, *31*, 1399. (c) Kohn, H.; Li, V.-S.; Tang, M.-s. *J. Am. Chem. Soc.* **1992**, *114*, 5501. (d) Kohn, H.; Li, V.-S.; Schiltz, P.; Tang, M.-s. *J. Am. Chem. Soc.* **1992**, *114*, 9218. (e) Li, V.-S.; Choi, D.; Tang, M.-s.; Kohn, H. *Biochemistry* **1995**, *34*, 7120. (f) Johnson, W. S.; He, Q. Y.; Tomasz, M. *Bioorg. Med. Chem.* **1995**, *3*, 851. (g) Li, V.-S.; Choi, D.; Tang, M.-s.; Kohn, H. *J. Am. Chem. Soc.* **1996**, *118*, 3765.

(8) (a) Hansen, M.; Hurley, L. H. *J. Am. Chem. Soc.* **1995**, *117*, 2421. (b) Sun, D.; Hansen, M.; Hurley, L. H. *J. Am. Chem. Soc.* **1995**, *117*, 2431. (c) Hansen, M.; Hurley, L. H. *Acc. Chem. Res.* **1996**, *29*, 249. (d) Prakash, A. S.; Moore, A. G.; Murray, V.; Matias, C.; McFayden, W. D.; Wickham, G. *Chem.−Biol. Interact.* **1995**, *95*, 17−28.

(9) (a) Jenkins, T. C.; Hurley, L. H.; Neidle, S.; Thurston, D. E. *J. Med. Chem.* **1994**, *37*, 4529. (b) Mountzouris, J. A.; Wang, J.-J.; Thurston, D.; Hurley, L. H. *J. Med. Chem.* **1994**, *37*, 3132.

(10) (a) Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **1993**, *115*, 5925. (b) Lee, C.-S.; Pfeifer, G. P.; Gibson, N. W. *Biochemistry* **1994**, *33*, 6024.

(11) (a) Seaman, F. C.; Hurley, L. H. *Biochemistry* **1993**, *32*, 12577. (b) Seaman, F. C.; Hurley, L. H. *J. Am. Chem. Soc.* **1996**, *118*, 10052−10064.

(12) Mitchell, M. A.; Kelly, R. C.; Wicnienski, N. A.; Hatzenbuhler, N. T.; Williams, M. G.; Petzold, G. C.; Slighton, J. L.; Siemieniak, P. R. *J. Am. Chem. Soc.* **1991**, *113*, 8994.

(13) (a) Li, L. H.; Kelly, R. C.; Warpehoski, M. A.; McGovren, J. P.; Gebhard, I.; DeKoning, T. F. *Invest. New Drugs* **1991**, *9*, 137. (b) Bhuyan, B. K.; Smith, K. S.; Adams, E. G.; Petzold, G. L.; McGovren, J. P. *Cancer Res.* **1992**, *52*, 5687. (c) Li, L. H.; DeKoning, T. F.; Kelly, R. C.; Krueger, W. C.; McGovren, J. P.; Padbury, G. E.; Petzold, G. L.; Wallace, T. L.; Ouding, R. J.; Prairie, M. D.; Gebhard, I. *Cancer Res.* **1992**, *52*, 4904.

(14) For example, versus i.p. implanted murine L1210 leukemia, a single i.v. injection of Bizelesin at 10 μg/kg cured (30-day survivors) 50% of the mice. For comparison, Adozelesin increased the life span of such mice by 94% at an optimum dose of 100 μg/kg but with no cures.

(15) Kelly, R. C., et al., in preparation

the expected bent DNA structure.[16] The central ureadiyl linker region of Bizelesin was found to play a decisive role in the rearrangement of the central AT step to form the Hoogsteen base-pair region[11b] and in cross-linking a seven- rather than a six-base-pair sequence containing a central GC base pair.[17] Because of these unexpected properties of six- and seven-base-pair cross-linked sequences, where the central urea linker plays a decisive role in the recognition of the cross-linked sequence, and because of the potential chemotherapeutic properties of (+)-CPI cross-linkers, further extensions of the linker region having extended minor groove cross-linking reach and more potential donor and acceptor groups were designed and synthesized.[15]

In this paper we make a comparison of the DNA reactivity and sequence specificity of these compounds with Bizelesin. The pyrrole compound has unique cross-linking reactivity with certain sequences not alkylated by the furan compound or Bizelesin, alongside an overall higher reactivity than either of the other drugs. A comparison of the consensus sequence analysis of Bizelesin and the pyrrole and furan compounds for duplex sequences reveals a preference of the pyrrole compound for specific GC-containing oligomers. Although Bizelesin reacts preferably with *bent* DNA sequences to form *straight* DNA adducts, the pyrrole compound reacts preferably with *straight* DNA and produces a *distorted* DNA adduct. Both Bizelesin and the new cross-linkers reported here face the same problem of having to convert a bent and distorted monoalkylation species to a substrate suitable for alkylation before cross-linking can occur. The solutions, however, are different. For the pyrrole compound, which requires a centrally positioned GC base pair, we demonstrate the importance of two hydrogen-bonding donor and acceptor pairs in the interaction between the linker regions of the pyrrole and the central GC base pair of the seven-base-pair cross-linked sequence in rearranging the bent monoalkylated intermediates so that they become suitable for cross-linking.

Results

In initial experiments, restriction enzyme fragments were used to determine the relative reactivities and sequence specificities of the pyrrole and furan compounds. On the basis of the results of these experiments, a series of 21-base-pair oligomers of defined sequences were designed to examine more directly the effect of sequence context on cross-linking efficiency, the preference for AT vs mixed AT−GC cross-linking sequence, and, finally, the apparent distortion of oligomers after cross-linking.

I. Studies Using Restriction Enzyme Fragments. A. Determination of the Sequence Specificities and Relative DNA Reactivities of the Pyrrole and Furan Compounds in Comparison to Bizelesin. Initially, a 5′ end-labeled 151-base-pair restriction enzyme fragment was used to compare the sequence selectivity and reactivity of the pyrrole and furan compounds with Bizelesin by use of a thermally induced strand breakage assay.[18] The results in Figure 2A show that the pyrrole compound has, in addition to the three Bizelesin alkylation sites, four other unique pyrrole alkylation sites (boxed to the right). In contrast, the furan compound is much less reactive than either of the other compounds and weakly alkylates just two of the common alkylation sites. A preliminary comparison of the sequence of the four unique pyrrole alkylation sites suggests that this compound has a higher GC tolerance than Bizelesin.

To follow up the initial conclusions made from the data in Figure 2A, a second restriction enzyme fragment (225 base pairs) with more potential (+)-CPI alkylation sites was prepared with a 5′-single-end label on the (+) and (−) strands in consecutive experiments. For each of the drugs, three different concentrations spanning a 100-fold range (0.7, 7, and 70 μM) were incubated with the 225-mer for similar periods of time, and the relative monoalkylation and cross-linking reactivities for each drug were determined as before. Phosphorimager analyses of each of the "lane 2's" in Figure 2B, containing DNA treated with the same concentration (7 μM) of all three drugs, showed that the relative amount of unmodified DNA remaining at the top of the gel is in the ratio of 24:60:1 (Bizelesin:furan compound:pyrrole compound). This shows that the pyrrole compound has a much higher reactivity toward DNA than either Bizelesin or the furan compound. By comparison of the alkylation sites at each of the lowest drug concentrations (0.7 μM, lane 1), the most reactive sequences for each drug were found to be either 5′-CTAAATA* (* represents the monoalkylation site) for both Bizelesin and the furan compound or 5′-TAGGTTA* (cross-linking site) for the pyrrole compound. The analysis of overall sequence selectivity of the three agents on this DNA fragment is as follows. First, Bizelesin shows the highest reactivity at the pure AT sequences, such as 5′-TAAATA*, 5′-ATAATA*, and 5′-TTTATA*. Second, while the pyrrole compound is quite reactive at sites containing the 5′-$(A/T)_3$A* consensus sequence, as are the other agents, it is even more reactive at those sites containing one or two G/C base pairs in the alkylation sequences that span seven base pairs, in accord with the results shown in Figure 2A. These highly reactive sites for the pyrrole compound are found in the sequences 5′-TTTCTTA*, 5′-TAGGTTA*, and 5′-GACGAAA*. The pyrrole compound can also alkylate guanine in the sequence 5′-AAACATG*, which is a unique monoalkylation site at the normally unreactive guanine. Finally, the furan compound, which has more than a 10-fold lower chemical reactivity, alkylates at the 5′-CTAAATA* sequence at the lowest drug concentration (lane 1) but reacts at other sequences, including 5′-GATAATA* and 5′-TAGGTTA*, at higher drug concentrations (lanes 2 and 3). Although the furan compound, like Bizelesin, retains the 5′-CTAAATA* sequence as the most reactive site, it is also quite reactive at the 5′-TAGGTTA* sequence, which might suggest a hybrid sequence selectivity between that of Bizelesin and the pyrrole compound.

A similar analysis was carried out on the opposite strand for each of the compounds. A quantitative comparison of the alkylation sites at a drug concentration of 7 μM is diagrammatically represented in Figure 3. Panels A, B, and C in Figure 3 are comparative data for Bizelesin and the furan and pyrrole compounds, respectively, but the furan scale is 10× lower than that for the other drugs. Analysis of this data is important in the determination of cross-linking vs monoalkylation sites and in the comparison of the relative reactivities at common alkylation sites for Bizelesin and the pyrrole and furan compounds. A summary of the results obtained from examination of the comparative data is as follows. First, most of the alkylation sites (24 sites out of 30) are common to all three drugs, and the pyrrole compound has four of the six unique alkylation sites (see asterisks in Figure 3). Second, site a′−a (5′-TATTTA*) is the most reactive six-base-pair cross-linking sequence for Bizelesin and the furan compound, site f′−f (5′-TAACCTA*) is a highly reactive seven-base-pair cross-linking sequence for both the furan and pyrrole compounds, and sites b′−b (5′-TTAGAAA*) and d′−d (5′-TAAACAA*) are quite reactive seven-base-pair cross-linking sequences for the pyrrole compound. Third, site e′−e (5′-TAAGAAA*), which is very similar to the unique seven-base-pair cross-linking site of

(16) (a) Thompson, A. S.; Hurley, L. H. *J. Am. Chem. Soc.* **1995**, *117*, 2371. (b) Thompson, A. S.; Hurley, L. H. *J. Mol. Biol.* **1995**, *252*, 86.

(17) Thompson, A. S.; Fan, J.-Y.; Sun, D.; Hansen, M.; Hurley, L. H. *Biochemistry* **1995**, *34*, 11005.

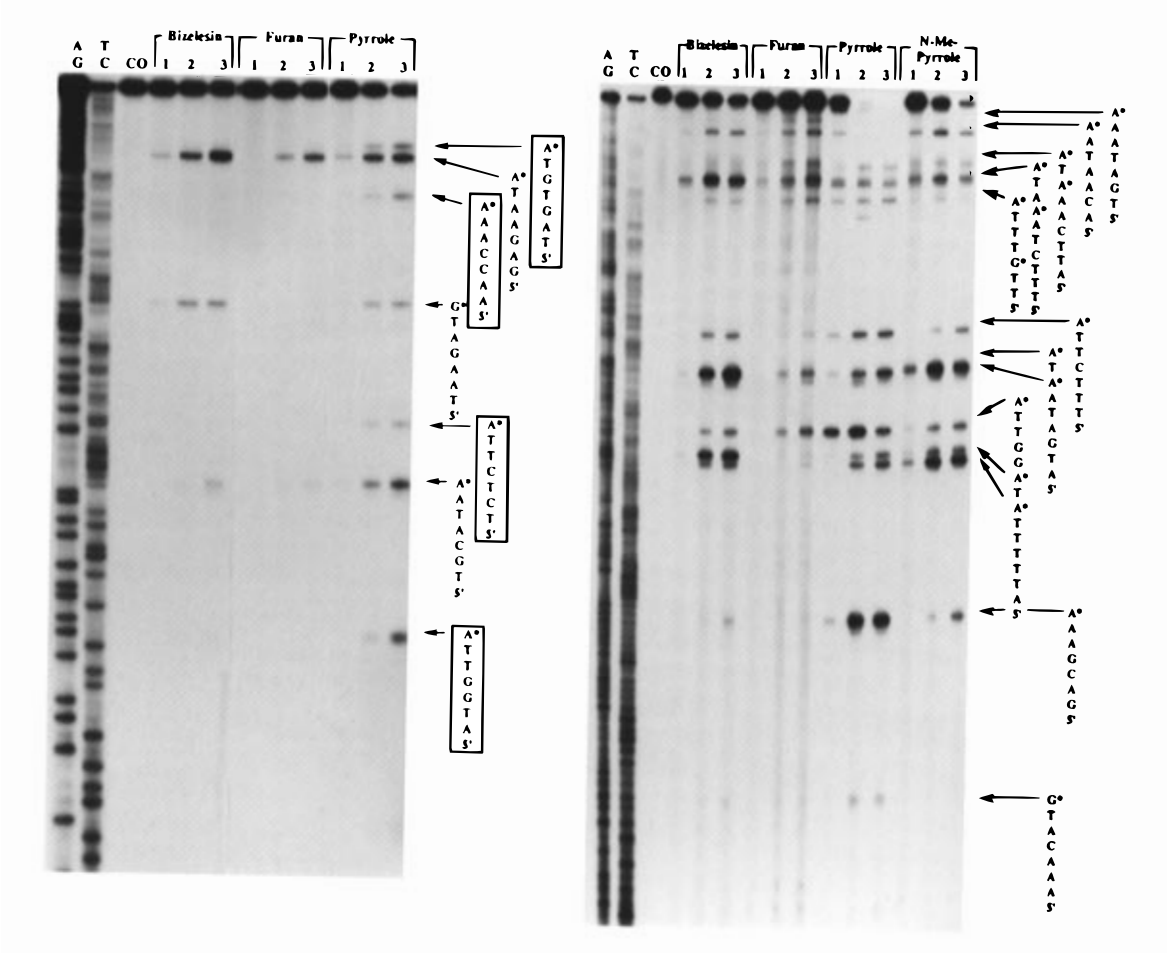


Figure 2. (A, left) Comparison of sequence specificity following thermal treatment of drug−DNA adducts of the furan and pyrrole compounds and Bizelesin. The DNA is an *Eae*I-*Hin*cII fragment (151 bp) from plasmid pBR322 that was 5′-end labeled on the top strand at the *Eae*I site. About 40 ng of DNA was modified with the indicated amount (see below) of each drug molecule and incubated at room temperature for 12 h. AG and TC represent the purine- and pyrimidine-specific chemical cleavage reactions, respectively. Lane CO is without drug treatment, and lanes 1−3 contain 0.14, 1.4, and 14 μM of three agents, respectively. The sequences alkylated are shown at the right of the gel, the asterisk indicates covalently modified adenines or guanines, and the boxes represent unique bonding sites for the pyrrole compound. (B, right) The DNA in this experiment is an *Ear*I-*Hin*dIII fragment (225 bp) from plasmid pBR322 that was 5′-end labeled on the bottom strand at the *Hin*dIII site. To compare chemical reactivity and sequence selectivity of the four compounds, a lower concentration (4 μg of DNA) and shorter incubation time (4 h) was used than for the 151-base-pair fragment. The clustered bonding sites at the top region of the autoradiogram were resolved by running the same sample on PAGE for a longer period of time. The DNA was exposed to three different concentrations of Bizelesin and the furan, pyrrole, and N-Me pyrrole compounds (0.7, 7, and 70 μM; lanes 1−3, respectively). The sequences alkylated are shown at the right of the gel; the asterisk indicates covalently modified adenines or guanines.

Bizelesin (5′-TTAGTTA*),[17] is a common seven-base-pair cross-linking site for three drugs, but the pyrrole compound shows the highest reactivity at this sequence. Fourth, in examining the sites and reactivities of the three compounds within the a′−c region, Bizelesin and the furan compound show higher reactivity with the consensus sequences 5′-TTTA* and 5′-AATA*, while the pyrrole compound has higher reactivity with the 5′-GAAA* and 5′-CTAA* sequences. This sequence preference of the three agents is characteristically observed in this region and is also mimicked in the other regions of the DNA fragment.

B. Analysis of Cross-Linking vs Mono Consensus Sequences and GC Content of High Reactivity Sites in Restriction Enzyme Fragments. A total of about 1 kb of DNA from restriction enzyme fragments obtained from pBR322 plasmid DNA has been screened for the covalent bonding sites of all three agents using three different concentrations of drug molecules (0.14, 1.4, and 14 μM). In most cases, only one of the two strands was examined, and the alkylation sites were grouped into either monoalkylating or potential cross-linking (i.e., having a thymine six and seven nucleotides to the 5′-side of the alkylated adenine for Bizelesin and the others, respectively) sites. The monoalkylating and cross-linking sites were divided into two groups based upon their comparative alkylation reactivities (high, 0.14 μM, and medium, 0.14−14 μM), but only the analyses of the high reactivity sites are shown in Table 1 (a complete listing is given in Supplementary Material). The furan compound has just a few alkylation sites at the lowest drug concentration, and therefore, corresponding alkylation sites for this agent with high reactivity are excluded from Table 1. In an examination of high reactivity cross-linking sites (Table 1), the pyrrole compound has a higher GC tolerance between the two covalently bound adenines (from position N^{+4} to N^{+1} for Bizelesin in 5′-T, N^{+4}, N^{+3}, N^{+2}, N^{+1} A* and from position N^{+5} to N^{+1} for the pyrrole compound in 5′-T, N^{+5}, N^{+4}, N^{+3}, N^{+2}, N^{+1} A*) than Bizelesin (i.e., 4.2% for Bizelesin and 22% for the pyrrole compound). For the equivalent positions in the monoalkylation sequences, GC contents for Bizelesin vs the pyrrole compound are 12.5% vs 16.5%, respectively. For Bizelesin to achieve high-reactivity cross-linking and monoalkylation, the intervening consensus sequences must be entirely A/T except for the N^{+4} position of the monoalkylation species, where

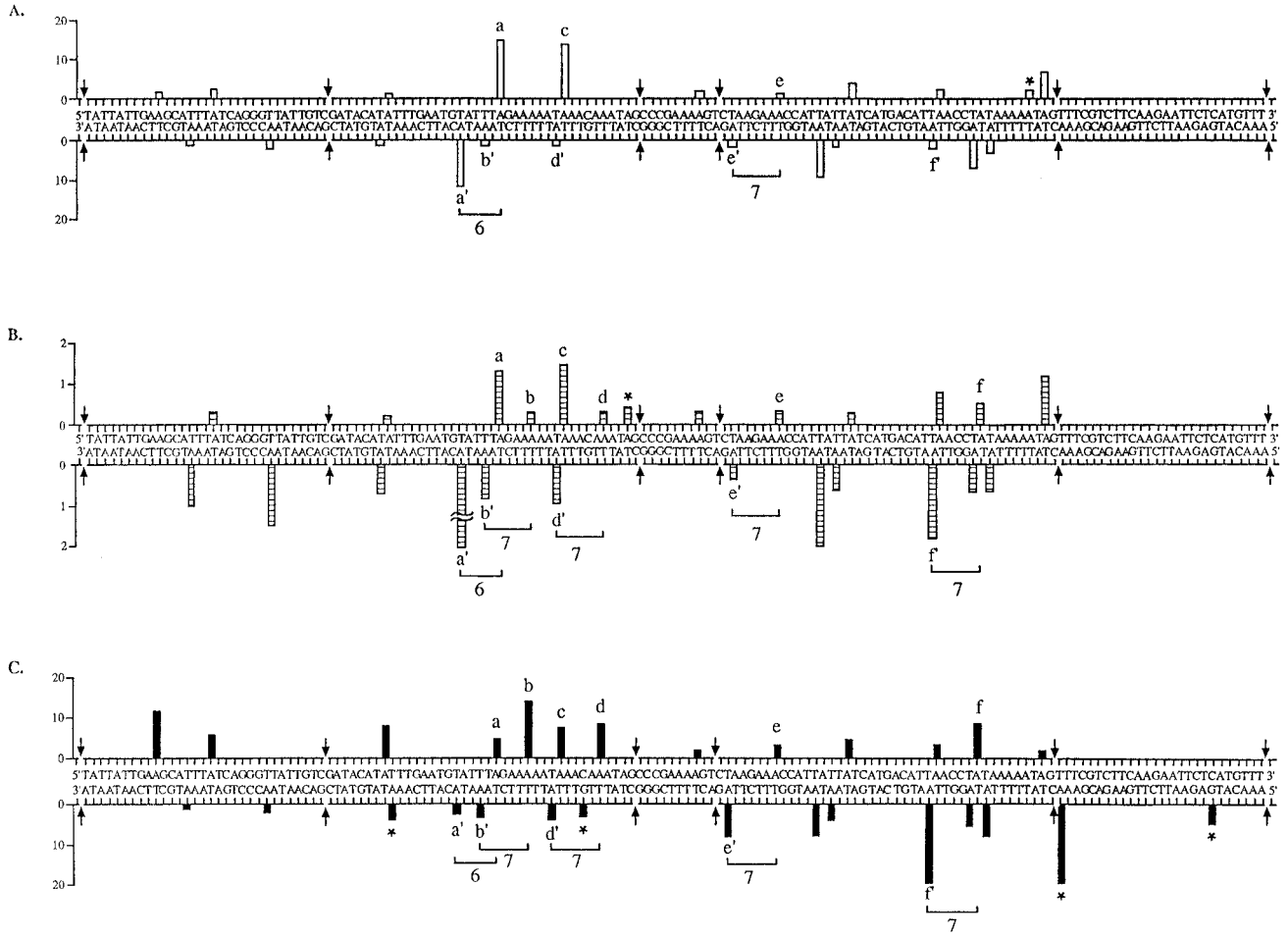


Figure 3. Diagrammatic representation of the relative reactivity of Bizelesin (A, open bars) and the furan (B, crossed bars) and pyrrole (C, filled bars) compounds at a 7 μM concentration of drug. The (−)-strand data are from Figure 2B and the (+)-strand data were obtained from a similar experiment, as described in Figure 2B, using an equivalent *Ear*I-*Hin*dIII fragment (225 bp) that was 5′-end labeled on the top strand at the *Ear*I site. The vertical axis represents the percentage of strand breakage at an indicated site relative to the total amount of radioactivity for each sample. A Molecular Dynamics Phosphorimager was used for the quantitative analysis. The alkylation sites of interest are labeled alphabetically, and the unique alkylation sites are indicated by asterisks. The brackets show the cross-linking sites, and the numbers below the brackets indicate the span of the cross-linking sites. The arrows on the horizontal axes indicate the sites of truncated DNA sequences of the DNA fragment. In plot B, the real value on the vertical axis for site a′ is 3. Note that the vertical axis scale for B is 10 times lower than that for A and C.

G is also included. The pyrrole consensus sequence is also A/T specific except for the central base pair, where in both the cross-linking and monoalkylating species there is almost an equal probability of A/T or G/C. So alongside the significantly higher G/C tolerance of the pyrrole compound there appears to be a strong preference for the GC base pair to be in the center of the seven-base-pair cross-linked species or at the equivalent +3 position of the monoalkylated species.

C. Sequence Specificity of the *N*-Methylpyrrole Compound. On the basis of the analysis of the comparative reactivity and sequence specificity of Bizelesin and the furan and pyrrole compounds, and in an effort to rationalize their differences, the *N*-methylpyrrole compound (U-107496, The Pharmacia Upjohn Co.) was synthesized.[15] This compound, which we abbreviate as "the N-Me pyrrole compound" (structure shown in Figure 1), is similar in structure to the pyrrole compound except that the pyrrole ring nitrogen is methylated. About 1 kb of DNA restriction enzyme fragments from plasmid pBR322 that had already been used for the study of sequence specificity of Bizelesin and the pyrrole and furan compounds was prepared again to examine the sequence specificity of the new analog containing the N-Me pyrrole linker. The alkylation sites for this compound were determined (listed in Supplementary Material) and compared with those for Bizelesin and the pyrrole and furan compounds. The results indicate that the alkylation sites for the N-Me pyrrole compound are highly overlapped with those for Bizelesin, and this compound is less reactive than the pyrrole compound but just as reactive as Bizelesin (Figure 2B). The amount of unmodified DNA at 7 μM of drug concentration (Figure 2B) was in the ratio of 24:17 for Bizelesin:N-Me pyrrole compound. An example of the differences in sequence specificity of Bizelesin, the furan, the pyrrole, and the N-Me pyrrole compounds is presented in Figure 2B. As before, the 5′-end-labeled 225-base-pair DNA fragment from plasmid pBR322 was exposed to three different concentrations (0.7, 7, and 70 μM) of the N-Me pyrrole compound. The alkylation sites for the N-Me pyrrole compound, in comparison to those for its parent pyrrole compound, show remarkably diminished reactivity at the two highly reactive sites for the pyrrole compound (5′-TAGGTTA* and 5′-GACGAAA) and increased reactivity at the sequences 5′-GATAATA* and 5′-ATTTTTA*. An investigation of a second DNA restriction enzyme fragment (results not shown) reveals that the N-Me pyrrole compound completely loses its reactivity at the unique bonding sites for the parent compound, such as 5′-AGTGTTA*, 5′-TATGGCA*, 5′-CTCCTTA*, and 5′-TCTCTTA*. Most of the alkylation sites at which the N-Me pyrrole compound loses its reactivity among the alkylation sites for its parent compound contain two or more G/C base pairs, which shows that the N-Me pyrrole compound has lower GC tolerance for alkylation than

Table 1. Consensus Sequences and GC Content (%) of High Reactivity Cross-Linking and Monoalkylating Sites for Bizelesin and the Pyrrole and N-Me Pyrrole Compounds[a]

	cross-linking		monoalkylating	
drug	consensus sequence	GC/B[b] (%)	consensus sequence	GC/B[b] (%)
Bizelesin	$\underline{T}\left(\frac{A}{T}\right)_4A^*$	4.2	A Pu$\left(\frac{A}{T}\right)_3A^*$ [c]	12.5
pyrrole	$\underline{T}\left(\frac{A}{T}\right)_2N\left(\frac{A}{T}\right)_2A^*$	22	Pu$\left(\frac{A}{T}\right)_2N\left(\frac{A}{T}\right)_2A^*$	16.5
N-Me pyrrole	$\underline{T}\left(\frac{A}{T}\right)_5A^*$	10	Pu$\left(\frac{A}{T}\right)_5A^*$	10

[a] Complete data sets are available in the supporting information. [b] The covalent bonding sites for all drugs are determined after screening a total of about 1 kb of DNA. The alkylation sites are grouped into either cross-linking ($\underline{T}N_{4\text{ or }5}A^*$) or monoalkylating [(A, G, or C)$N_{4\text{ or }5}A^*$] sites. The intervening sequence is either a four or five nucleotide sequence positioned between the 5′-side first base and the 3′-side alkylated adenine. The total number of DNA bases (B) and number of G or C bases (GC) in the intervening sequences (5′-T, N^{+4}, N^{+3}, N^{+2}, N^{+1} A* for Bizelesin or 5′-T, N^{+5}, N^{+4}, N^{+3}, N^{+2}, N^{+1} A* for the pyrrole or N-Me pyrrole compound) were determined, and the GC content is obtained from the equation GC content (%) = (GC/B) × 100. [c] Pu = purine.

its parent compound. The percentage of GC bases accepted by the N-Me pyrrole compound at cross-linking vs monoalkylation sites is shown in Table 1. In an examination of cross-linking sites, the N-Me pyrrole compound conforms to the consensus sequence, which is represented by all A/Ts, rather than consensus of the parent pyrrole compound [5′-T(A/T)$_2$N-(A/T)$_2$A*]. The GC content between the two covalently bound adenines (from position N^{+5} through N^{+1}) is 10%, which is intermediate between Bizelesin and the pyrrole compound. The high-reactivity monoalkylation sites for the N-Me pyrrole compound also conform to the same consensus sequence and GC contents as those for the cross-linking sites, and these are intermediate between Bizelesin and the pyrrole compound. In summary, the chemical reactivity and sequence specificity of the N-Me pyrrole compound are very similar to those of Bizelesin.

II. Studies Using The 21-Base-Pair Oligomers. A. Relative Cross-Linking and Monoalkylation Reactivities of Bizelesin and the Pyrrole and Furan Compounds in Select Cross-Linking Sequences Contained in 21-Base-Pair Oligomers. The results from the studies using restriction enzyme fragments pointed strongly toward a higher tolerance of the pyrrole compound for G/C base pairs in high-reactivity cross-linking and monoalkylation sequences. In particular, a tolerance for a *central* GC base pair in the cross-linked sequence suggested that a GC base pair might be *preferred* over the equivalent species with an AT base pair. To gain further evidence for this possibility and explore the molecular basis of the sequence selectivities of Bizelesin and the furan and pyrrole compounds, a series of several 21-base-pair oligomer sequences were designed on the basis of the results of the restriction enzyme fragments studies. (The selected 21-base-pair sequences are shown in the first column of Table 2.) Following the drug bonding reactions using the 21-base-pair oligomers labeled on either the (+)- or the (−)-strand, denaturing gel analysis (Figure 4, parts A and B) was used to separate single-stranded unmodified oligomer species (U) from both the cross-linked (Cr) and monoalkylated oligomer species (M). As examples, the results of two experiments in which 21-base-pair oligomers containing the potential cross-linking sequences 5′-TTAATTA* and 5′-TAAGTTA* were incubated with Bizelesin (A), the furan compound (B), or the pyrrole compound (C) are shown in Figure 4, parts A and B, respectively. In these examples, the oligomers 5′-TTAATTA*-3′ (lane A in Figure 4A) and 5′-TTAGTTA*-3′ (lane C in Figure 4B) occur almost entirely as cross-linked species for Bizelesin and the pyrrole compound, respectively, while mixed products (Cr and M) or unreacted oligomer (U) result from the other reactions (lanes B and C in Figure 4A and lanes A and B in Figure 4B).

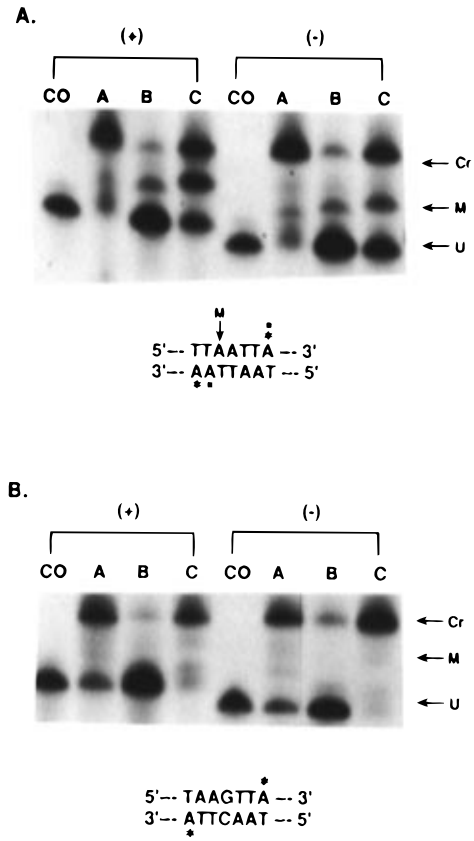


Figure 4. Autoradiograms of a 12% denaturing polyacrylamide gel showing the cross-linked adduct (Cr), monoalkylated species (M), and unreacted single-strand DNA (U) after reaction of 5′-end-labeled 21-base-pair oligomers with A (Bizelesin), B (the furan compound), and C (the pyrrole compound), respectively. CO is non-modified DNA and (+) and (−) refer to upper and lower strands of DNA. A summary of the alkylation sites obtained from the thermal treatment of each species (Cr, M, and U) is shown in each of the partial duplex sequences that appear below each gel in panels A and B. The arrow indicates the site of monoalkylation, the asterisks represent common seven-base-pair cross-linking sites for three agents, and the squares refer to the six-base-pair cross-linking site for Bizelesin.

Densitometric analyses of autoradiograms similar to those shown in Figure 4, A and B, were carried out to quantify each of the species (Cr, M, and U), and a summary of the results is shown in Table 2. Each species was also isolated from the denaturing gel and subjected to thermally induced strand breakage followed by PAGE to determine the covalent sites. As expected, Bizelesin (Table 2) forms six- or seven-base-pair cross-linked adducts at the 5′-T(A/T)AATTA* sites of the first group of oligomers and seven-base-pair cross-linked adducts at the equivalent sites of the other groups of oligomers, while the furan compound (Table 2) and the pyrrole compound (Table 2) form seven-base-pair cross-linked adducts at the equivalent sites of all of the oligomers shown in Table 2. Bizelesin shows higher cross-linking preferences for 5′-T(A/T)AATTA* sequences (average 81%), while the pyrrole compound shows higher cross-linking preference for 5′-T(A/T)$_2$(G/C)TTA* sequences (average 94%). To summarize the data obtained from the second and third groups of oligomers in Table 2, the pyrrole

Table 2. Relative Cross-Linking Reactivities of Bizelesin and the Pyrrole and Furan Compounds in Selected 21-Base-Pair Oligomers[a]

alkylation sequences[b]	percent of cross-linking[c]		
	bizelesin[d]	furan compd	pyrrole compd
5′-TTAATTA*-3′	72 ± 4	6 ± 1	45 ± 3
5′-TAAATTA*-3′	89 ± 4	19 ± 2	43 ± 4
average[d]	81	13	44
5′-TAAGTTA*-3′	54± 2	11 ± 5	94 ± 4
5′-TATGTTA*-3′	55 ±1	5 ± 1	92 ± 1
5′-TTAGTTA*-3′	40 ± 2	7 ± 1	95 ± 1
5′-TTACTTA*-3′	46 ± 6	8 ± 2	95 ± 2
average[e]	49	8	94
5′-TAAITTA*-3′	73 ± 1	16 ± 3	71 ± 4
5′-TAGATTA*-3′	19 ± 2	1	42 ± 6
5′-TGAATTA*-3′	36 ± 5	2	15 ± 3

[a] Results are from Figure 4, panels A and B, and equivalent experiments with other 21-base-pair oligomers. [b] Only the potential cross-linking sequences within the 21-base-pair oligomers (5′-TGGCCGGGG**TNNNNNA***CGGGT-3′) are shown. [c] Relative percent of cross-linked species vs monoalkyl and unmodified species in Figure 4, taking into account the differences between (+)- and (−)-strands. [d] Percent of cross-linking at the sites 5′-T(A/T)AATTA*-3′. [e] Percent of cross-linking at the sites 5′-T$(A/T)_2$(G/C)TTA*-3′.

compound shows the highest cross-linking efficiency at sequences that contain a G/C base pair in the center of the cross-linking sites and loses its reactivity remarkably as the G/C base pair is positioned away from the center (e.g., 5′-TAGATTA* and 5′-TGAATTA*). With the 5′-TAAITTA* sequence, containing an I/C base pair mismatch in the center of the cross-linking site, the pyrrole compound reveals some reduced reactivity (about 25% decrease), while the cross-linking reactivity of Bizelesin is restored to the same level as that found for the sequence 5′-T$(A/T)_2$ATTA*. The furan compound has a much lower cross-linking reactivity with all of the sequences.

B. Comparative Kinetics of Mono vs Cross-Linking Alkylation of the 21-Base-Pair Oligomers by Bizelesin and the Furan, Pyrrole, and N-Me Pyrrole Compounds. To confirm the rather decisive differences in sequence preference of the three agents shown in the 21-base-pair oligomer study (Figure 4 and Table 2), another 21-base-pair oligomer sequence was designed, this time containing overlapped high-reactivity cross-linking sites for Bizelesin and the pyrrole compound 5′-TTAATTA*GTTA*-3′ (i.e., a combination of 5′-TTAATTA* and 5′-TTAGTTA*). Kinetics experiments to determine the time-course alkylation at this overlapped sequence by these compounds were carried out, and the results are shown in Figure 5, parts A and B. The (+)- and (−)-strand-labeled oligomers were treated with each drug, and thermal cleavage was carried out directly on material isolated from the time-course reaction. For a representative comparison, the kinetics of alkylation on both strands by the pyrrole compound is shown in Figure 5A. In examining the autoradiograms, the material migrating at the strand breakage sites (site A*s) represents the total amount of either cross-linked species or monoalkylation species, and the predominant cross-linking by the pyrrole compound of the 5′-TTAGTTA* sequence [(+)- and (−)-strands in Figure 5A] is shown. The unexpected minor monoalkylation at 5′-CTTAA* by the pyrrole compound [Figure 5A, the first 5′-side A* in the (+) strand] caused some minor errors in quantitation of the alkylated species at the two downstream adenines (A*s) (5′-TTAA*TTA*GTTA*).

The results of the time-course alkylation by the pyrrole compound from Figure 5A are plotted in Figure 5B alongside the results of similar experiments (unpublished results) of Bizelesin and the furan and N-Me pyrrole compounds. The left and right panels show the cross-linking kinetics at the 5′-TTAATTA* and 5′-TTAGTTA* sequences, respectively. A comparison of the left- and right-hand panels for Bizelesin and the N-Me pyrrole compound shows their remarkable cross-linking preference for the 5′-TTAATTA* sequence. The furan compound shows a much slower alkylation reaction than the other agents but still reveals a cross-linking preference for the 5′-TTAATTA* sequence. The pyrrole compound shows the fastest initial alkylation kinetics and occupies the 5′-TTAGTTA* cross-linking site preferentially. In a summary of the overall cross-linking selectivities, Bizelesin and the furan compound form seven- or six-base-pair cross-linked adducts at either 5′-TTAATTA* (major) or 5′-TTAGTTA* (minor), and the final ratio of cross-linking efficiency at the two sites is about 75% vs 5% (15:1) for Bizelesin and 40% vs 2% (20:1) for the furan compound. The pyrrole compound produces six- or seven-base-pair cross-linked adducts at both sites and a minor amount of monoalkylation products at 5′-CTAATTA* in the (−) strand, and the final ratio of cross-linking efficiency at 5′-TTAGTTA* vs 5′-TTAATTA* is approximately 80% vs 10% (8:1). The N-Me pyrrole compound forms six- or seven-base-pair cross-linked adducts at 5′-TTAATTA* and monoalkylation products at 5′-TTAGTTA* and 5′-CTAATTA* in the (+)- and (−)-strands, respectively, and the final ratio of alkylation at 5′-TTAATTA* vs 5′-TTAGTTA* is about 65% vs 13% (5:1). Because *only* the 5′-strand breakage site is revealed in a 5′-end-labeled oligomer, in order to check for any error in quantitation of material resulting from two competing strand breakage reactions in the same DNA molecule, the same series of experiments were conducted, using another 21-base-pair oligomer containing the overlapped competitive sequence in reverse order: 5′-TTAGTTA*ATTA*-3′. Similar quantitative results for the sequence preference were obtained for all four agents (results not shown). In conclusion, the results of the experiments with the overlapped cross-linking sequences confirm those found in oligomers having just one available binding site (Table 2); i.e., while Bizelesin prefers the 5′-TTAATTA* sequence, as does the N-Me pyrrole and furan compounds, the pyrrole compound strongly prefers the 5′-TTAGTTA* sequence.

C. Determination of the Overall Distortive Effects of Cross-Linkage of Duplexes by Bizelesin and the Furan and Pyrrole Compounds. In previous experiments with Bizelesin, we have shown that the monoalkylated bent DNA sequences are converted into straight DNA structures in the case of six-base-pair cross-linked structures. To provide some initial insight into the final overall structures of the cross-linked furan and pyrrole duplexes, two 21-base-pair oligomers containing the unique reactive cross-linked sequences 5′-TAAATTA* and 5′-TAAGTTA* (panels A and B, respectively, in Figure S1 from Supplementary Material) were selected for studies to determine the overall distortive effects of cross-linking. While self-ligation of the non-drug-modified 21-mer containing the pure AT sequence (5′-TAAATTA) showed retardation of gel mobility, which is indicative of overall bending of DNA,[19] the 21-mer oligomer containing the GC base pair at the center of the seven-base-pair cross-linking sequence has no apparent intrinsic bending (Supplementary Material). Upon cross-linking of the 21-base-pair AT oligomer by all three of the ligands, the intrinsic bending is eliminated, while for the 21-base-pair GC oligomer the effects of cross-linking by the three ligands are different. While cross-linking by Bizelesin does not appear to appreciably change the bending value of ligated GC oligomers, the pyrrole has a significant distortive effect and the furan is intermediate in its distortive effects. Subsequent experiments with the pyrrole compound were attempted to determine the locus of the

(18) Reynolds, V. L.; Molineux, I. J.; Kaplan, D. J.; Swenson, D. H.; Hurley, L. H. *Biochemistry* **1985**, *24*, 6228.
(19) Koo, H.-S.; Wu, H.-M.; Crothers, D. M. *Nature* **1986**, *320*, 501.

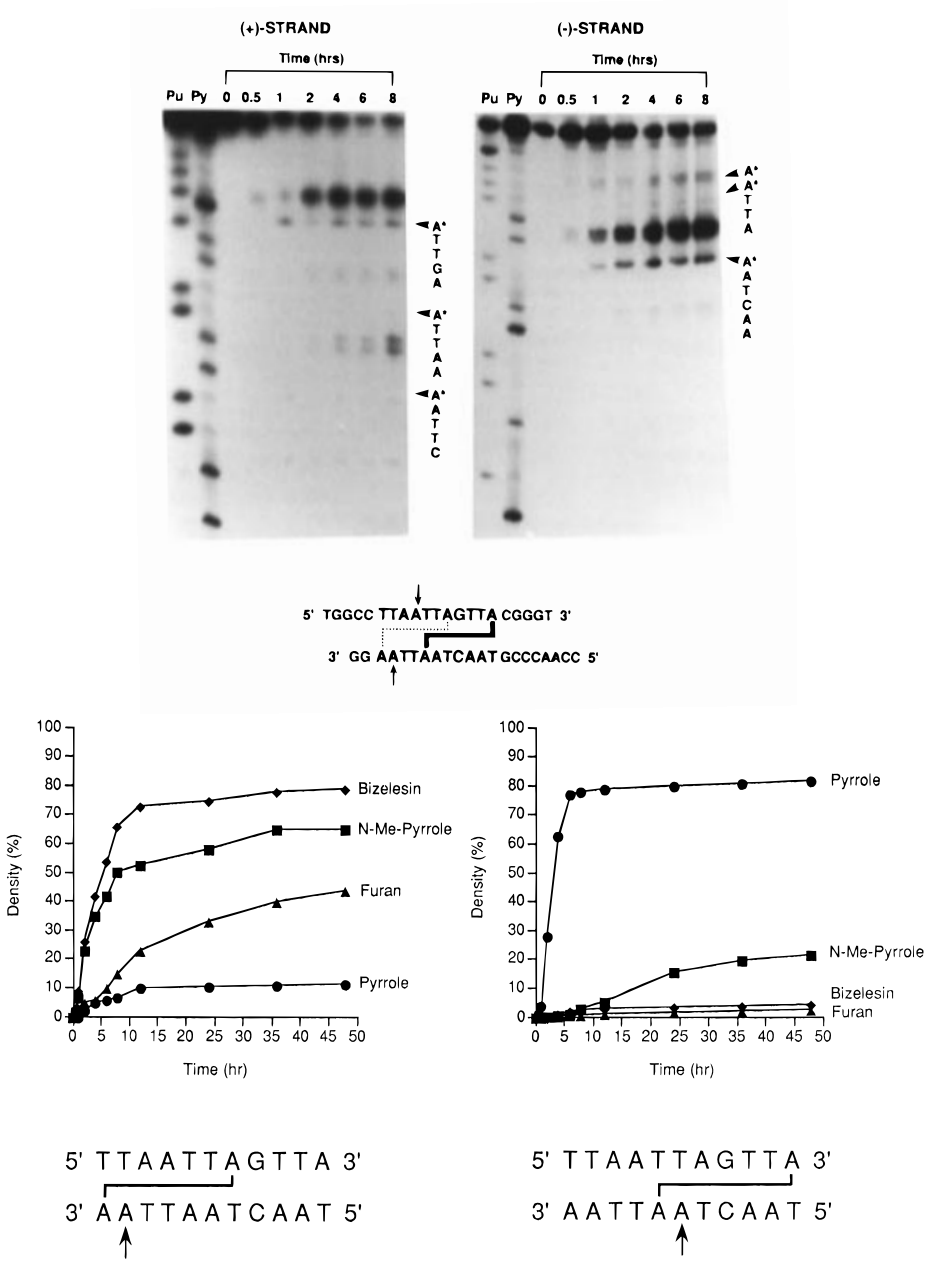
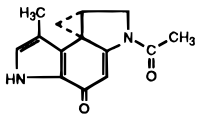


Figure 5. Kinetics of mono vs cross-linked adduct formation by Bizelesin and the furan, pyrrole, and N-Me pyrrole compounds on the 21-mer oligomer duplex containing the overlapping sequence 5′-TTAATTA*GTTA*-3′. (A, Top) The results of PAGE analysis after thermal treatment of DNA modified with the pyrrole compound on the (+)- and (−)-strands, respectively. The results shown are from the early time periods (0.5−8 h), while the actual time course was up to 48 h. The arrowheads point to the position of products that result from two successive β-eliminations that give rise to a band equivalent to the Maxam−Gilbert A sequences product.[5d] Because of insufficient thermal treatment, the majority of the product runs higher, due to incomplete degradation of the product resulting from depurination. (B, Bottom) Densitometric tracings obtained from four pairs of autoradiograms similar to those shown in panel A for Bizelesin and the pyrrole, furan, and N-Me-pyrrole compounds were used to determine the percentage of the strand breakage at a given site, and from this data the percentage of the two types of cross-linked adducts for each of the four agents was determined (the two cross-linked products are shown below the graphs). Arrows point to Bizelesin cross-linkage sites on the (−) strand.

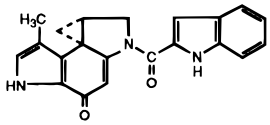
distortive effect and direction of possible bending in the GC 21-mer oligomer, but these led to electrophoretic mobility patterns that were difficult to interpret (unpublished results), although speculative analysis suggests that the pyrrole compound winds the helix of the GC oligomer the equivalent of one base pair, and the direction of the "bending-like distortion" was toward the minor groove (unpublished results).

Discussion

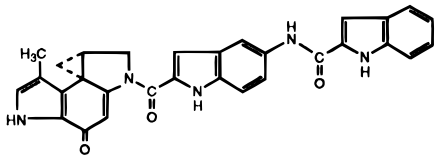
The pyrrole and furan compounds described in this paper were originally derived from the (+)-CPI unit of (+)-CC-1065, which is a minor groove sequence specific alkylating agent. Structurally they are most closely related to the DNA−DNA cross-linker Bizelesin by substitution of the ureadiyl linkage by the more elaborate bisamido pyrrole or furan linkers (see Figure 1). In comparison to Bizelesin, the pyrrole is more biologically potent[15] but, most significantly, now has an added requirement for a GC base pair not found in the Bizelesin consensus sequence. Furthermore, the position of this GC base pair for high reactivity is specified at the central position of the seven-base-pair cross-linked sequence (e.g., 5′-TAAGTTA*), and this GC base pair cannot be replaced by an IC base pair. Since this central GC base pair is not required by either the N-Me pyrrole or furan compound, and inosine cannot substitute for guanine,



(+)-CPI



(+)-CPI-I



(+)-CPI-I_2

Figure 6. Structures of (+)-CPI, (+)-CPI-I, and (+)-CPI-I_2.

this is suggestive of H-bonding recognition involving donor−acceptor pairs between the minor groove 2-NH_2 group of guanine and the bisamido pyrrole linker (see later).

The Chemical Evolution of a (+)-CPI-Derived Drug That Has a Requirement for a Positional GC Base Pair within the Cross-Linked AT Sequence. For high reactivity, the covalent reaction of (+)-CC-1065 with DNA has a strict requirement for an AT-rich region immediately upstream of the alkylation site (i.e., 5′-AAA* or TTA*).[5b] A surprising and very significant finding was that the sequence specificity of the tri-subunit (+)-CPI-I_2 could be specified by *just* the (+)-CPI unit (Figure 6).[5c] Consequently, in sequence specificity experiments in which this trisubunit compound was compared to bi- and monosubunit compounds [(+)-CPI-I and (+)-CPI], it was found that the (+)-CPI subunit had sufficient structural information to dictate the sequence specificity of the entire trisubunit molecule. The unambiguous conclusion from these experiments is that it was the covalent bonding reaction, *not* the binding interactions, of the noncovalently binding subunits that dictated the overall sequence specificity of the (+)-CPIs.[5b−d,f,20] This is in sharp contrast to the (−)-CPI drugs, where the noncovalent reactions play the predominant role.[5d] The dilemma of how the covalent reaction could mediate a sequence selectivity that included a sequence in which the drug *did not overlap* was partially solved when it was discovered that (+)-CPI alkylation of DNA required a distorted DNA structure.[21] Furthermore, the distortion was characterized as a bent DNA structure very similar in conformational character to the intrinsic bending of A-tracts.[22] Indeed, the 3′ adenine of A-tracts was found to be a highly reactive site for monoalkylation,[22a] as are other sequences (e.g., AGTTA*) that have an unusual propensity to form bent DNA structures.[22c,d] Thus, the relative flexibility of sequences at and to the 5′-side of alkylation sites was pinpointed as an important factor for sequence recognition,[5b,22c,d] as was also catalytic activation by DNA,[23] leading to an overall rate acceleration at specific sites.[24] Alkylation of DNA by (+)-CC-1065 therefore has a strict requirement for both catalytic activation enforced by a precisely positioned binding pocket for the (+)-CPI unit and for upstream conformational flexibility in DNA, since the presumed transition state intermediate for the alkylation resembles the final bent DNA structure. As will be described later, both of these requirements (catalytic activation and conformational flexibility) have important implications for the sequence specificity of cross-linkers such as Bizelesin and the new, more elaborate compounds reported here.

In its simplest form, interstrand cross-linking can be envisaged as two consecutive monoalkylations occurring in close proximity on opposite DNA strands. Since Bizelesin and the pyrrole and furan compounds are made up of two (+)-CPI units in a self-complementary molecule, if the stereoelectronic environment for the second alkylation is similar to the first, then the cross-linking reaction should proceed very rapidly after monoalkylation. A simple test of this principle for two successive alkylations by the component monoalkylating units of the P(1,4)-Bs and (+)-CPIs led to quite different structural outcomes. For the P(1,4)Bs, two tomaymycin molecules are able to alkylate a 12-mer duplex on *opposite* strands four base pairs apart, predictive of the nondistorted structure of the DSB-120 cross-linked DNA structure.[9,25] However, for successive alkylations by two (+)-CPI-I units on the Bizelesin cross-linking sequences 5′-TAATTA* or 5′-TAAAAA*, the alkylations occurred on the same strands, i.e., 5′-TAA*TTA* and 5′-TAA*AAA*, with no detectable product foretelling the Bizelesin interstrand cross-linked pattern.[26] Clearly there is an inherent difference between these two cross-linking systems that can be traced back to the distortive vs nondistortive nature of the monoalkylated products; that is, while the P(1,4)Bs and (+)-CPIs both require a B-form duplex for reaction, it is *only* the (+)-CPIs that produce a distortion propagated to the upstream side of the adduct that provides an environment unsuitable for the cross-linking reaction on the opposite strand.

All cross-linkers derived from two (+)-CPI subunits have to face the common problem of how to convert a bent DNA structure into a duplex substrate suitable for cross-linking. For each of the two previous six-base-pair Bizelesin cross-linked sequences examined, the solution to this distortion-based problem is different. For the sequence 5′-TAAAAA*, there is a unique solution based upon the equilibrium that exists between a straight B-form structure and a bent non-B-form structure.[16] In this case, the straight B-form structure is selectively trapped out as the Bizelesin cross-linked form. For the 5′-TAATTA* sequence, again there is a unique solution, but in this case it involves a drug-assisted rearrangement of the two central AT base pairs that are trapped out either as Hoogsteen base pairs or as a non-base-paired step within a normal B-form DNA.[11] Crucially, in both cases (5′-TAAAAA* and 5′-TAATTA*) the cross-linkage site is restored to B-form prior to the second alkylation reaction. Therefore, it appears that cross-linkage of duplexes that proceed through a distorted monoalkylation adduct

(20) This point has been disputed in the past (ref 5, g−m), but the results and interpretation of this simple experiment are difficult to dismiss, despite complex and confusing arguments to the contrary.

(21) Hurley, L. H.; Needham-VanDevanter, D. R.; Lee, C.-S. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6412.

(22) (a) Lin, C. H.; Sun, D.; Hurley, L. H. *Chem. Res. Toxicol.* **1991**, *4*, 21. (b) Lin, C. H.; Hill, G. C.; Hurley, L. H. *Chem. Res. Toxicol.* **1992**, *5*, 167. (c) Sun, D.; Lin, C. H.; Hurley, L. H. *Biochemistry* **1993**, *30*, 4487. (d) Hurley, L. H.; Sun, D. *J. Mol. Recog.* **1994**, 7, 123.

(23) Lin, C. H.; Beale, J. M.; Hurley, L. H. *Biochemistry* **1991**, *30*, 3597.

(24) Warpehoski, M. A.; Harper, D. E. *J. Am. Chem. Soc.* **1994**, *116*, 7573.

(25) (a) Boyd, F. L.; Cheatham, S. F.; Remers, W.; Barkley, M. D.; Hurley, L. H. *Biochemistry* **1990**, *29*, 2387. (b) Wang, J.-J.; Hill, G. C.; Hurley, L. H. *J. Med. Chem.* **1992**, *35*, 2995. (c) Bose, D. S.; Thompson, A. S.; Ching, J.; Hartley, J. A.; Berardini, M. D.; Jenkins, T. C. *J. Am. Chem. Soc.* **1992**, *114*, 4939. (d) Bose, D. S.; Thompson, A. S.; Smellie, M.; Berardini, M. D.; Hartley, J. A.; Jenkins, T. C.; Neidle, S.; Thurston, D. E. *J. Chem. Soc., Chem. Commun.* **1992**, *20*, 1518.

(26) Seaman, F. C.; Chu, J.; Hurley, L. H. *J. Am. Chem. Soc.* **1996**, *118*, 5383.

requires an additional step in which the drug either facilitates a rearrangement of the duplex or awaits the non-drug-assisted reemergence of a normal B-form duplex before the second alkylation can occur readily. Thus , the energetically highly unfavored reaction of a distorted monoalkylated adduct is rendered more favorable, even though the final duplex conformation cannot be the idealized bent DNA structure produced after monoalkylation. It is the immobilization of the second (+)-CPI unit at the latent cross-linking site by the initial monoalkylation event that allows for normally rare and unexpected conformational forms to be selectively trapped out by these cross-linking reactions.

The facilitated rearrangement of the 5′-TAATTA* sequence by the interaction of the linkage of Bizelesin with the central AT base pair provides the impetus to design more elaborate linkers with presumably increased possibilities for interactions with suitably positioned minor groove H-bond acceptor and donor groups. These linkages would likely afford even more complex interactions with AT and GC base pairs, perhaps facilitating even more elaborate rearrangements, giving rise to even greater selectivities. Of the three new linker forms examined, clearly the bisamido pyrrole is the more intriguing species.

Bizelesin cross-linkage of six-base-pair sequences results in a straight DNA species. This is in contrast to the pyrrole compound in which the seven-base-pair cross-linked species is a distorted DNA structure. The precise nature of this distorted species awaits full characterization, but preliminary conclusions from analysis of the one- and two-dimensional NMR data on an 11-mer duplex containing the seven-base-pair pyrrole cross-linked sequence 5′-TTAGTTA*-3′ show disruption of the H1′ to aromatic walks and upfield-shifted imino protons for the central three base pairs, suggestive of a non-base-paired region, and a nonstacked cytosine, indicative of an overall gross distortion of the central region of the duplex. Although it is premature to speculate on the precise structure of the distortion trapped out by cross-linking the seven-base-pair pyrrole compound, it is clear that the distortion is localized to the center of the duplex and is not an A-tract-like bend.

On the basis of the comparative abilities of the furan, pyrrole, and N-Me pyrrole compounds to cross-link the 5′-TNNGNNA* sequence and the requirement for the exocyclic 2-amino group of guanine, some speculative role for the bisamido pyrrole linker can be ascribed in the induced rearrangement of the central region of the cross-linked species. Clearly the pyrrole NH is essential for this rearrangement, since neither the furan nor the N-Me pyrrole can facilitate this rearrangement. This requirement, alongside that of the 2-amino group of guanine, implies a role for two donor−acceptor pairs in the rearrangement. Molecular modeling provides a clue as to how the initial molecular interactions between the bisamido pyrrole and the 5′-AGT step minor groove substituents might occur. A DOCK search was carried out to identify the best conformer of the pyrrole compound (selected from a conformer library generated using a torsional grid search) to fit into a B-form 11-mer duplex containing the 5′-TTAGTTA-3′ sequence. The highest scoring conformer was obtained from the output of the DOCK searches, manually docked into the 5′-TTAGTTA*-3′ site of the DNA, and then covalently linked to N3 of the 3′-side adenine. The unfavorable contacts in this initial model of monoalkylated adduct were removed by conjugate gradient minimization without NMR distance constraints, and the molecular interactions between the bisamido pyrrole and 5′-AGT step DNA minor groove substituents were identified. As shown in Figure 7, the amide (the one closer to the alkylated CPI-I) carbonyl oxygen

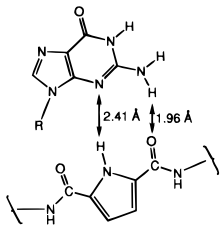


Figure 7. The arrangement of the pyrrolo−carbonyl unit of the pyrrole compound and central guanine of DNA in the preliminary model of the monoalkylated adduct. The atoms positioned in proximal distance are indicated by arrows and measured distances (Å) between those atoms are presented.

and the pyrrole NH groups were closely docked into the central guanine, which is consistent with strong hydrogen bonding. In this monoalkylated adduct, the distance between the methenyl carbon to be alkylated in the second alkylation arm and the N3 of adenine on the opposite strand was about 4 Å, and thus severe structural distortion is required before cross-linking can occur. While this preliminary modeling study does not provide an obvious explanation for the subsequent rearrangement, it does provide a rationalization for uniqueness of the pyrrole compound and the requirement for the central GC base pair in the cross-linked sequence. Solving the structure of the seven-base-pair cross-linked species will be essential to providing a firmer basis for understanding this requirement for a mixed AT−GC sequence.

Conclusions

In this article we have presented a rationale for how the pyrrole compound shown in Figure 1 is able to specifically cross-link sequences with high reactivity that have a mixed AT−GC content. We propose that this specificity is dependent upon a rearrangement or distortion of the duplex that results from an interaction between the bisamido pyrrole linker of the ligand and the central GC base pair of the duplex. This central distortion is required to counteract the bending associated with the monoalkylation product that initially renders the cross-linking site an unsuitable substrate for alkylation. While monoalkylation may occur at many sites with the correct consensus sequence for (+)-CPI reaction (i.e., 5′-$(A/T)_2$A*), the cross-linking reaction has a more restricted sequence requirement, because only DNA substrates that can be restored to a B-form structure around the target cross-linking site become available for cross-linkage. Immobilization of the ligand bound at the monoalkylation site allows exploration of a full range of duplex conformations for possible structures suitable for cross-linking. This approach represents a parallel evolution of cross-linking sequence recognition molecules to the 1:1 and 2:1 distamycin motif[4] that are already available for mixed sequence recognition.

Experimental Procedures

Chemicals and Enzymes. Bizelesin was a gift from The Pharmacia Upjohn Company, and electrophoretic reagents (acrylamide, ammonium persulfate, bisacrylamide, and *N*,*N*,*N*′,*N*′-tetramethylethylene diamine) were purchased from Biorad. Plasmid DNA (pBR322), calf intestinal phosphatase (CIP), and restriction endonuclease were from New England Biolabs. T4 polynucleotide kinase and T4 ligase were from United States Biochemical Co. (γ-^{32}P)-ATP and X-ray film were from Amersham and intensifying screens developing and chemicals were from Kodak. The pyrrole, furan, and N-Me pyrrole compounds were synthesized as will be described in a subsequent manuscript.[15]

DNA Restriction Fragments. Plasmid DNA was digested with the appropriate restriction endonuclease, dephosphorylated with CIP, and labeled with (γ-^{32}P)-ATP and T4 polynucleotide kinase at the 5′-end.

After the second restriction enzyme digestion, uniquely end-labeled DNA fragments were isolated on an 8% non-denaturing polyacrylamide gel.

Drug Treatments of DNA and Determination of Drug-Modification Sites. The reaction between drugs and 5′-end-labeled restriction fragments was conducted in 1 mM Tris-HCl, 10 mM NaCl, pH 7.4, at room temperature for 4 or 12 h, and unreacted drug molecules were removed by phenol/$CHCl_3$ extraction followed by ethanol precipitation. DNA pellets were dissolved in distilled water, incubated in a 95 °C water bath for 15 min to give strand breakage at the drug modification sites, and then lyophilized. Equivalent amounts of radioactivity for each sample were loaded onto an 8% denaturing sequencing gel. After electrophoresis, gels were dried over DE81 paper and exposed to X-ray film.

Preparation of Oligonucleotides. A series of oligonucleotides (Table 2 and Figure 6) were synthesized on an Applied Biosystems solid-phase synthesizer using the phosphoramidite method. The oligomers were deprotected with concentrated ammonium hydroxide by heating at 55 °C overnight. The ammonium hydroxide solution was evaporated under vacuum, and the dried pellets were redissolved in distilled water.

5′-^{32}P-End-Labeling of Oligonucleotides. Individual strands of the 21-mer (Table 2) were labeled with (γ-^{32}P)-ATP and T4 polynucleotide kinase and hybridized to an excess of unlabeled complementary strand. The resulting 5′-^{32}P-end-labeled duplexes were loaded onto an 8% nondenaturing polyacrylamide gel, located on the gel using autoradiography, excised from the gel, minced with a blade, and extracted with annealing buffer (10 mM Tris-HCl, 100 mM NaCl, pH 7.4).

Drug Modification of DNA Oligomers and Purification of Drug−DNA Adducts Using Denaturing Polyacrylamide Gel Electrophoresis (PAGE). Duplexes (about 50 ng) were modified with drug (final concentration, 140 μM) in annealing buffer at room temperature for 3 days, and unreacted drug molecules were removed by phenol/$CHCl_3$ extraction and ethanol precipitation. DNA pellets were dried under vacuum, redissolved in 15 μL of alkaline tracking dye [10 mM NaOH/80% (v/v) formamide], heated at 70 °C for 1 min just before loading samples on a prewarmed 12% denaturing polyacrylamide gel, and electrophoresed to separate cross-linked DNA (Cr) from monoalkylated (M) and unreacted (U) DNA. Each DNA species was located on the gel using autoradiography, and the autoradiograms were scanned with an Ultrascan XL laser densitometer (LKB 2202) to measure the percentage of each DNA species relative to the total amount of DNA. Drug-modified species (Cr and M) were extracted from the gel, dissolved in distilled water, and heated at 95 °C for 30 min to induce strand breakage. Cross-linking and monoalkylation sites for each drug were determined using 20% denaturing sequencing gel electrophoresis.

Kinetics Experiments. Fifty nanograms of duplex was treated with drug (final concentration, 140 μM) in 100 μL of annealing buffer. Five microliter aliquots of sample was isolated at a given time, immediately mixed with 2 μg/μL of calf thymus DNA, and frozen in −70 °C to stop the reaction. Every sample was incubated in a 95 °C water bath for 30 min to determine the drug-modification sites and quantify the amount of alkylation.

Kinasing, Ligation, and Nondenaturing Gel Electrophoresis. The oligonucleotides (Figure 6), each mixed with its complementary strand, were 5′-labeled with (γ-^{32}P)-ATP according to the procedure previously described.[27] Kinased duplexes were reacted with each drug at room temperature for 3 days, and cross-linked DNA adducts were separated using 20% denaturing preparative gel electrophoresis. Cross-linked and non-cross-linked DNA oligomers were self-ligated by T4 ligase to make multimers.[27] The ligated products were run on an 8% nondenaturing polyacrylamide gel at room temperature, and the ligation ladder was located by autoradiography.

Acknowledgment. Research has been supported by the Welch Foundation and the Public Health Service (CA-49751). We thank David Bishop for preparing, proofreading, and editing the manuscript.

Supporting Information Available: Tables of high, medium, and low reactivity sites for Bizelesin and the furan, pyrrole, and N-Me pyrrole compounds with associated consensus sequences, and a figure (S1) in which R_L is plotted against ligated oligomer length for drug modified 21-mers (8 pages). See any current masthead page for ordering and Internet access instructions.

JA961678A

(27) Lee, C.-S.; Sun, D.; Kizu, R.; Hurley, L. H. *Chem. Res. Toxicol.* **1991**, *4,* 203.