Solution Structure of the Complex between the Head-to-Tail Dimer of Calicheamicin γ_1^I Oligosaccharide and a DNA Duplex Containing d(ACCT) and d(TCCT) High-Affinity Binding Sites

Giuseppe Bifulco,[†,‡] Aldo Galeone,[‡] K. C. Nicolaou,*[,†,§] Walter J. Chazin,*[,⊥] and Luigi Gomez-Paloma*[,†,‡]

Contribution from the Departments of Chemistry and Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California 92093, and Dipartimento di Chimica delle Sostanze Naturali, Universita' degli studi di Napoli "Federico II", via D. Montesano 49, Napoli 80131, Italy

Received November 14, 1997

Abstract: The head-to-tail dimer of the calicheamicin oligosaccharide domain exhibits substantially higher DNA binding affinity and sequence selectivity and greater bioactivity than the monomer from which it is derived. To determine the structural basis for these functional properties, the solution structure of the 1:1 complex between the head-to-tail dimer of the calicheamicin oligosaccharide and the oligonucleotide duplex d(GCACCTTCCTGC)·d(GCAGGAAGGTGC) has been solved by restrained molecular dynamics calculations using NMR-derived distance and torsion angle constraints. The final input data consisted of 562 internuclear distance and 114 dihedral angle constraints, an average of 27 constraints per residue. In contrast to observations made for a complex between a DNA duplex and the head-to-head dimer of calicheamicin oligosaccharide, the head-to-tail dimer exhibits a unique binding mode in the DNA minor groove. A comparative analysis of the carbohydrate−DNA interactions at the two different binding sites explains at the atomic level how calicheamicin derivatives are able to effectively recognize both d(ACCT) and d(TCCT) sites. This study brings deeper insight into the factors governing DNA-binding affinity and the sequence preferences of calicheamicin and its derivatives.

Introduction

Despite the vast potential of new therapeutic modalities such as antisense and gene therapies, small molecule ligands remain as the mainstay of pharmaceutical agents in use today. Molecules that recognize and bind to DNA have been widely studied, and some have found clinical applications, for example, in cancer chemotherapy. One promising class of these agents are those that not only bind to DNA but also perform chemistry on this substrate. The enediynes, which include calicheamicins, esperamicins, dynemicin A, neocarzinostatin chromophore, kedarcin chromophore, and C1027 chromophore, form one family of extremely potent natural products of this type.

Calicheamicin γ_1^I (**1**)[1] (Scheme 1) has the strongest cytotoxicity among the natural enediynes, and it exhibits an intriguing mode of binding to DNA.[2,3] Two chemical functionalities have been shown to be responsible for its high cytotoxic activity (Scheme 1): a molecular device, the enediyne aglycon that upon activation forms a phenyl diradical, which leads to the cleaveage of the DNA, and a binding moiety (the carbohydrate domain) that directs the reaction to certain sequences. Previous structural studies from this and other laboratories have shown that the oligosaccharide moiety binds in the DNA minor groove.[4] The binding to duplex DNA actually occurs in a selective manner, with high affinity for four specific, primarily polypurine/polypyrmidine sites: d-TCCT·d-AGGA, d-TTTT·d-AAAA, d-TCTC·d-AGAG, d-ACCT·d-AGGT.[5]

The methyl glycoside of calicheamicin (**2**) has been shown to inhibit binding of transcription factors to sequences containing the TCCT high-affinity binding site and to interfere with the corresponding DNA transcription in vivo.[6] As part of the effort to explore and build upon this remarkable observation, we have designed and synthesized oligomers of the calicheamicin oligosaccharide domain.[7] The concept of linking two or more DNA binding agents to generate higher binding affinity, greater

[†] Department of Chemistry, The Scripps Research Institute.
[‡] Universita' degli studi di Napoli "Federico II".
[§] University of California at San Diego.
[⊥] Department of Molecular Biology, The Scripps Research Institute.

(1) (a) Nicolaou, K. C.; Dai, W. M. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 1387. (b) Nicolaou, K. C.; Smith, A. L.; Yue, E. W. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5881.

(2) De Voss, J. J.; Townsend, C. A.; Ding, W. D.; Morton, G. O.; Ellestad, G. A.; Zein, N.; Tabor, A. B.; Schreiber, S. L. *J. Am. Chem. Soc.* **1990**, *112*, 9669.

(3) Hangeland, J. J.; De Voss, J. J.; Health, J. A.; Townsend, C. A.; Ding, W. D.; Ashcroft, J. S.; Ellestad, G. A. *J. Am. Chem. Soc.* **1992**, *114*, 9200.

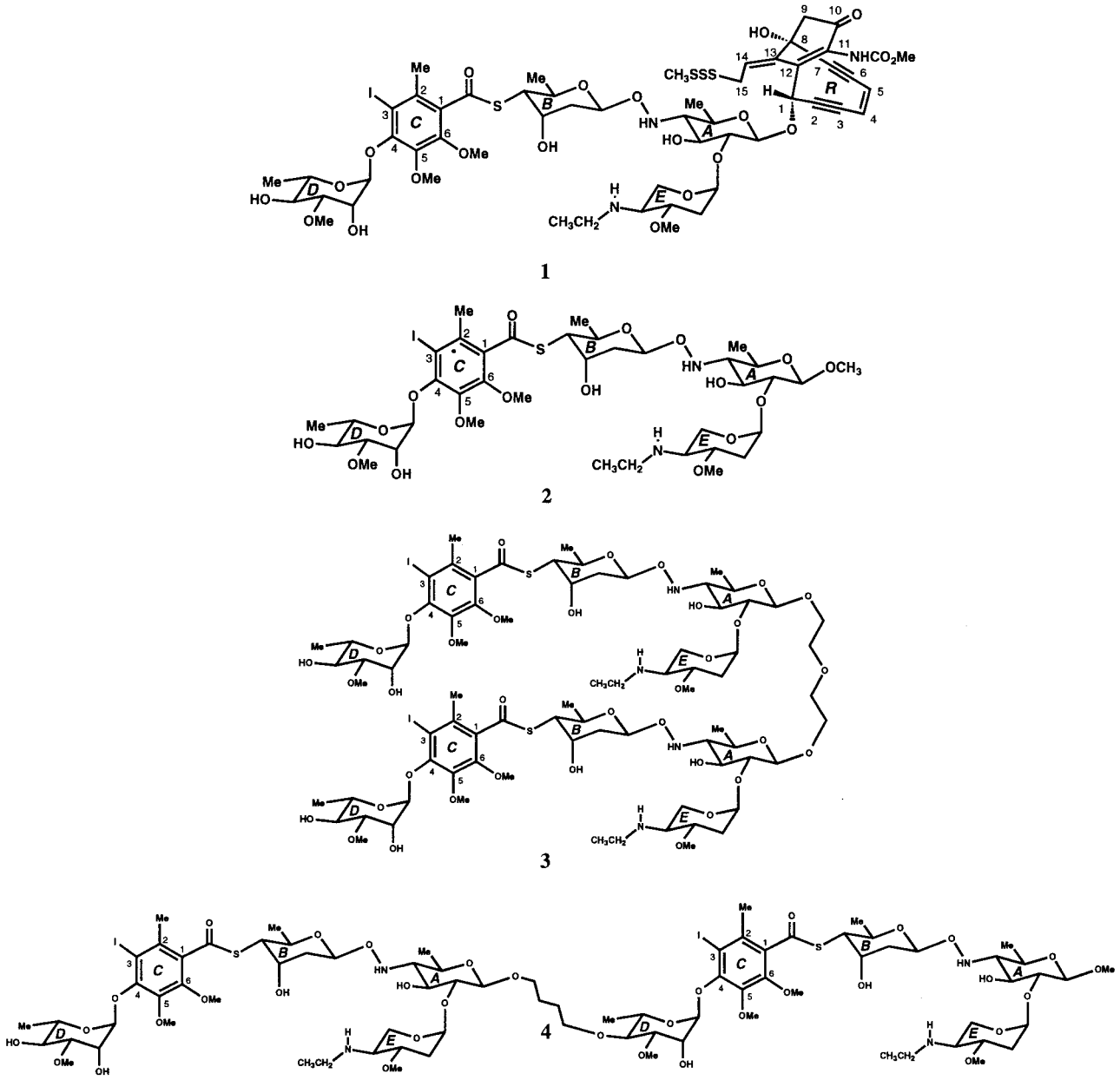
(4) (a) Walker, S. L.; Murnick, J.; Kahne, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 7954. (b) Walker, S. L.; Andreotti, A. H.; Kahne, D. E. *Tetrahedron* **1994**, *50*, 1351. (c) Gomez Paloma, L.; Smith, J. A.; Chazin, W. J.; Nicolaou, K. C. *J. Am. Chem. Soc.* **1994**, *116*, 3697. (d) Ikemoto, N.; Kumar, R. A.; Ling, T. T.; Ellestad, G. A.; Danishefsky, S. J.; Patel, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10506. (e) Smith, J. A.; Gomez-Paloma, L.; Case, D. A.; Chazin, W. J. *Magn. Reson. Chem.* **1996**, *34*, S147. (f) Kumar, R. A.; Ikemoto, N; Patel, D. J. *J. Mol. Biol.* **1997**, 265, 187.

(5) (a) Zein, N.; Sinha, A. M.; McGabren, W. J.; Ellestad, G. A. *Science* **1988**, *240*, 1198. (b) Walker, S.; Landovitz, R.; Ding, W. D.; Ellestad, G. A.; Kahne, D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4608.

(6) Ho, S. N.; Boyer, S. H.; Schreiber, S. L.; Danishesfky, S. J.; Crabtree, G. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9203.

S0002-7863(97)03910-3 CCC: $15.00
Published on Web 07/11/1998

Scheme 1. Structures of Calicheamicin (1), Calicheamicin Oligosaccharide Domain (2), Head-to-Head Calicheamicin Oligosaccharide Dimer (3), and Head-to-Tail Calicheamicin Oligosaccharide Dimer (4)



sequence selectivity, or longer recognition sites has already been exploited for a variety of DNA binding ligands. Distamycin is probably the most well-studied example; analogues were shown to have increased affinity many years ago, and ultimately, composite distamycin-based agents capable of high sequence specificity have been developed.[8] The calicheamicin analogues examined in this study represent a unique effort to explore the DNA ligand oligomerization concept using a carbohydrate motif.

The head-to-head[7a] (**3**) and head-to-tail[7b] (**4**) dimers (Scheme 1) of calicheamicin oligosaccharide domain have been shown

5'-G C **A C C T T C C T** G C-3'
3'-C G **T G G A A G G A** C G-5'

Figure 1. DNA sequence with the dual head-to-tail binding site in bold and the two subsites identified by the arrows.

to possess superior binding properties and improved bioactivity relative to the monomer. The dimers exhibit increased binding affinity ($K_D \sim 10^{-6}$ M for the monomer vs $K_D \sim 10^{-9}$ for the dimers) and site specificity (100−1000-fold higher than the monomer vs the theoretical increase of 64-fold),[9] as well as more than 10-fold greater inhibition of transcription factor binding and transcriptional activity.[10]

In this paper, we report the results of the NMR structural analysis of the complex between the head-to-tail dimer (HTD)

(7) (a) Nicolaou, K. C.; Ajito, K.; Komatsu, H.; Smith, B. M.; Li, T.; Egan, M. G.; Gomez-Paloma, L. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 576. (b) Nicolaou, K. C.; Ajito, K.; Komatsu, H.; Smith, B. M.; Bertinato, P.; Gomez-Paloma, L. *Chem. Commun.* **1996**, 1495.

(8) (a) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1985**, *107*, 5528. (b) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564. (c) Reviewed in: Wemmer, D. E.; Dervan, P. B. *Curr. Opin. Struct. Biol.* **1997**, *7*, 355. (d) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468. (e) Hélène, C. *Nature* **1998**, *391*, 436.

(9) (a) Li, T.; Zeng, Z.; Estevez, V. A.; Baldenius, K. U.; Nicolaou, K. C.; Joyce, G. F. *J. Am. Chem. Soc.* **1994**, *116*, 3709. (b) Liu, C.; Smith, B. M.; Ajito, K.; Komatsu, H.; Gomez Paloma, L.; Li, T.; Theodorakis, E.; Nicolaou, K. C.; Vogt, P. K. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 940.

(10) Nicolaou, K. C.; Smith, B. M.; Ajito, K.; Komatsu, H.; Gomez Paloma, L.; Tor, Y. *J. Am. Chem. Soc.* **1996**, *118*, 2303.

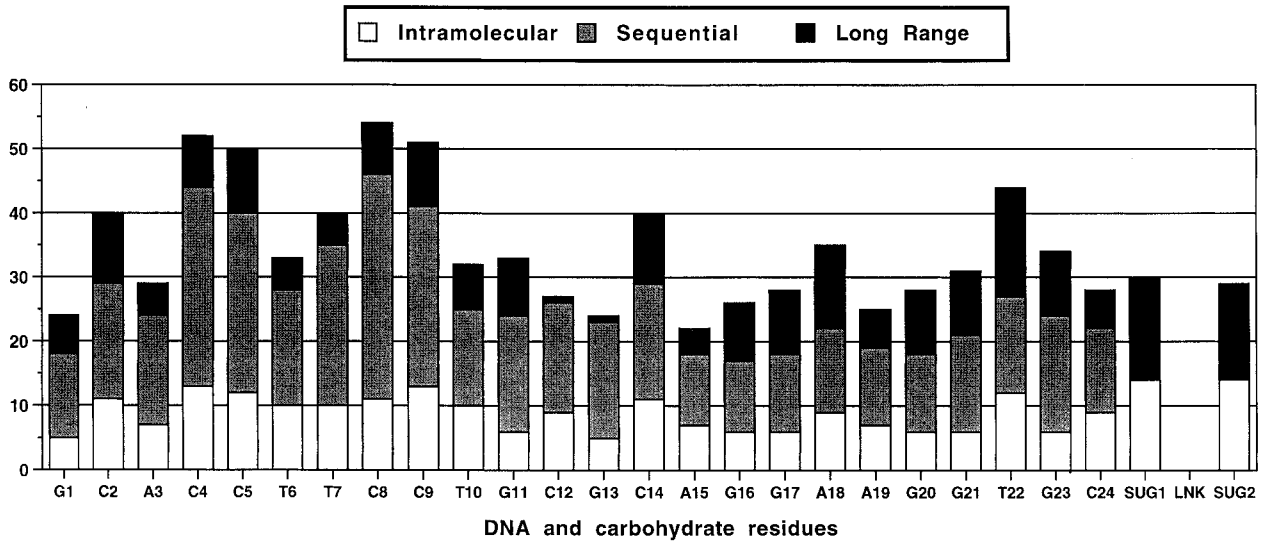


Figure 2. Plot of all the NMR-derived distance constraints used in the structure calculation of the HTD−DNA complex ordered by type.

(**4**) and duplex DNA (see Figure 1). This structure complements the previous study of a complex of the head-to-head dimer bound to a DNA duplex,[11] providing further insight into the molecular basis for the binding properties of these unique dimeric, carbohydrate-based DNA binding agents and into primary structure−function relationships for the calicheamicins. These results also contribute to a growing effort to understand the fundamental principles of biological recognition of carbohydrates.

Experimental Section

Synthesis of the Oligonucleotide. The two complementary oligonucleotides 5′-GCACCTTCCTGC-3′ and 5′-GCAGGAAGGTGC-3′ were synthesized on a Beckmann 200A automatic synthesizer using phosphoramidite chemistry. Both oligomers were purified by ion-exchange HPLC on a Partisil 10 SAX column, eluting with 20% aqueous CH_3CN and a linear gradient of KH_2PO_4 from 10 mM to 0.35 M at pH = 7.0. The oligos were then desalted by gel filtration on a Biogel P2 column.

Preparation of the Complex. The two complementary strands were mixed in water in equimolar amounts and annealed by heating at 90 °C for 5 min, followed by slow cooling to room temperature. The NMR sample was prepared by lyophilizing the duplex two times from 99.6% D_2O and then dissolving the lyophilized material in 500 μL of 10 mM phosphate buffer containing 10 mM NaCl and 0.1 mM EDTA at pH 7.0 in 99.996 D_2O. For experiments to assign the labile protons, the sample was lyophilized and redissolved in 90% H_2O/10% D_2O. The final concentration of the duplex was 1.5 mM.

The ligand−DNA complex was prepared according to the following procedure. Known amounts of the ligand dissolved in MeOH were added to 1 mL of a 0.75 mM solution of the duplex, stirred at 25 °C for 10 min, and then lyophilized and dissolved in 99.96% D_2O. The extent of the titration was monitored by examining the DNA 6H/8H resonances in the ^{1}H NMR spectrum. The final solution was lyophilized and redissolved in 500 μL of 99.996% D_2O. To examine the labile protons of the complex, the solution was lyophilized and redissolved in 500 μL of a 90% H_2O/10% D_2O solution.

NMR Experiments. All NMR experiments were performed on a Bruker AMX2-500 spectrometer. The temperature was 27 °C for both the free DNA duplex and the ligand−DNA complex. All spectra were acquired in the phase-sensitive mode. The transmitter was placed on the solvent resonance, and the TPPI method was used to achieve frequency discrimination in the ω_1 dimension.[12] The standard pulse sequence and phase cycling were used for 2Q[13] and PE-COSY[14] spectra. A total of 64 scans/t_1 value were acquired for the 2Q (t_{mix} = 30 ms, t_{1max} = 50 ms) and PE-COSY (t_{1max} = 80 ms). A TOCSY spectrum was acquired using the DIPSI-2 sequence[15] for spin locking with t_{mix} = 70 ms, 64 scans/t_1 and t_{1max} = 40 ms. NOESY[16] spectra from H_2O solutions were recorded with the last pulse replaced by a jump and return composite sequence.[17] NOESY spectra from D_2O solution were recorded with saturation of the residual HOD resonance during the preparation and mixing periods and a Hahn-echo to improve the quality of the baseline.[18] The NOESY spectra were acquired with mixing times of 50 and 200 ms, 64 scans/t_1, and t_{1max} = 50 ms.

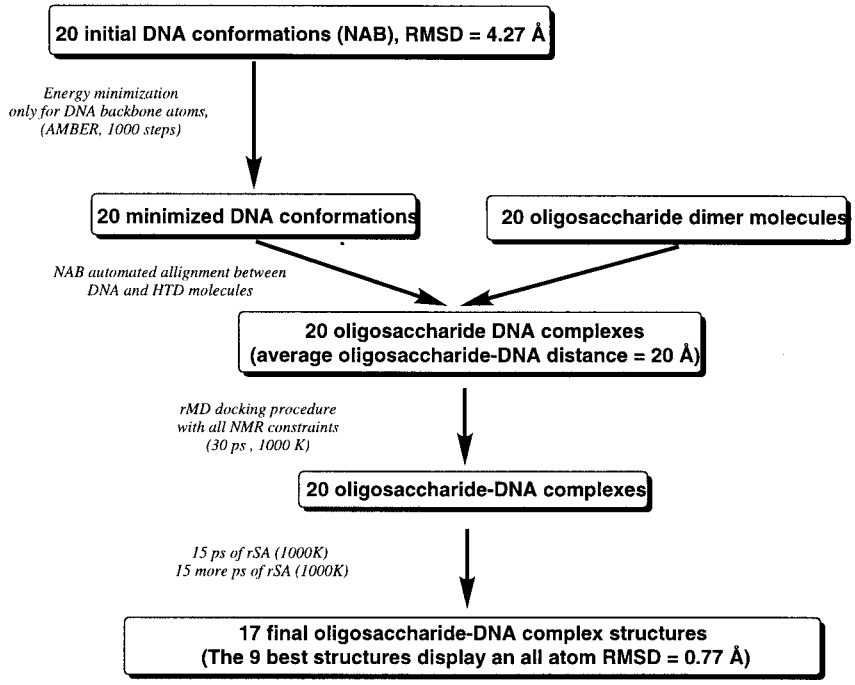
The NMR data were processed on a SGI INDIGO^2 workstation using FELIX 95 software (Biosym-MSI, San Diego, CA).

Restrained Molecular Dynamics Calculations. The structures of the complex were generated by a protocol involving several iterations through cycles of simulated annealing docking and refinement (Scheme 2). All restrained energy minimizations (rEM) and restrained molecular dynamics calculations (rMD) were performed on a Convex C-240 metacluster using the SANDER module of the AMBER 4.1 software package.[19] The potential functions for distance and dihedral constraints were flat between the given upper and lower bounds and rise parabolically outside of these bounds. These parabolic functions are turned smoothly into a linear function using a special feature of SANDER in order to avoid large violations. No explicit solvent molecules were included in these calculations, so a distance-dependent dielectric and reduced net charges on the phosphate oxygens (scaled by 0.2) were used to partially compensate for the absence of solvent. A distance cutoff of 9 Å was set for all nonbonded interactions. The force constant for both distance and dihedral constraints was set to 32 kcal/mol·Å^2.

The starting conformation of the calicheamicin oligosaccharide dimer was generated by linking two oligosaccharide molecules (derived from our structural model of the calicheamicin−DNA complex) through a saturated four carbon chain linker. Twenty different DNA starting conformations were constructed using Nucleic Acid Builder (NAB),[20] which allows generation of coordinates for nonstandard nucleic acid structures with a range of helical parameters. The initial geometries

(11) Bifulco, G.; Galeone, A.; Gomez-Paloma, L.; Nicolaou, K. C.; Chazin, W. J. *J. Am. Chem. Soc.* **1996**, *118*, 8817.
(12) Marion, D.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 500.
(13) Braunschweiler, L.; Bodenhausen, G.; Ernst, R. R. *Mol. Phys.* **1983**, *48*, 535.
(14) Mueller, L. *J. Magn. Reson.* **1987**, *72*, 191.
(15) Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *77*, 274.
(16) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95.
(17) Plateau, P.; Guéron, M. J. *J. Am. Chem. Soc.* **1982**, *104*, 7310.
(18) Davis, D. G. *J. Magn. Reson.* **1989**, *81*, 603.
(19) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 4.1, University of California, San Francisco, CA, 1994.
(20) (a) Macke, T. Ph.D. Thesis, The Scripps Research Institute, 1995.

Scheme 2. Protocol Used for the Calculation of the Solution Structure of the HTD−DNA Complex



were chosen with the stipulation of their being within the family of right-handed DNA helices. These structures sampled a range of conformations, exhibiting an RMSD over all atoms of 4.27 Å. The variation in their *x*-displacement, inclination, twist, and rise is listed in Table S1 (Supporting Information).

The docking procedure consisted of an "automated" alignment of the oligosaccharide and DNA molecules using NAB, followed by 30 ps of rMD docking, as described previously.[4e] The alignment between the oligosaccharide molecule with each DNA conformation was achieved using the following procedure: (a) six pairs of DNA and ligand atoms were chosen among the residues that were more likely to be involved in the binding (based on intermolecular NOEs), and (b) the two molecules were positioned so that all ligand atoms are separated from DNA atoms by at least 20 Å. The distance between ligand and DNA atoms was selected assuming that a short distance is more likely to generate van der Waals (VDW) clashes, while a larger value would have compromised the rate of convergence of the subsequent steps of rSA. After the two molecules were positioned, the complex was obtained by performing 30 ps of rSA (1000 K) with a very slow ramping up of the force constant of the penalty function for the experimental constraints (0.1−32 kcal/mol-Å^2). These docked structures were subjected to another 20 ps cycle of rSA (1000 K). These were then arranged in order of increasing experimental constraint violation energy, and the nine lowest violation energies were selected for detailed conformational analysis and display in Figure 3. The coordinates of these nine solution structures are being deposited with the nucleic acids section of the Brookhaven Protein Data Bank.

Figure 3. Ensemble of nine conformations representing the three-dimensional solution structure of the HTD−DNA complex. The DNA is displayed in light gray, the ligand subunits in black, and the linker between the two subunits of the ligand in dark gray.

Results and Discussion

^{1}H NMR. The general strategies for sequence-specific assignment of the ^{1}H resonances of small DNA duplexes have been amply reviewed,[21] and the specific protocols used here for the free DNA and ligand−DNA complex have been described.[22] The shorthand notation described by Wüthrich[21b] is used to specify DNA interproton distances and corresponding NOEs. For the free duplex, all cytosine 5H−6H and thymine 5Me−6H resonances were readily identified by scalar connectivities in 2Q, PE-COSY, and TOCSY spectra. All 24 discrete 1′H−2′H−2″H−3′H spin subsystems were identified in the 2Q, PE-COSY, and TOCSY spectra and were confirmed by inspection of the NOESY spectra. The sequential resonance assignments were obtained from the 50 ms mixing time NOESY

(21) (a) Wemmer, D. E.; Reis, B. R. *Annu. Rev. Phys. Chem.* **1985**, *36*, 105. (b) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986. (c) Patel, D. J.; Shapiro, L.; Hare, D. *Annu. Rev. Biophys. Chem.* **1987**, *16*, 423. (d) Reid, B. R. *Q. Rev. Biophys.* **1987**, *20*, 1.

(22) (a) Chazin, W. J.; Wüthrich, K.; Hyberts, S.; Rance, M.; Denny, W. A.; Leupin, W. *J. Mol. Biol.* **1986**, *190*, 439. (b) Chazin, W. J.; Rance, M.; Chollet, A.; Leupin, W. *Nucleic Acid Res.* **1991**, *19*, 5507. (c) Chen, S.; Leupin, W.; Rance, M.; Chazin, W. J. *Biochemistry* **1992**, *31*, 4406.

Table 1. ^{1}H Chemical Shifts in the HTD−d(GCACCTTCCTGC)·d(GCAGGAAGGTGC) Complex at pH = 7.0, T = 300 K

DNA Chemical Shifts[a] (ppm)

residue	N1H, N3H	NH2	2H, 5H, 5-Me	6H, 8H	1′H	2′H	2″H	3′H	4′	5′H/5″H
G1				7.74	5.95	2.53	2.72	4.98	4.31	
C2		6.37, 8.46	5.40	7.51	5.41	2.23	2.46	4.91	4.20	
A3			7.92	8.30	6.52	2.75	3.12	5.12	4.52	
C4		6.26, 8.05	5.19	7.05	6.06	1.85	2.89	4.93	4.15	3.63, 3.97
C5		6.49, 7.81	5.46	7.35	5.66	1.85	2.76	4.65	4.55	4.23
T6	13.30		1.59	7.49	6.29	1.91	2.73	4.57	4.21	
T7	13.88		1.59	7.32	6.25	2.11	2.88	4.99	4.28	
C8		6.57, 8.19	5.56	7.40	6.16	1.94	2.94	4.94	4.21	
C9		6.55, 8.00	5.49	7.38	5.50	2.00	2.64	4.59	4.47	3.73, 4.30
T10	13.28		1.61	7.44	5.94	1.95	2.47	4.68	4.05	
G11	12.87			7.93	6.05	2.47	2.79	5.02	4.41	
C12			5.53	7.48	6.21	2.20	2.21	4.55	4.08	
G13				7.74	5.95	2.46	2.72	5.07	4.22	
C14		6.48, 8.54	5.49	7.52	5.61	2.18	2.48	4.90	4.22	
A15			7.68	8.14	6.28	2.71	3.10	5.08	4.46	
G16	13.33			7.49	5.44	2.51	2.56	4.99	4.41	
G17	12.67			7.96	5.99	2.63	2.79	4.87	4.24	
A18			7.37	7.99	6.04	2.65	2.83	4.87	4.27	3.75, 3.72
A19			7.71	7.93	6.25	2.48	3.05	4.86	4.42	
G20	13.28			7.59	5.48	2.46	2.53	4.96	4.36	
G21	12.92			7.94	6.13	2.47	2.74	5.10	4.38	
T22	14.09		1.22	7.00	6.06	2.08	2.54	4.83	3.46	3.09, 3.59
G23	13.00	8.05		7.96	5.98	2.59	2.79	4.96	4.15	
C24			5.49	7.51	6.27	2.22	2.22	4.55	4.13	

Ligand Chemical Shifts[a] (ppm)

residue	H1	H2ax	H2eq	H3	H4	H5	Me	*N*-CH_2CH_3	OMe
A	4.31	3.32		4.41	2.46	3.97	1.54		
B	5.24	1.98	2.39	4.39	3.78	4.18	1.44		
C							2.43		3.87 (C5)−4.10 (C6)
D	5.19		4.68	4.09	3.78	4.39	1.50		3.64
E	5.36	1.69	2.80	3.96	3.30	4.25		3.18−1.39	3.63
A′	4.31	3.32		4.46	2.44	4.01	1.56		
B′	5.28	1.96	2.51	4.28	3.73	4.14	1.47		
C′							2.46		3.87 (C5)−4.10 (C6)
D′	5.35		4.65	4.04	3.67	4.26	1.44		3.64
E′	5.39	1.71	2.78	4.01	3.29	4.14		3.24−1.32	3.65

[a] The chemical shifts are referenced to 3-(trimethylsilyl)propionic-*2,2,3,3-d*$_4$ acid, sodium salt, using the HOD resonance previously calibrated in stock buffer solution.

spectrum. The combination of d_i(6,8;2′), d_s(2″;6,8), d_i(5Me;6), d_s(6,8;5Me) NOEs was sufficient to obtain complete sequence-specific resonance assignments. The relative intensities of characteristic NOE cross-peaks [e.g., d_i(6,8;2′), d_s(2″;6,8) ≫ d_i(6,8;2″), d_s(2′;6,8) and d_i(5Me;6) > d_s(6,8;5Me)] and the cross-peak patterns in the PE-COSY spectrum clearly indicated that the free duplex is a B-form duplex with C2′-endo-like sugar conformations.

Upon the binding of HTD (**4**) to the DNA, a small broadening of the line widths of all DNA and saccharide resonances was observed. The observation of one set of resonances over the full range of temperatures (10−37 °C) suggested a unique binding mode of the ligand to the DNA, in contrast to what was observed in the case of the HHD (**3**)−DNA complex, where two binding modes were detected.[11] The analysis of 2Q, TOCSY, and NOESY experiments enabled sequence-specific assignments to be made for the ligand and each DNA strand. The resonance assignments are provided in Table 1. The relative intensities of cross-peaks in the 50 ms NOESY indicated that in the presence of HTD (**4**) d(GCACCTTCCTGC)·d(GCAGGAAGGTGC) remains as a B-form duplex with $C_{2'}$-endo-like sugar conformations. However, a unique pattern of NOEs and scalar connectivities has been observed around the second cytosine of each recognition site (i.e., C_5 and C_9), suggesting that the deoxyribose ring conformation must be somewhat altered for these residues.

NMR-Derived Distance and Torsion Constraints. The chemical shift dispersion and resonance line widths of the HTD (**4**)−DNA complex were highly favorable relative to our experience with previously studied calicheamicin−DNA complexes. This facilitated both resonance assignment and accurate cross-peak volume measurements and ultimately led to a relatively large number of distance constraints. NOESY cross-peaks that could be unambiguously assigned for the first round of calculations were categorized into weak, medium, and strong on the basis of the cross-peak intensities and then converted into distance constraints with upper bounds of 5.0, 4.0, and 3.0 Å, respectively. The structures obtained using these constraints were used as input for hybrid relaxation matrix back-calculations (MARDIGRAS 5.1[23]) to set target distances more precisely and assign additional constraints. Cross-peak intensities used for these calculations were taken from three NOESY spectra τ_{mix} = 50 and 200 ms in D_2O and τ_{mix} = 200 ms in H_2O/D_2O (9:1). The target distances obtained from this hybrid relaxation matrix approach were adjusted by adding 10% of the distance to set the upper bound, to account for experimental and systematic errors. Note that the newer RANDMARDI[23b] protocol would obviate the need for this correction. Forty hydrogen-bonding restraints were also included on the basis of characteristic imino

(23) (a) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1990**, *87*, 475. (b) Liu, H.; Spielmann, H. P.; Ulyanov, N. B.; Wemmer, D. E.; James, T. L. *J. Biomol. NMR* **1995**, *6*, 390.

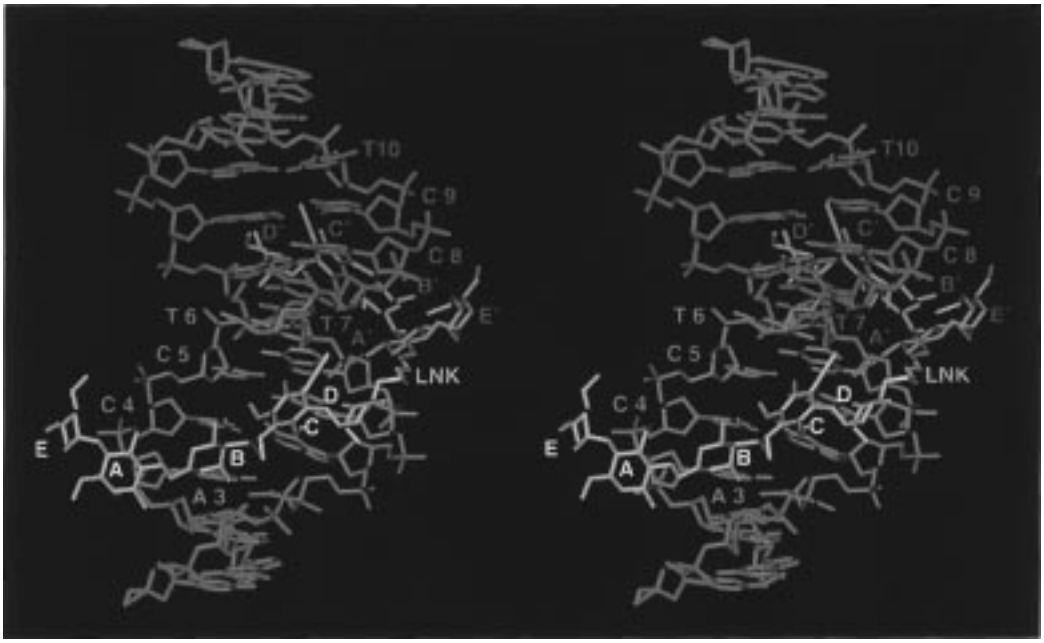


Figure 4. Stereoview of the HTD−DNA complex showing the details of the ligand and DNA binding site. The two binding sites are shown in green and red with the rest of the DNA in blue, and the polypyrimidine strand ($A_3CCTTCCT_{10}$) labeled. The two oligosaccharide subunits are shown in yellow and orange with the linker between the two subunits in white.

and amino ^{1}H resonances in the appropriate regions of the NMR spectrum. These were enforced for the G/C base pairs by constraining the N4−O6 distance to 3.01, H42−O6 to 2.35 Å, N1−N3 to 3.05 Å, and H1−N3 to 2.35. For the A/T base pairs, the N1−N3 distance was constrained to 2.92 Å and N1−H3 to 2.22 Å. A total of 562 proton−proton distance constraints were used for the final round of structure calculations. These are listed in Table S2 (Supporting Information), and their distribution is shown in Figure 2.

Dihedral angle constraints for the deoxyribose sugar rings were obtained by measuring $J_{1'-2'}$ and $J_{1'-2''}$ coupling constants in the PE-COSY spectrum and estimating the magnitude of $J_{2'-3'}$ in the 2Q spectrum. The coupling constant values were converted into PPA (pseudorotational phase angles), which in turn were used to constrain the angles ν_0, ν_1, ν_2, ν_3, and ν_4. The PPA constraint values ranged between 90° and 190°, as compared to ca. 160° for standard B-DNA geometry ($C_{2'}$-endo), ca. 20° for standard A-DNA geometry ($C_{3'}$-endo), and 120°−160° for most experimentally determined structures of free duplexes. Additional loose dihedral angle constraints on the backbone angles α, γ, ϵ, and ζ were determined using the approach described by Reid and co-workers.[24] Typical constraint ranges were 240°−360°, 20°−100°, 120°−200°, and 240°−360°, respectively. In all, 114 torsion angle constraints were assigned, bringing the total of NMR-derived experimental constraints to 676 (~27 constraints per residue, counting the ligand as one residue).

Solution Structure of the HTD−DNA Complex. Direct evidence for the nature of the binding of the oligosaccharide (**4**) to DNA was first obtained from the intermolecular NOESY cross-peaks (Table 2). The localization of HTD (**4**) in the minor groove of the DNA duplex is evident in Figures 3 and 4. Figure 3 displays the ensemble of nine structures used to represent the HTD (**4**)−DNA complex, and structural statistics for this ensemble are provided in Table 3. The high quality of these structures is reflected in the small numbers of violations of the experimental constraints, the small magnitude of these residual violations, and the large negative molecular energies in the AMBER force field. The central binding site is well defined with an all-atom RMSD from the mean structure of 0.77 Å. The ends of the DNA are less well defined, as are the

Table 2. Intermolecular NOEs Observed at 300 K for the 1:1 Complex of d(GCACCTTCCTGC)·d(GCAGGAAGGTGC) and HTD (**4**)

proton-1 (HTD)	proton-2 (DNA)	size[a]
A-Me	**C**24-4′H	s
B-Me	**C**5-4′H	s
B-H1	**G**23-4′H	m
B-H3	**T**22-3H	s
C-2-Me	**C**5-1′H	s
C-6-OMe	**T**22-4′H	s
C-6-OMe	**T**22-5′H	m
C-6-OMe	**T**22-5″H	m
D-H1	**T**22-4′H	s
D-H1	**T**22-5′H	s
D-H1	**T**22-5″H	s
D-H2	**T**22-4′H	s
D-H2	**T**22-5′H	m
D-H2	**T**22-5″H	m
D-OMe	**G**21-4′H	m
D-OMe	**G**20-2H	w
A′-Me	**G**20-4′H	m
B′-Me	**C**9-4′H	w
B′-H1	**A**19-4′H	s
B′-H1	**A**19-5′H	s
B′-H2′	**A**18-2H	s
B′-H2″	**A**18-2H	m
C′-2-Me	**C**9-1′H	m
C′-6-OMe	**A**18-4′H	w
C′-6-OMe	**A**18-5′H	w
D′-H1	**A**18-5′H	s
D′-H1	**A**18-5″H	w
D′-H2	**A**18-5′H	m
D′-H2	**A**18-5″H	s
D′-OMe	**G**17-4′H	m
D′-OMe	**G**16-2H	w

[a] w = weak (5.0 Å), m = medium (4.0 Å), s = strong (3.0 Å). This semiquantitative categorization is based on the volume of the NOESY cross-peaks (τ_{mix} = 50 and 200 ms) and was applied only for intermolecular NOEs.

(24) Kim, S. G.; Lin, L. J.; Reid, B. R. *Biochemistry* **1992**, *31*, 3564.

Table 3. Summary of the Molecular Energies, RMSDs, and Residual Violations of the Ensemble of the 9 Final Structures of the 1:1 Complex of HTD (**4**) and d(GCACCTTCCTGC)·d(GCAGGAAGGTGC)

molecular energies (kcal)	
E_{amber}	-926.1 ± 12.5
E_{viol}	8.9 ± 1.7
distance violations (Å)	
$0.01 < d < 0.20$	23.9 ± 2.1
$0.20 < d$	0.4 ± 0.7
dihedral angle violations (deg)	
$2 < \theta < 6$	3.56 ± 1.26
$6 < \theta < 10$	0.22 ± 0.44
average rmsd, all heavy atoms in the binding site (Å)	
HTD + DNA (pairwise)	1.16
HTD + DNA (from the mean)	0.77

Table 4. Intermocular H Bonds[a] Observed in the Solution Structure of the 1:1 Complex of HTD (**4**) and d(GCACCTTCCTGC)·d(GCAGGAAGGTGC)

donor	acceptor
B-3O*H*	$\mathbf{T_{22}}$-*O*2
$\mathbf{G_{21}}$-NH_2	**C**-SC=*O*
D-2O*H*	$\mathbf{T_{22}}$-*O*1P
B′-3O*H*	$\mathbf{A_{18}}$-*N*3
$\mathbf{G_{17}}$-NH_2	**C′**-SC=*O*
D′-2O*H*	$\mathbf{A_{18}}$-*O*1P

[a] The H bonds are defined on the basis of a distance cutoff of 3.0 Å.

oligosaccharide residues E and E′, which face out away from the DNA and are highly exposed to solvent.

To facilitate the discussion of the structure of the complex at the atomic level, a tube diagram of the representative structure (that closest to the mean) is shown in Figure 4 with color coding and labeling of the binding site. The binding of the HTD (**4**) molecule in the DNA minor groove is stabilized by the existence of many favorable intermolecular interactions. Among these are six hydrogen bonds (Table 4), two salt bridges between the positively charged nitrogen atoms of rings E and E′ and the O1P oxygen atoms of residues C_4 and C_8, respectively, and a large buried intermolecular surface with extensive van der Waals contacts. In addition, each of the iodine atoms of HTD (**4**) exhibits a hydrogen-bonding-like interaction with a guanine NH_2 in the binding site as observed in several other studies on DNA complexes of calicheamicin-based ligands.[4,11]

A detailed analysis of the DNA geometry was obtained using NEWHELIX software.[25] The helicoidal parameters, sugar pucker, and backbone torsion angles indicate that the oligonucleotide occupies a conformation in the B-DNA family. By way of example, the mean values and standard deviation of *x*-displacement, inclination, twist, and rise are provided in Table 5. Some small local perturbations from typical B-DNA structure are observed, with the most significant being a shift in the deoxyribose ring conformation of the second cytosine (in direction 5′ → 3′) of each binding site (i.e., C_5 and C_9) toward $O_{4'}$-endo sugar pucker (PPA = 90°). This shift appears to be due to the high steric demand of the HTD iodine atom in ring C. A similar deoxyribose ring pucker for the corresponding DNA residue has been reported in other calicheamicin−DNA complexes.[4] Although the experimental observables have been interpreted in terms of a perturbed sugar ring conformation, it should be noted that a similar structural result could be obtained if the sugar rings occupied an equilibrium distribution of $C_{2'}$-endo and $C_{3'}$-endo conformations. In fact, the currently available information is insufficient to distinguish unambiguously between the static and dynamic interpretations. However, the small difference in energy between $O_{4'}$-endo and $C_{2'}$-endo and $C_{3'}$-endo low-energy sugar ring conformations (2 kcal/mol or less)[26] suggests that the shift to the higher energy PPA could be readily accommodated.

(25) Dickerson, R. E. *NEWHELIX Release 1/20/93*, Molecular Biology Institute, UCLA, 1993.

Table 5. Mean and Standard Deviation of a Selection of Representative Helical Parameters of the Ensemble of the 9 Final Structures of the 1:1 Complex of HTD (**4**) and d(GCACCTTCCTGC)·d(GCAGGAAGGTGC)

base pair	X displacement	inclination	twist	rise
1	-0.12 ± 0.53	12.54 ± 4.60	9.08 ± 5.70	7.43 ± 0.46
2	-0.70 ± 0.51	14.52 ± 3.10	38.77 ± 2.59	2.78 ± 0.19
3	-1.16 ± 0.33	15.13 ± 2.60	32.94 ± 2.12	2.95 ± 0.20
4	-1.19 ± 0.25	15.47 ± 2.69	39.04 ± 1.52	2.76 ± 0.20
5	-1.41 ± 0.21	14.42 ± 2.06	36.34 ± 2.31	3.08 ± 0.19
6	-1.57 ± 0.35	12.94 ± 1.55	33.81 ± 3.71	3.24 ± 0.17
7	-1.67 ± 0.33	13.23 ± 1.47	36.24 ± 3.30	3.21 ± 0.17
8	-1.47 ± 0.39	13.44 ± 1.23	39.30 ± 2.28	2.76 ± 0.20
9	-1.31 ± 0.53	12.34 ± 1.90	39.21 ± 4.54	2.77 ± 0.21
10	-1.53 ± 0.49	10.16 ± 2.86	34.64 ± 1.34	2.98 ± 0.18
11	-1.38 ± 0.38	7.49 ± 4.30	35.41 ± 2.20	3.16 ± 0.25
12	-0.50 ± 0.43	7.77 ± 5.07		
A-DNA	−3.5	19.0	33	2.6
B-DNA	0.0	2.0	36	3.4

Comparative Analysis of Calicheamicin Oligosaccharide Binding in the TCCT and ACCT Recognition Sites. The TCCT and ACCT sites to which the two subunits of the HTD (**4**) dimer are bound correspond to two of the four high affinity DNA sequences for calicheamicins; hence, it is clearly of interest to compare the structures of the two sites. We find that the intermolecular interactions within each of the dimer subunits are very similar. This similarity even extends to the position where the sequences are different: a hydrogen bond is formed from the carbohydrate B-3OH of each subunit, either to the thymine O2 in the ACCT site (5′-ACCT-3′/5′-AGG$\mathbf{T_{22}}$-3′) or the adenine N3 in the TCCT site (5′-TCCT-3′/5′-AGG$\mathbf{A_{16}}$-3′). These homologous interactions are displayed in Figure 5, where the similarity in the position of the T_{22}−O2 and A_{18}−N3 is clearly evident. Thus, both AT and TA base pairs can be bound interchangeably at the 5′ end of the binding site because both have a hydrogen-bond acceptor (the carbonyl O2 of thymine or the ring N3 of adenine) protruding into the center of the DNA minor groove.

The conformations of the two subunits of the HTD (**4**) dimer and their positioning within their DNA binding sites are also very similar, even though the sites are different (Figure 4). In the ACCT site, the HTD A and B rings are proximate to the A_3/T_{22} and C_4/G_{21} base pairs, and the amino sugar E is outside of the minor groove. The hexasubstituted aromatic ring C occupies a position between the two guanine residues G_{21} and G_{20} and appears to be closer to the purine than to the pyrimidine strand. Part of sugar residue D is actually outside the DNA minor groove with C-1, C-2 and the ring oxygen in close proximity to the phosphate backbone of the purine strand. The positioning of the ligand in the TCCT site, and the oligosaccharide−DNA interactions for the carbohydrate unit, are virtually identical to the ACCT site and are also similar to that observed in previous studies conducted on complexes between calicheamicin-based monomers and TCCT-containing oligonucleotides.[4]

Implications for Sequence Selectivity of the Calicheamicin Oligosaccharide. This work brings deeper insight into the

(26) (a) Olson, W. K.; Sussman, J. L. *J. Am. Chem. Soc.* **1982**, *104*, 270. (b) Olson, W. K. *J. Am. Chem. Soc.* **1982**, *104*, 278.

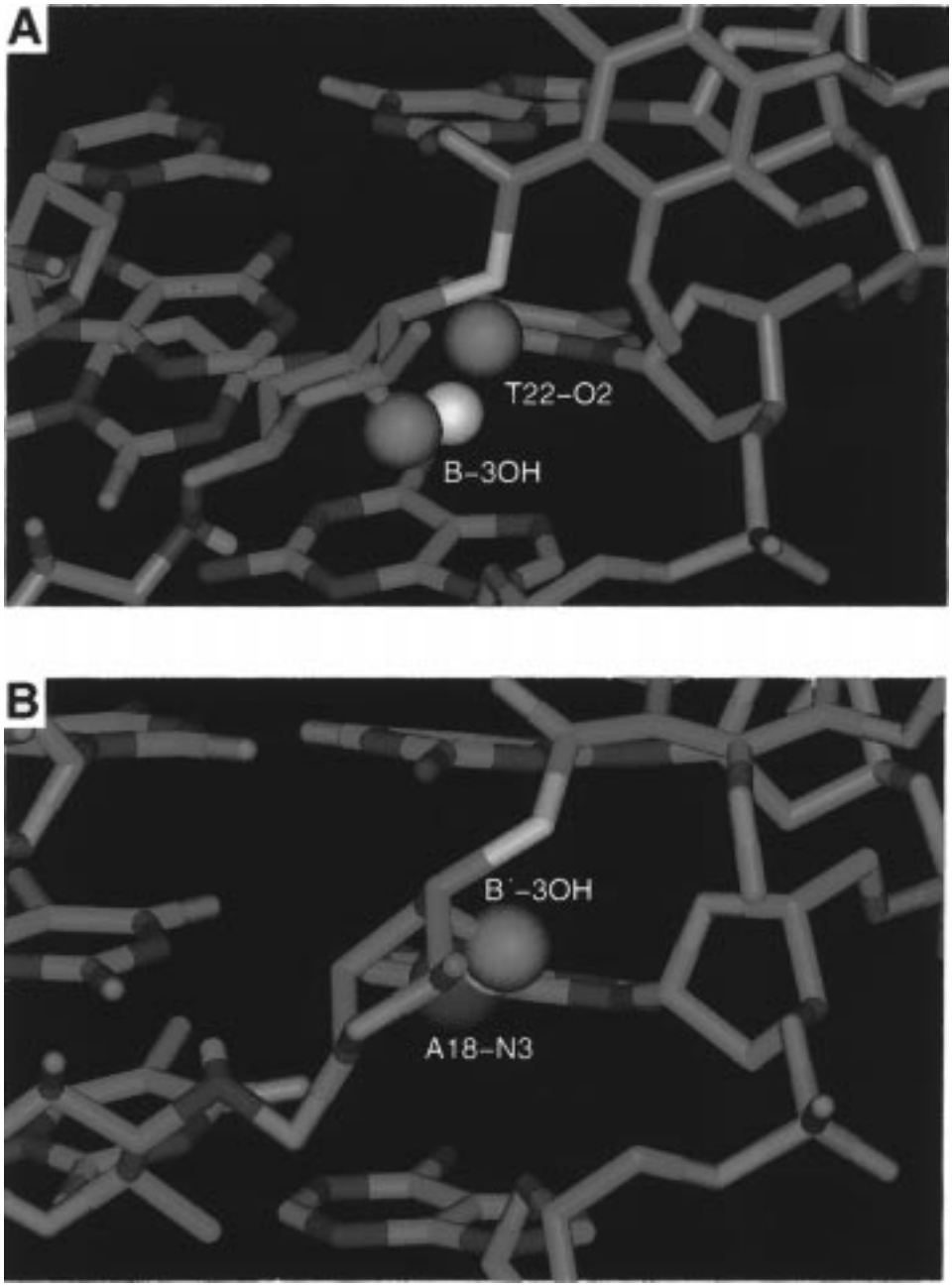


Figure 5. Close-up view of the homologous hydrogen-bonding interactions of HTD to different DNA functional groups in the HTD−DNA complex. The hydrogen bonds between (A) T_{22}−O2 and the 3OH of ring B and (B) between A_{18}−N3 and the 3OH of ring B′ are highlighted with spheres colored red for oxygen, blue for nitrogen, and gray for hydrogen.

factors governing the DNA binding affinity of calicheamicin and its derivatives. The comparison of the structures and the carbohydrate−DNA interactions at the two different binding sites explains at the atomic level how calicheamicin derivatives are able to effectively recognize both d(ACCT) and d(TCCT) sites. A number of specific hydrogen-bonding and salt bridge interactions have been identified for both recognition sites. Three sequence-specific intermolecular interactions are present for each DNA site: (a) the H-bond between the carbonyl group connecting carbohydrate B and C rings and the G_{17}−NH_2 (and correspondingly, G_{21}−NH_2 in the other site); (b) the H-bond between B−3OH and A_{18}−N3 (and T_{22}−O2); and (c) the interaction between the C ring iodine atom and a guanine amino group G_{16}−NH_2 (and G_{20}−NH_2), whose importance in the binding affinity of these carbohydrates has been well established.[8,27] Thus, intermolecular interactions appear to play an important role in the sequence preferences of DNA binding observed for these compounds.

Structural distortions away from standard geometry are expected in the binding of any strucurally complex ligand, as the two molecules optimize their structures to maximize interaction energy. The conformational adaptability of duplex DNA is clearly evident from the wide range of perturbations observed in structures of complexes with proteins and natural product ligands.[28] The structural perturbations of the d(GCACCTTCCTGC)·d(GCAGGAAGGTGC) duplex in the HTD (**4**) complex are quite modest overall, with the most extreme effect at C_5 and C_9, whose sugar ring pseudorotation angles shift on the order of 50°. The energy landscape of the

(27) Bailly, C.; Waring, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 7311.

sugar ring conformations found in duplex DNA is fairly shallow, with the $O_{4'}$-endo maximum only 1.5−2 kcal/mol less stable than the $C_{2'}$-endo and $C_{3'}$-endo minima.[26] Thus, the energetic penalty to pay for C_5 and C_9 occupying a less favorable conformation is small and is outweighed by the favorable hydrogen bond, salt bridge, and van der Waals interactions. These are important considerations for understanding the factors directing the propensity for calicheamicins to bind certain DNA sites.

The results reported here and in other structural studies have highlighted the importance of specific and nonspecific intermolecular interactions for the binding of the calicheamicin oligosaccharide in the DNA minor groove. In addition, some degree of induced fit of the DNA site to ligand (and possibly of the ligand to the DNA) is likely. Kahne and co-workers have proposed that DNA conformational adaptability is a primary factor determining the DNA sequence selectivity of calicheamicin.[4a,c] In this view, binding to the recognition site will be facilitated in regions possessing greater intrinsic conformational flexibility. Unfortunately, there is little if any direct experimental evidence available on the sequence-dependence of conformational propensities and relative flexibility of duplex DNA. Furthermore, in the structure reported here, we observe no substantial distortions from B-DNA geometry, so the requirement for induced fit or indirect read-out appears to be minimal. This is true, even though the two ligands are constrained by their covalent linkage! Thus, we do not interpret the relatively small (50°) shift in sugar pucker as supportive of the "indirect readout" hypothesis. In contrast, the great similarity in the intermolecular interactions in the two binding sites of the HTD complex suggests that sequence selectivity *might* be due to the specific interactions not being available in the requisite locations in DNA duplexes with nonoptimal sequences. Solvation effects may also play an important role.[29] Further structural studies of the other high-affinity DNA binding sites and molecular modeling of lower affinity sites are in progress to test this hypothesis and ultimately establish the relative importance of the factors directing the propensity of the calicheamicin oligosaccharide to bind to duplex DNA.

(28) Neidle, S. DNA Structure and Recognition. In *Focus*; Rickwood, D., Ed.; Oxford University Press Inc.: New York, 1994.

(29) Misra, V. K.; Sharp, K. A.; Friedman, R. A.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 245.

Conclusions

The NMR solution structure has been determined for the 1:1 complex between the head-to-tail dimer of the calicheamicin oligosaccharide and the duplex oligonucleotide d(GCACCTTCCTGC)·d(GCAGGAAGGTGC) that contains two different DNA binding sites (ACCT and TCCT). The nanomolar affinity associated with the bidentate mode of binding of the HTD dimer to the DNA duplex combined with the snug fit of the ligand in the DNA minor groove strongly suggest that the HTD molecule possesses a very high degree of complementarity to the duplex containing the dual d(ACCTTCCT) binding site. The ability of each of the oligosaccharide subunits to assume correct positioning for effective binding to the DNA sequence is in turn attributed to the favorable attributes of the linker. Thus, the flexible, open-chain linkers of the HHD (**3**) and HTD (**4**) dimers appear to enable effective localization of each subunit into the respective binding site. On the basis of these results, an effort is currently underway to design corresponding ligands with rigid linkers that, in addition to ensuring correct positioning of each unit, bear specific structural elements for further stabilization of the ligand in the DNA minor groove.

Acknowledgment. L.G.P. thanks the "Programma per la breve mobilità di Professori e Ricercatori" sponsored by the University of Naples "Federico II" (Italy) for financial support during his 2-month leave at The Scripps Research Institute. Support for this research was provided to K.C.N. by National Institutes of Health and The Scripps Research Institute and to W.J.C. by the American Cancer Society (Grant Nos. DHP-123, FRA-436). We thank Sudip Parikh, Melanie Nelson, and Dr. Randy Ketchem for assistance in the preparation of figures.

Supporting Information Available: Three figures showing expanded regions of NOESY spectra of the HTD−DNA complex (Figures S1−S3). Two tables listing helical parameters used to generate initial DNA conformations (Table S1) and all NMR derived distance constraints used in the structure calculation (Table S2) (16 pages, print/PDF). See any current masthead page for ordering information and Web access instructions.

JA973910Y