Molecular Basis for the DNA Sequence Selectivity of Ecteinascidin 736 and 743: Evidence for the Dominant Role of Direct Readout via Hydrogen Bonding

Frederick C. Seaman* and Laurence H. Hurley*

Contribution from the Drug Dynamics Institute, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712

Received August 27, 1998

Abstract: The marine natural product ecteinascidin 743 (Et 743) is currently in phase II clinical trials. We have undertaken parallel structural and modeling studies of an Et 743−(N2-guanine) 12-mer DNA adduct and an adduct involving the structurally related Et 736 of the same sequence in order to ascertain the structural basis for the ecteinascidin−DNA sequence selectivity. In contrast to the C-subunit differences found in Et 736 and Et 743, they have identical A−B-subunit scaffolds, which are the principal sites of interaction with DNA bases. These identical scaffolds generate parallel networks of drug−DNA hydrogen bonds that associate the drugs with the three base pairs at the recognition site. We propose that these parallel hydrogen bonding networks stabilize the Et 736 and Et 743 A- and B-subunit prealkylation binding complex with the three base pairs and are the major factors governing sequence recognition and reactivity. The possibility that a unique hydrogen-bonding network directs the course of sequence recognition was examined by first characterizing the hydrogen-bonding substituents using ^{1}H NMR properties of the exchangeable protons attached to the hydrogen-bond donor and other protons near the proposed acceptor. Using these experimental findings as indicators of hydrogen bonding, Et 736−12-mer duplex adduct models (binding and covalent forms) containing the favored sequences 5′-AGC and 5′-CGG were examined by molecular dynamics (MD) in order to evaluate the stability of the hydrogen bonds in the resulting conformations. The MD-generated models of these favored sequences display optimal donor/acceptor positions for maximizing the number of drug−DNA hydrogen bonds prior to covalent reaction. The results of MD analysis of the carbinolamine (binding) forms of the sequences 5′-GGG (moderately reactive) and 5′-AGT (poorly reactive) suggested reasons for their diminished hydrogen-bonding capability. These experimental and modeling results provide the structural basis for the following sequence specificity rules: For the target sequence 5′-XGY, the favored base to the 3′-side, Y, is either G or C. When Y is G, then a pyrimidine base (T or C) is favored for X. When Y is C, a purine (A or G) is favored for X.

Introduction

Ecteinascidin 743 is one of a series of structurally related antitumor compounds isolated from *Ecteinascidia turbinata* (Ete).[1] This isoquinoline alkaloid has completed phase I clinical trials.[2−5] The biosynthesis of this antitumor agent incorporates three tetrahydroisoquinoline precursors into the A, B, and C subunits (Figure 1). Et 743 differs from most other ecteinascidins and related alkaloids by the structure of its C subunit, which is attached to the rigid bis(tetrahydroisoquinoline) A−B-subunit "scaffold" via a flexible 10-membered lactonic ring. For example, the Et 743 tetrahydroisoquinoline C subunit is replaced in Et 736 (Figure 1) and its N12 demethyl analogue, Et 722,[6] by a tetrahydro-β-carboline C-subunit. Et 736 and Et 722 also differ from Et 743 in that the former agents show a higher level of activity in vivo in mice against P388 leukemia, while the latter drug shows higher activity against B16 melanoma, Lewis lung carcinoma, M5076 ovarian sarcoma, and MX1 human mammary carcinoma xenograph.[6] Et 743 and Et 736 were proposed to react with DNA via nucleophilic attack by the exocyclic amino group of guanine.[7] Two-dimensional ^{1}H NMR analysis of the Et 743 guanine−DNA adduct of the 12-mer duplex containing the target sequence 5′-AGC (underline indicates site of covalent modification) confirmed the covalent linkage site[8] and the previously predicted drug orientation in the minor groove[6] with the methylenedioxy moiety directed toward the 3′-side and the A subunit positioned to the 5′-side

* Address correspondence to either author. Tel: (512) 471-4841. Fax: (512) 471-2746. E-mail: dg-dna@mail.utexas.edu.

(1) (a) Rinehart, K. L.; Holt, T. G.; Fregeau, N. L.; Stroh, J. G.; Keifer, P. A.; Sun, F.; Li, L. H.; Martin, D. G. *J. Org. Chem.* **1990**, *55*, 4512−4515. (b) Sakai, R.; Jares-Erijman, E. A.; Manzanares, I.; Elipe, M. V. S.; Rinehart, K. L. *J. Am. Chem. Soc.* **1996**, *118*, 9017−9023.

(2) Jimeno, J. M.; Faircloth, G.; Cameron, L.; Meely. K.; Vega, E.; Gómez, A.; Fernandez Sousa-Faro, J. Ma.; Rinehart, K. *Drugs Future* **1996**, *21*, 1155−1165.

(3) Bowman, A.; Twelves, C.; Hoekman, K.; Simpson, A.; Smyth, J.; Vermorken, J.; Höppener, F.; Beijnen, J.; Vega, E.; Jimeno, J.; Hanauske, A.-R. *Ann. Oncol.* **1998**, *9* (Suppl. 2), 119.

(4) Villalona-Calero, M.; Eckhardt, S. G.; Weiss, G.; Campbell, E.; Hidalgo, M.; Kraynak, M.; Beijnen, J.; Jimeno, J.; Von Hoff, D.; Rowinsky, E. *Ann. Oncol.* **1998**, *9* (Suppl. 2), 118.

(5) Cvitkovic, E.; Mekranter, B.; Taamma, A.; Goldwasser, F.; Beijnen, J. H.; Jimeno, J.; Riofrio, M.; Vega, E.; Misset, J. L.; Hop, P. *Ann. Oncol.* **1998**, *9* (Suppl. 2), 119.

(6) Sakai, R.; Rinehart, K. L.; Guan, Y.; Wang, H.-J. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 114656−11460.

(7) Pommier, Y.; Kohlhagen, G.; Bailly, C.; Waring, M.; Mazumder, A.; Kohn, K. W. *Biochemistry* **1996**, *35*, 13303−13309.

(8) Moore, B. M., II; Seaman, F. C.; Hurley, L. H. *J. Am. Chem. Soc.* **1997**, *119*, 5475−5476. On page 5476 in this article, the authors incorrectly identified the HB1 and HB3 acceptors as O2 and O1′, respectively, of 21T instead of 20T.

10.1021/ja983091x CCC: $15.00
Published on Web 12/03/1998

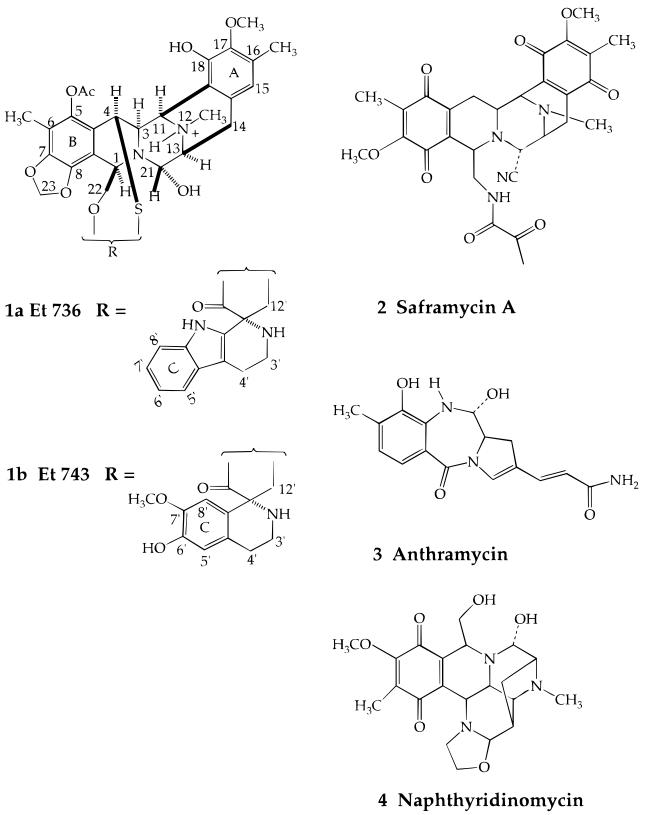

Figure 1. Structures of Et 736 (**1a**) and Et 743 (**1b**), saframycin A (**2**), anthramycin (**3**), and naphthyridinomycin (**4**).

of the attachment point. On the basis of the protonated state of the Et 743−DNA adduct, a reaction mechanism was proposed by which the carbinolamine form of the drug (**5** in Figure 2) underwent dehydration in the minor groove to give an iminium ion (**6** in Figure 2) suitable for nucleophilic attack by the amino group of an adjacent guanine base.[9]

Two ecteinascidin studies yielded important findings relating to sequence specificity. In the first, on the basis of modeling of Et 743− and Et 729−(N2-guanine) DNA adducts, a hydrogen-bonding network linking the drug to a four-base-pair region was proposed.[10] The second study demonstrated the importance of a three-base sequence (central guanine and two flanking bases) in determining the Et 743 sequence recognition and the resulting reactivity of the central guanine.[7] As pointed out by Pommier and co-workers in their gel electrophoresis study and supported by our preliminary NMR studies, the favored base to the 3′-side of the target guanine must be either G or C. Our results indicate that when this 3′-side base is guanine, a pyrimidine base (C or T) is then favored for the base to the 5′-side of the target guanine. When this 3′-side base is cytosine and its pairing partner on the opposite strand is guanine, the favored base to the 5′-side of the target guanine is a purine (G or A). These findings lead to the proposal that a hydrogen-bonding network involving specific DNA base donors/acceptors directs the course of sequence recognition. This proposal serves as the focus for our experimental and modeling analysis.

Our first goal was to gather experimental evidence for hydrogen bonding associating DNA with the covalently bound drug. The hydrogen-bonding behavior of this complex was characterized using the exchangeable ^{1}H NMR data from the Et 736−DNA adduct and by comparing these data to the less complete data of the Et 743 duplex adduct. Comparison of exchangeable and nonexchangeble ^{1}H NMR data for the Et 736− and Et 743−DNA adducts of identical 12-mer duplexes revealed very similar chemical shifts and NOESY cross-peak patterns for protons in the region of the A−B-subunit scaffold's contact with DNA. This suggests identical hydrogen-bonding networks for the two DNA adducts. On the basis of these experimental findings, the second goal was to ascertain the degree to which this hydrogen-bonding potential was sequence dependent. Hydrogen-bonding stability was evaluated using MD analyses of Et 736−DNA duplex models of two highly reactive sequences, 5′-AGC and 5′-CGG, one moderately reactive sequence, 5′-GGG, and one poorly reactive sequence 5′-AGT (Figure 3). MD analyses (both solvated and in vacuo) were conducted for the Et 736 covalent adduct form and the carbinolamine binding form associated with the highly reactive sequences embedded in 12-mer duplexes. The objective of this sequence comparison was to ascertain if the two different favored (i.e., highly reactive) base sequences[7] possess structures that optimize hydrogen-bonding association in comparison to other nonfavored sequences.

The MD results for the Et 736 binding form yielded insight into those structural properties that determine the range of reactivity. A stable hydrogen-bonding network was identified for the sequence 5′-AGC from both experimental findings and modeling. Similar network stability of the alternative favored 5′-CGG sequence was shown by modeling. However, these favored sequences appear to represent two different optimal binding modes. Modeling of the 5′-GGG sequence suggests that hydrogen bonding between the drug N12H donor and O2 of the cytosine paired with the guanine to the 5′-side of the target guanine occurs less readily than the comparable hydrogen bond in the favored sequences. This is expressed in the MD study by a lower frequency of hydrogen bonding and suboptimal hydrogen-bonding distances. The postulated instability of the 5′-AGT drug binding complex was reflected by the erratic behavior witnessed during MD trajectories for this drug positioned near the target guanine in the minor groove. All of these observations buttress the view that the different hydrogen bonding profiles of the triplet sequences determine their relative drug binding stability and reactivity.

Results

Parallel ^{1}H NMR studies were conducted on Et 736− and Et 743−(N2-guanine) DNA adducts using the same self-complementary 12-mer duplex with the favored sequence, 5′-AGC (Figure 3), containing a central target guanine. Aside from C-subunit-related conformational differences (resulting mostly from the lack in Et 736 of the Et 743 C subunit's DNA backbone hydrogen-bonding potential), our data demonstrated identical patterns of complex hydrogen bonding in the minor groove between DNA and the A−B-subunit scaffold. We propose that these parallel hydrogen-bonding networks linking the common Et 736 and Et 743 A and B subunits to a three-base-pair region are the major factors governing the similar sequence recognition and reactivity behavior.

^{1}H NMR Data Demonstrate a 3′-Offset of NOE Contacts on Both the Covalently Modified and Unmodified Duplex Strands Extending in Both Instances from the Site of Drug Attachment. All of the nonexchangeble ^{1}H NMR signals and virtually all of the exchangeable ^{1}H NMR signals of the Et 736−12-mer duplex DNA adduct were assigned (Table 1). The D_2O and H_2O/D_2O NOESY and COSY cross-peaks generated by

(9) Moore, B. M., II; Seaman, F. C.; Wheelhouse, R. T.; Hurley, L. H. *J. Am. Chem. Soc.* **1998**, *120*, 2490−2491.

(10) Guan, Y.; Sakai, R.; Rinehart, K. L.; Wang, A. H.-J. *J. Biomol. Struct. Dyn.* **1993**, *10*, 793−818.

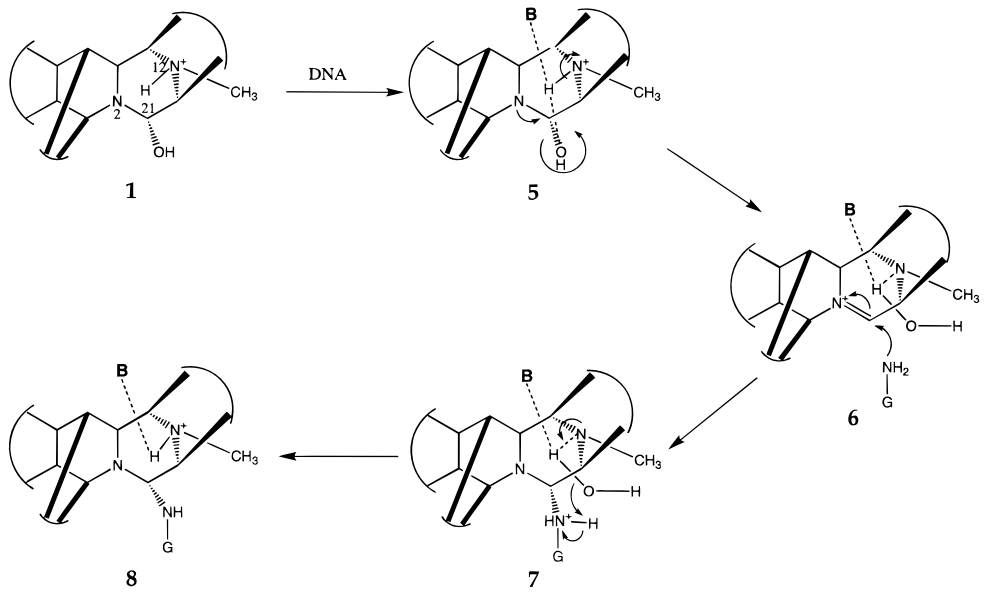

Figure 2. Proposed mechanism for the catalytic activation of the ecteinascidin carbinolamine prior to covalent bond formation with N2 of guanine. The 12NH of the carbinolamine (**5**) catalyzes the dehydration of C21, yielding the iminium ion (**6**). Nucleophilic attack by GN2 on **6** results in the expulsion of the transiently bound H_2O, which contains the proton released by guanine N2 (**7**). The resulting adduct (**8**) retains a protonated N12 (adapted from ref 9). Hydrogen bonds are depicted by dotted lines, and **B** corresponds to a DNA base hydrogen bond acceptor (e.g., thymine O2).

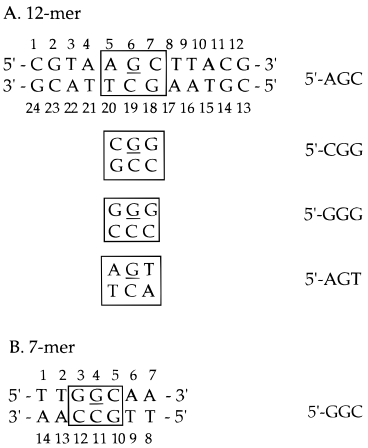

Figure 3. DNA duplexes used in NMR and modeling analyses include (A) 12-mer sequences incorporating the high reactivity 5′-AGC and 5′-CGG sequences, the moderate reactivity 5′-GGG sequence, and the low reactivity 5′-AGT sequence, and (B) a 7-mer sequence with the high reactivity 5′-GGC triplet. Target sequences are enclosed in boxes.

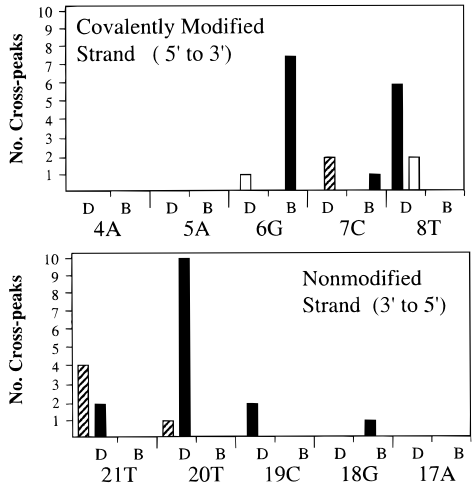

Figure 4. Histograms showing the number of NOESY cross-peaks observed between Et 736 and the 5′-AGC 12-mer duplex illustrate the 3′-side offset experienced by both the covalently modified DNA strand (upper panel) and unmodified strand (lower panel). Cross-peaks are distinguished by their origin from Et 736 A (diagonally hatched bars), B (black bars) or C subunit (white bars) protons and by the involvement of either DNA deoxyribose (D) or base (B) protons.

these signals reveal an extensive pattern of drug-to-DNA cross-peaks (see the table in the Supporting Information) characterized by a 3′-side offset "footprint" for both DNA strands (Figure 4). From the site of covalent drug attachment, 6G, there extend numerous cross-peaks to the 3′-side nucleotides, 7C and 8T on the covalently modified strand and 20T and 21T on the unmodified strand (Figure 4). To examine this pattern of drug−DNA associations, a preliminary computer model was generated by docking Et 736 adjacent to 6G and energy-minimizing (AMBER 4.1) the covalent adduct (Figure 5A). Analysis of the NOESY cross-peak distribution for the A, B, and C subunits of this model (Figures 4 and 5B) reveals a concentration of B- and C-subunit cross-peaks to the 3′-side nucleotides on the covalently modified strand and A- and B-subunit cross-peaks to the 3′-side nucleotides on the unmodified strand.

Et 736− and Et 743−DNA Duplex Adducts Share Similar ^{1}H NMR Profiles for the A and B Subunits. Et 736−(N2-guanine) DNA duplex adduct drug-to-DNA NOESY cross-peaks for the A and B subunits (white and yellow cross-peaks in Figure 5B) are virtually identical to those identified for the Et 743 adduct of the same duplex. These shared patterns are important because they document similar "fits" for the A−B-subunit scaffold of the drug in the minor groove. In addition to parallel cross-peak patterns, the chemical shifts of the DNA protons in the A−B scaffold region document similar shift effects caused by Et 736 and Et 743. An example of this shared effect is the radical upfield shifts experienced by 19C H1′, H2″, and H4′ after covalent drug attachment relative to the corresponding chemical shifts of 7C H1′, H2″, and H4′ in both the covalently modified strand and the two strands of the unmodified 2-fold symmetrical duplex (Figure 6; Figures 1 and 2, Supporting Information).

Covalent Attachment of Et 736 to the 12-Mer Duplex Produces Modest Duplex Distortion in That Region Closely Associated with the C Subunit. Comparison of NMR proper-

Table 1. Chemical Shifts of Nonexchangeable and Exchangeable Protons of (A) the Et 736 Adduct of 5′-AGC 12-mer DNA Duplex Strands 1 (C1−G12) and 2 (C13−G24), (B) Et 736 Covalent Form, and (C) the Unmodified 12-mer Duplex (Self-Complementary C1−G12)

A.

	H8/H6	H1′	H2′	H2′′	H3′	H4′	H5′′	H5′	H2	H7	H1	H3	N2Hb	N2He	N4Hb	N4He
C1	7.642	5.771	2.012	2.427	4.697	4.057	3.715	3.715								
G2	7.915	5.942	2.666	2.783	4.961	4.351	4.092	4.004			12.781		6.800	6.800		
T3	7.158	5.591	2.012	2.378	4.858	4.341	4.194	4.093		1.504		13.236				
A4	8.227	6.035	2.783	2.949	5.059	4.419	4.175	4.082	6.719							
A5	7.837	5.850	2.354	2.598	5.010	4.307	4.263	4.223	7.353							
G6	7.793	6.025	2.637	2.793	5.029	4.404	4.218	4.169			13.661		9.719			
C7	7.339	5.879	2.08	2.236	4.570	4.482	4.238	4.111							7.910	6.440
T8	7.236	5.976	2.402	1.836	4.887	3.706	4.224	4.043		1.484		13.872				
T9	7.334	5.718	2.148	2.456	4.937	4.163	4.214	3.994		1.606		13.661				
A10	8.301	6.167	2.695	2.837	5.039	4.429	4.170	4.028	7.524							
C11	7.275	5.649	1.836	2.266	4.766	4.141	4.238	4.058							8.220	6.610
G12	7.905	6.123	2.343	2.578	4.639	4.150	4.048	4.048								
C13	7.631	5.771	2.012	2.427	4.697	4.057	3.716	3.716								
G14	7.964	5.976	2.666	2.783	4.971	4.351	4.092	4.004			12.778		6.800	6.800		
T15	7.246	5.688	2.124	2.495	4.888	4.189	4.243	4.092		1.514		13.280				
A16	8.271	5.952	2.798	2.949	5.063	4.409	4.184	4.101	6.851							
A17	8.018	6.055	2.651	2.871	5.029	4.473	4.243	4.218	7.358							
G18	7.475	5.601	2.451	2.451	5.015	4.336	4.263	4.218			12.887		8.670	8.670		
C19	7.343	4.863	1.548	1.709	4.565	3.828	4.272	4.106							8.070	6.190
T20	7.642	6.162	2.539	2.656	5.049	4.575	4.238	3.935		1.675		13.976				
T21	7.329	5.679	2.012	2.334	4.888	3.828	4.419	4.092		1.738		13.671				
A22	8.267	6.103	2.641	2.788	4.971	4.341	4.199	4.101	7.368							
C23	7.241	5.625	1.841	2.271	4.766	4.106	4.194	4.048							8.220	6.610
G24	7.895	6.113	2.343	2.578	4.639	4.150	4.048	4.048								

B.

Et 736 Subunit:											
A	H15	C16Me	17OMe	C18OH	H14A	H14B					
	6.95	2.43	3.69	9.94	3.45	3.36					
B	H13	H21	H11	H1	H3	H4	12Me	AcMe	C6Me	H23A	H23B
	5.08	4.52	4.88	4.57	4.02	4.22	2.81	2.37	1.98	6.54	5.93
C	N2′H	H3′A	H3′B	H4′A	H4′B	H5′	H6′	H7′	H8′		
	10.08	3.54	3.15	2.77	2.64	7.52	7.12	7.28	7.47		

C.

	H8/H6	H1′	H2′	H2′′	H3′	H4′	H5′	H5′′	H2	H7	H1	H3	N2Hb	N2He	N4Hb	N4He
C1	7.640	5.784	2.019	2.427	4.697	4.073	3.708	3.708								
G2	7.972	5.966	2.664	2.772	4.966	4.353	4.008	4.094			12.791					
T3	7.221	5.590	1.997	2.320	4.847	4.149	4.095	4.170		1.492		13.342				
A4	8.224	5.859	2.718	2.869	5.052	4.374	4.030	4.127	7.031							
A5	8.023	5.945	2.589	2.804	5.020	4.407	4.192	4.192	7.429							
G6	7.547	5.687	2.449	2.557	4.890	4.342	4.084	4.181			12.720					
C7	7.290	5.870	2.040	2.481	4.676	4.374	4.127	4.202							7.890	6.360
T8	7.434	6.020	2.126	2.546	4.848	4.170	4.084	4.127		1.543		13.960				
T9	7.393	5.730	2.083	2.427	4.880	4.127	4.094	4.116		1.707		13.615				
A10	8.319	6.182	2.696	2.836	5.020	4.428	4.116	4.202	7.526							
C11	7.280	5.654	1.836	2.267	4.762	4.245	4.116	4.148							8.240	6.610
G12	7.885	6.128	2.341	2.567	4.633	4.149	4.062	4.159								

ties of the unmodified 12-mer duplex and its Et 736−(N2-guanine) DNA adduct (Table 1 and Figure 6) indicates that DNA protons in the vicinity of the drug display significant chemical shift changes resulting from the shielding/deshielding effects of the substituted aromatic ring system of Et 736. However, unbroken ^{1}H−^{1}H NOESY cross-connectivity patterns (e.g., Figures 1 and 3, Supporting Information) indicate that only a modest level of duplex structural perturbation results from introduction of the drug. The data indicate that duplex alkylation at 6G causes the single set of NMR signals of the unmodified self-complementary duplex to split into two distinct sets of signals (Table 1) corresponding to the covalently modified (C1−G12) and unmodified (C13−G24) strands. Other than protons in the immediate vicinity of drug attachment, the only significant effects of covalent attachment of Et 736 to DNA are in the chemical shifts that change in the backbone of the covalently modified DNA strand, specifically in the area adjoining the C subunit. For example, the 8T and 9T ^{31}P NMR chemical shifts (−1.38 and −1.42 ppm, respectively; Figure 4, Supporting Information) are the only ones that shift far upfield relative to their positions in the unmodified duplex spectrum (−0.62 and −0.45 ppm). A second significant change resulting from the close proximity to the C subunit is the upfield chemical shifts of the 8T deoxyribose ^{1}H signals (Figure 6; Figure 2, Supporting Information) relative to their chemical shifts in the unmodified duplex. This effect corresponds to the shielding pattern predicted

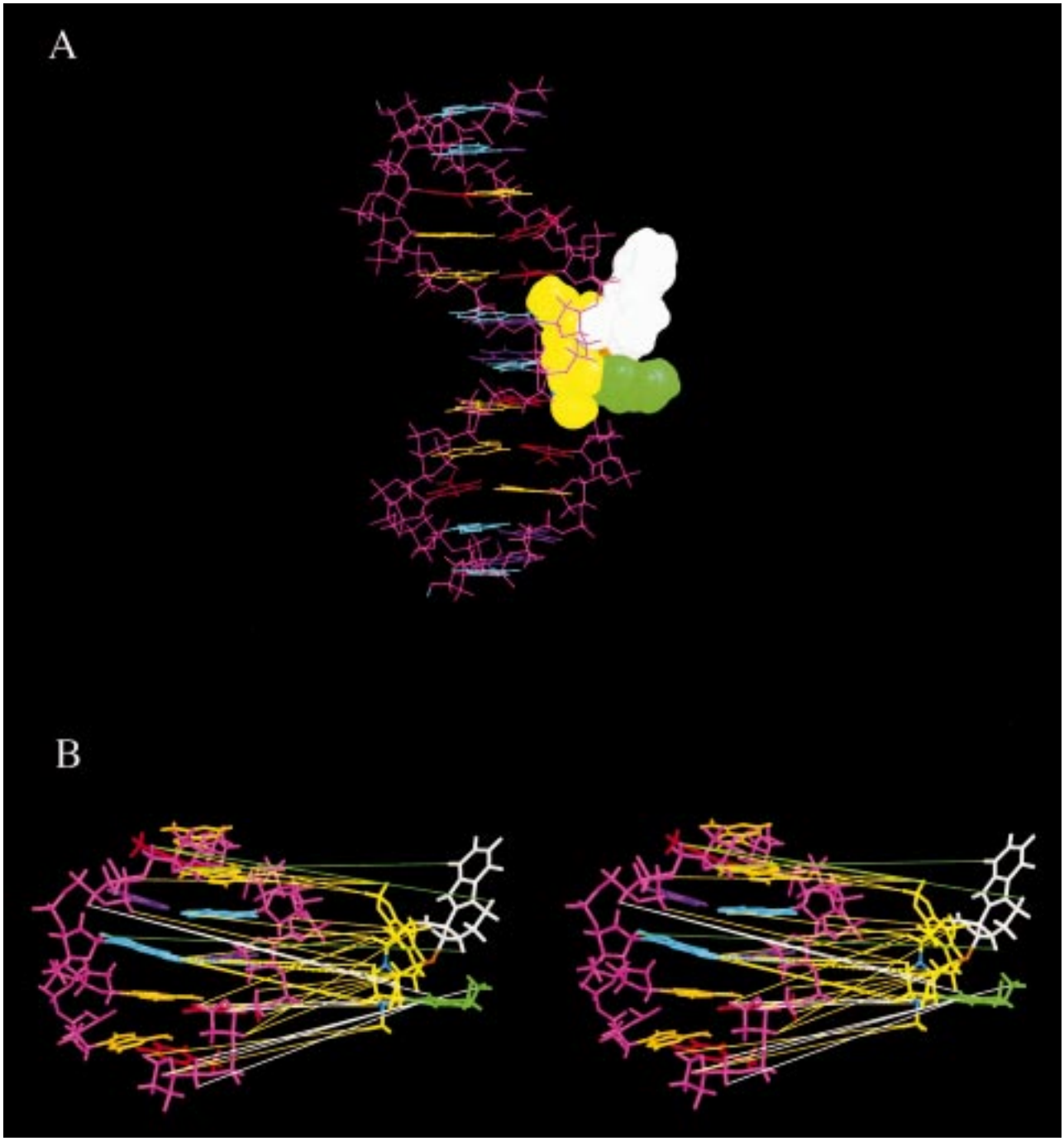

Figure 5. Energy-minimized structure of the Et 736−DNA 12-mer duplex adduct showing (A) a conic view of the three Et 736 drug subunits A (green), B (yellow), and C (white) and their positions in the minor groove and (B) a stereoview of the central region of the adduct showing NOESY cross-peaks as lines linking drug (extracted from minor groove) and DNA protons. Connectivity lines are white, yellow, and green for A, B, and C subunits, respectively.

from the model of the Et 736−DNA adduct structure: 8T H2″ > H4′ > H1′ > H2′ > H3′ (Table 1); i.e., H2″ is the nearest neighbor of the C-subunit ring system, followed by the next most closely positioned proton, H4′, etc. These 8T ^{31}P and ^{1}H NMR chemical shift changes are consistent with a DNA backbone structure that "adjusts" to its juxtaposition with the bulky Et 736 C subunit (Figure 5A) without any radical conformational perturbation.

Identification of the Hydrogen-Bonding Network between the A−B-Subunit Scaffold of Et 736 and Duplex DNA. For those exchangeable ^{1}H NMR signals implicated in drug−DNA hydrogen bonding, more data are available from the Et 736−(N2-guanine) DNA adduct than the Et 743−(N2-guanine) DNA adduct of the same 12-mer duplex. The most significant ^{1}H NMR signals and associated NOE cross-peaks are those providing experimental proof for exchangeable protons associated with hydrogen-bond donors in the drug−DNA hydrogen-bond network. Fortunately, the extensive NOESY and COSY cross-peak arrays for the Et 736 duplex adduct permit a detailed description of this network. Those data available for both Et 736 and Et 743 indicate that these drugs produce the same HB1−HB4 hydrogen bonds with the 5′-AGC site (Figure 7). However, a comparison of the NMR data of the Et 743 12-mer adduct with the Et 736 adduct data reveals different positions for the C subunits of Et 743 and Et 736 relative to the drug's A and B subunits and to the neighboring backbone of the

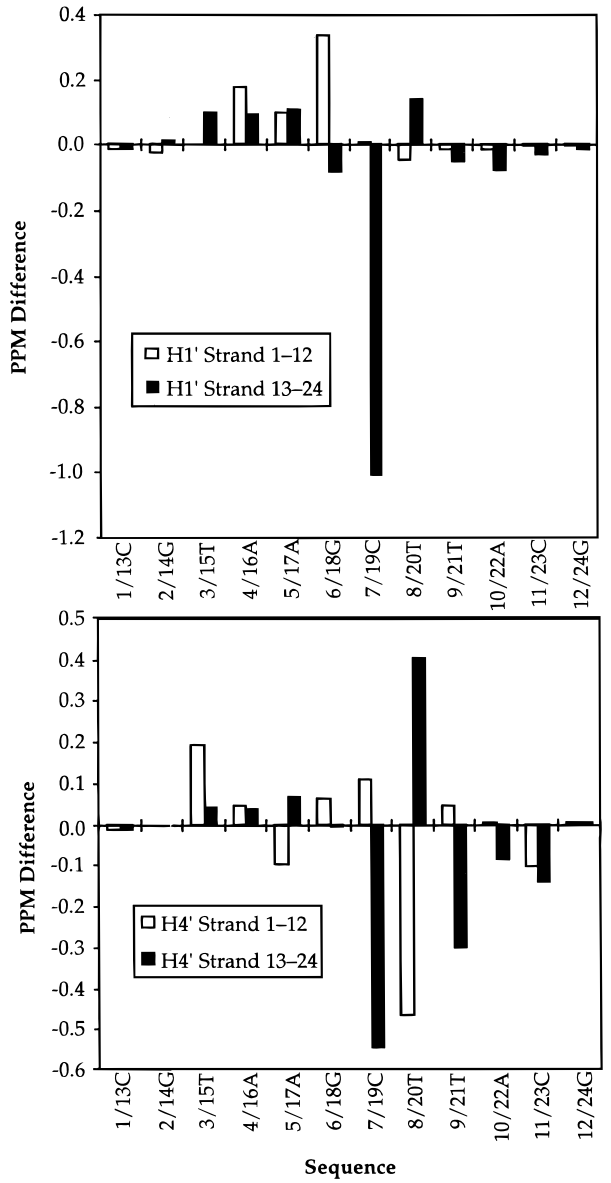

Figure 6. Chemical shift profile of the deoxyribose H1′ (upper panel) and H4′ (lower panel) protons expressed as the difference between the chemical shifts of the covalently modified (1−12) and unmodified (13−24) strands of the Et 736−DNA 12-mer duplex adduct and the corresponding chemical shifts of the unmodified 5′-AGC 12-mer duplex.

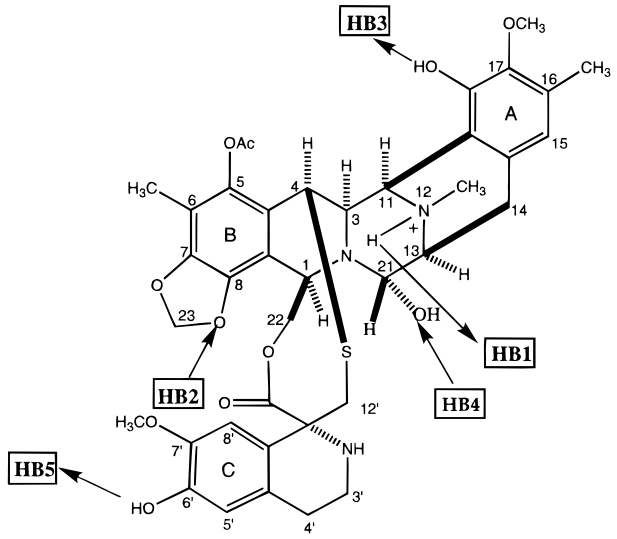

Figure 7. Sites of Et 743 (carbinolamine form) and DNA hydrogen-bonding interaction (HB1−HB5). Direction of arrow indicates the path from donor to acceptor.

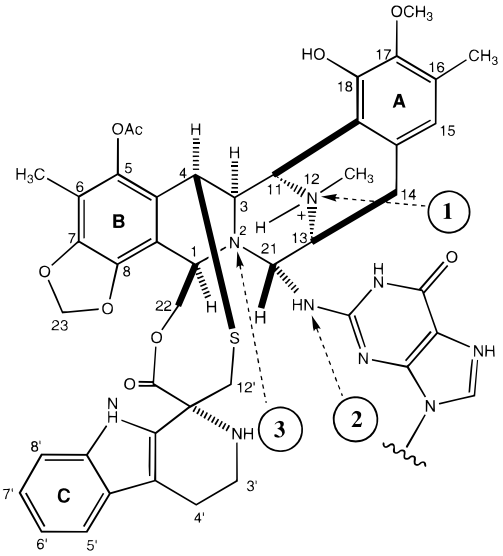

Figure 8. Structure of Et 736 showing the three possible protonation sites (circled 1−3).

covalently modified DNA strand. These differences result in part from the absence in the Et 736 12-mer duplex adduct of the hydrogen bond produced between the Et 743 C subunit's C6′-OH and the 9T O1P phosphate oxygen of the covalently modified DNA strand (HB5, Figure 7). The differences between the Et 743 and Et 736 duplex adducts will be addressed in a subsequent publication (Seaman and Hurley, manuscript in preparation).

Any tenable proposal of a role for hydrogen bonding in the sequence recognition of DNA by Et 736 must conform to the experimental results relating to potential donor/acceptor moieties in the minor groove. Those adduct NOESY data relating to HB1−HB3 hydrogen-bonding patterns (Figure 7) are summarized below.

A. Evidence That the N12*H* of the B Subunit Hydrogen Bonds to 20T O2 in the Unmodified DNA Strand (HB1 in Figure 7). Circumstantial evidence that N12 is the site of adduct protonation in the Et 736 12-mer duplex adduct rather than one of the alternative sites (2 or 3 in Figure 8) is documented in the Supplementary Information, Figure 6A,B. These data show an exchangeable signal at 7.97 ppm that is characterized by NOESY (H_2O/D_2O) cross-peaks to N12 methyl, H11, and H3 signals of Et 736 and to 6G N2H, 5A H2, and 20T H1′ (not shown) of DNA. These NOESY cross-peaks pinpoint N12 as the probable location of protonation but do not unequivocally exclude other possible sites. Unequivocal proof of N12 protonation was obtained from a COSY (H_2O/D_2O) cross-peak between the 7.97 ppm signal and the N12 methyl signal (Figure 4, Supporting Information). After this discovery, a similar pattern of N12 protonation was observed in the equivalent Et 743 adduct,[8] confirming that the protonated N12 component of the hydrogen-bonding network is shared by the two 12-mer duplex adducts. Specifically, this axially protonated N12 places a hydrogen-bond donor next to the O2 acceptor of the thymine (20T), which is paired with the adenine (A5) adjacent to the covalently modified G6. The previously mentioned set of NOESY cross-peaks (Figure 6A,B, Supporting Information) indicates that N12H is directed toward the predicted 20T O2

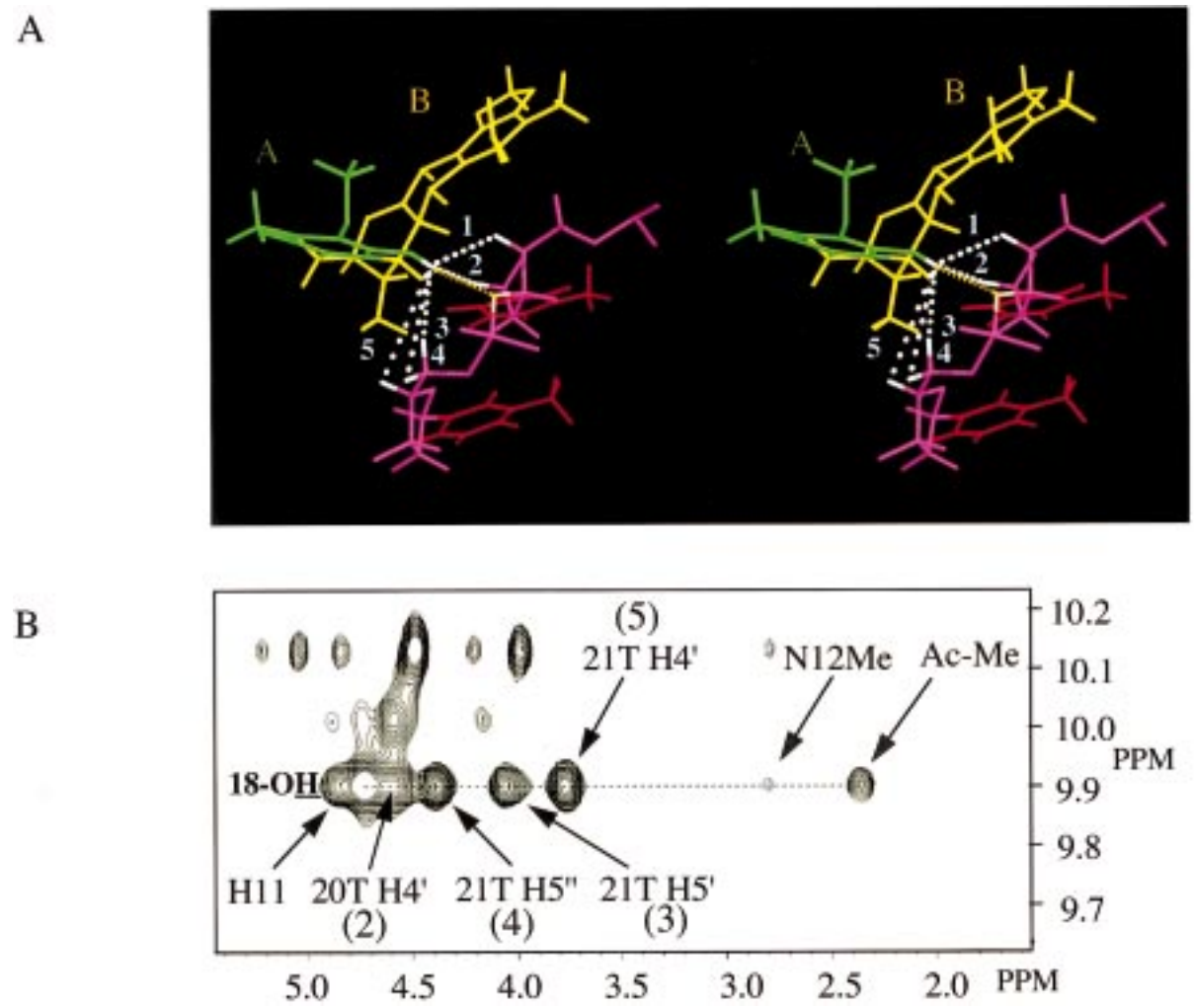

Figure 9. (A) Stereoview of the modeled duplex adduct depicts the Et 736 C18 hydroxyl proton's hydrogen bond to 20T O3′ (yellow dotted line) and cross-peaks to neighboring 20T and 21T protons (white dotted lines). Crosspeaks 3 and 4 refer to the C5′ protons far from and near to C18, respectively. (B) NOESY 2D-spectral region showing Et 736 A-subunit C18 hydroxyl proton cross-peaks to 20T and 21T sugar protons and other drug protons. Numbers 2−5 correspond to cross-peaks shown in A.

acceptor. Hydrogen bonding of N12H is further supported by its NMR signal's sharpness and downfield chemical shift (7.97 ppm).

B. Evidence That the 18G N2*He* Hydrogen Bonds to the C8-O-C23 Hydrogen-Bond Acceptor of the B Subunit (HB2 in Figure 7). NOESY data for the 7C−18G base pair in the 5′-AGC target sequence indicate normal base pairing. For example, the 7C N4Hb and N4He chemical shifts (Table 1) and cross-peaks with 18G H1 (Figure 7A,B, Supporting Information) are typical of a base-paired cytosine. Likewise, the 18GH1 chemical shift and its NOESY cross-peaks are characteristic of normal base pairing. This base pairing also includes hydrogen bonding of 18G N2Hb to 7C O2 (Figure 7A, Supporting Information). Experimental evidence for hydrogen bonding in the 12-mer DNA duplex adduct between the other 18G amino proton, N2He, and the B-subunit methylenedioxy acceptor (C8-*O*-C23) is the 18G amino proton pair's signal downfield shift to 8.67 ppm from its typical chemical shift at approximately 6.8 ppm in a GC base pair of an unmodified duplex (Figure 7A,B, Supporting Information). Hydrogen bonding of both 18G N2He with C8-*O*-C23 and 18G N2Hb with 7C O2 results in the unusual downfield shift of the broadened 18G amino signal. The intense Et 736 H23A and H23B cross-peaks with 8T H4′, H23A cross-peak with 7C H2″, and H23A and H23B cross-peaks with 8T H5″ (see the table in the Supporting Information) provide additional support for the juxtaposition and correct positioning of this hydrogen-bond donor (18G N2)/acceptor (C8-*O*-C23) pair.

C. Evidence That the C18-O*H* Moiety of the A Subunit Hydrogen Bonds to O3′ in 20T in the Unmodified Strand (HB3 in Figure 7). The intensity, sharpness, and chemical shift of the C18-O*H* proton signal of Et 736 in the 12-mer duplex adduct indicate that this proton is hydrogen bonded. Examination of Et 736−DNA 12-mer duplex adduct models shows that the Et 736 C18 hydroxyl function is near four possible DNA hydrogen-bond acceptors (20T O1′[O4′], 20T O3′, 21T O1P, 21T O4′) and one on the drug (C5-acetate carbonyl oxygen). The last option requires no deformation of Et 736 other than bond rotations to yield suitably oriented C5-acetate and C18 hydroxyl groups, and this hydrogen bond presumably exists in the prebinding form of the free drug. Hydrogen bonding in the free drug between C18-OH and the B-subunit acetate group orients the acetate two-carbon chain so that it projects outward from the minor groove with the acetate's methyl group sitting above a hydrophobic region (20T C4′ and C5′) on the backbone of the unmodified strand. The model of the free Et 736 suggests that this side-chain orientation may facilitate the entry of the B subunit into the minor groove prior to covalent bond formation.

The NOESY cross-peak and chemical shift data indicate that, after DNA adduct formation, C18-O*H* of Et 736 is directed

toward the 20T O3′ hydrogen-bond acceptor. This orientation and the implied hydrogen bonding are indicated by NOESY cross-peaks (Figure 9A,B and the table in the Supporting Information) for the exchangeable C18-O*H*. The intensity of its cross-peaks to 21T H4′, H5″, and H5′ suggests that the hydrogen-bonded C18-O*H* is closer to the sugar of this neighboring nucleotide than would be possible with hydrogen bonding to 20T O4′, as predicted in our earlier modeling study.[8] The downfield chemical shifts expressed by the 20T H1′, H2′, H2″, H3′, and H4′ signals (Figure 6; Figures 1−3, Supporting Information; Table 1) suggest that these protons experience the deshielding effect of the neighboring Et 736 C18 hydroxyl oxygen, possibly with some contribution from the localized deformation of the DNA backbone. Deshielding that is focused on 20T H2″, H2′, and H3′ is incompatible with C18-OH hydrogen bonding to 20T or 21T O4′ but is consistent with hydrogen bonding to 20T O3′. These data are also inconsistent with a 21T O1P acceptor, which is accessible only if the strand 2 backbone is distorted to bring this *O*1P acceptor into the proper orientation in relation to C18-O*H*. Finally, the presence of the HB1 hydrogen bond restricts the range of C18-O*H* targeting of potential acceptors. Whatever the sequence of drug-to-DNA hydrogen-bonding events, HB3 targeting of 20T O3′ fits comfortably with the HB1 targeting of 20T O2. Modeling indicates that proper orientation of the HB1 donor/acceptor pair positions the HB3 donor in close proximity to the O3′ acceptor site. When this 12-mer is modified by Et 736, it produces virtually identical NOESY cross-peak patterns for the 19C, 20T, and 21T regions as produced by modification by Et 743. These data suggest a similar HB3 association for both the Et 736− and Et 743−12-mer duplex adducts. While HB1 is a sequence (base)-dependent association, the HB3 association with the DNA backbone is less obviously dependent upon sequence.

D. Evidence That the B Subunit C21-*O*H of the Carbinolamine Binding Form of Et 736 Is Hydrogen Bonded to 6G N2*H* of DNA (HB4 in Figure 7). While not retained in the Et 736 12-mer duplex adduct and, thus, not available for experimental analysis, the proposed HB4 hydrogen bond fits as an integral part of the HB1−HB4 hydrogen-bond network (Figure 7) in the modeled Et 736 (carbinolamine form)−DNA association. Prior to alkylation, modeling suggests that the experimentally evaluated HB1, HB2, and HB3 hydrogen bonds direct the Et 736 C21 hydroxyl group toward 6G N2He and correctly orient the donor (6G N2) and acceptor (O21) for generating the HB4 hydrogen bond. The presence of HB4 is essential to the carbinolamine reaction mechanism proposed for adduct formation (Figure 2).[9]

Molecular-Modeling Studies Characterize the Structural Basis for the Rules Governing the Three-Base-Pair Sequence Preferences for Et 736 and Et 743. A. Previous Theoretical Molecular Modeling Studies. In the earlier modeling of Et 743− and Et 729−DNA 7-mer duplex adducts (Figure 3), the following drug−DNA hydrogen bonds were predicted for the 5′-GGC target sequence: (1) B-subunit N12 *H* to C12 O2 (HB1), (2) B-subunit C8-O-C23 to G10 N2*He* (HB2), (3) A-subunit C18−O*H* to A13 O4′ (HB3), and (4) C-subunit C6′-O*H* to A7 O2P5′ (HB5).[10] These predicted interactions and the NMR-based proposals for the Et 736− and Et 743−DNA interactions described here are in agreement, except that our experimental data relating to HB3 do not support the earlier proposed role of O4′ as an acceptor. Despite the differences between the DNA sequences and drug structures of the experimentally examined Et 736−12-mer duplex adduct and the modeled 7-mer structures, the Et 736 data indicate areas of solid agreement. The possibility of a common ecteinascidin theme for HB1−HB4 hydrogen-bonding interactions (Figure 7) suggested the need for a more intensive molecular-modeling study. The goals of this expanded study were to determine (1) if computer models of hydrogen-bonding interactions of Et 736 and Et 743 5′-CGG and 5′-AGC adducts agreed with the experimental results for HB1−HB4 and (2) whether these models provide a structural rationale for the proposed ecteinascidin sequence specificity rules.

B. Experimentally Based Molecular Modeling Studies. Despite the sequence differences, inspection of Et 736 5′-CGG and 5′-AGC 12-mer duplex adduct models shows that both display donor/acceptor pairs consistent with suitably positioned drug−DNA hydrogen bonds (HB1−HB4). The structures also suggest that the presence of a downstream guanine amino HB2 donor on either the covalently modified (i.e., 5′-CGG) or unmodified strand (i.e., 5′-AGC) controls (because of the rigid nature of the B subunit) the correct "positioning" of the HB1 donor relative to either an upstream purine acceptor (i.e., 20G N3 of 5′-CGG) or pyrimidine acceptor (i.e., 20T O2 of 5′-AGC) on the unmodified strand (Figure 10). To further evaluate these Et 736−DNA 12-mer duplex adduct models and to address the question regarding what factors govern the HB1/HB2 positioning, MD analysis was conducted on 12-mer sequences containing either 5′-AGC or 5′-CGG associated with the carbinolamine binding form of Et 736. In each instance, starting structures were generated by docking Et 736 into the minor groove such that C21 was positioned near the target guanine N2. Solvated rMD (100 ps) of this complex was performed using constraints that combined the standard Watson−Crick hydrogen-bond constraints with HB1, HB2, and HB3 distance constraints derived from NOESY-based ^{1}H−^{1}H distance information. Using the product of this initial rMD trajectory as the starting structure for solvated MD analysis, the degree of stability of drug−DNA complexes was evaluated by analyzing hydrogen bonding between DNA and drug in both in vacuo and solvated systems (Table 2). The stability of the complex was measured by noting those hydrogen bonds that remain intact throughout the last 20 ps of the trajectory following removal of the constraints.

C. Modeling of the Precovalent Binding Form of Et 736 with 5′-AGC and 5′-CGG. For Et 736 in the precovalent binding form there is stable HB1−HB4 hydrogen bonding for 5′-AGC and 5′-CGG binding sequences. In vacuo and solvated MD results for the Et 736 carbinolamine noncovalent binding form with the 5′-AGC target sequence in the 12-mer consistently show a stable hydrogen-bonding pattern involving four sites within the DNA minor groove: 20T O2 (HB1), 18G N2H (HB2), 20T O3′ (HB3), and 6G N2H (HB4). Although it is difficult to ascertain the precise sequence of binding events, one can postulate that 18G N2He (HB2 donor; Figure 10, top hydrogen bond) orients the drug in the minor groove so as to position the HB1 donor (Et 736 12*N*H) opposite its thymine O2 acceptor (Figure 10, bottom hydrogen bond). Optimal HB1 donor/acceptor orientation results in an HB3 donor (C18-*O*H) that is properly aligned in relation to its acceptor (20T O3′). The hydrogen bonds HB1, HB2, and HB3 collectively position the carbinolamine C21 hydroxyl group (HB4 acceptor) opposite the 6G N2H amino functional group (HB4 donor; Figure 10, middle hydrogen bond). The stability of the 5′-AGC binding complex of hydrogen bonds (Figure 11, upper panel) was examined by in vacuo and solvated MD. The results (Table 2A) indicate that all four hydrogen bonds remain intact throughout the latter unconstrained portion of the trajectory (100 ps, temperature ramp 0 to 600 to 300K).

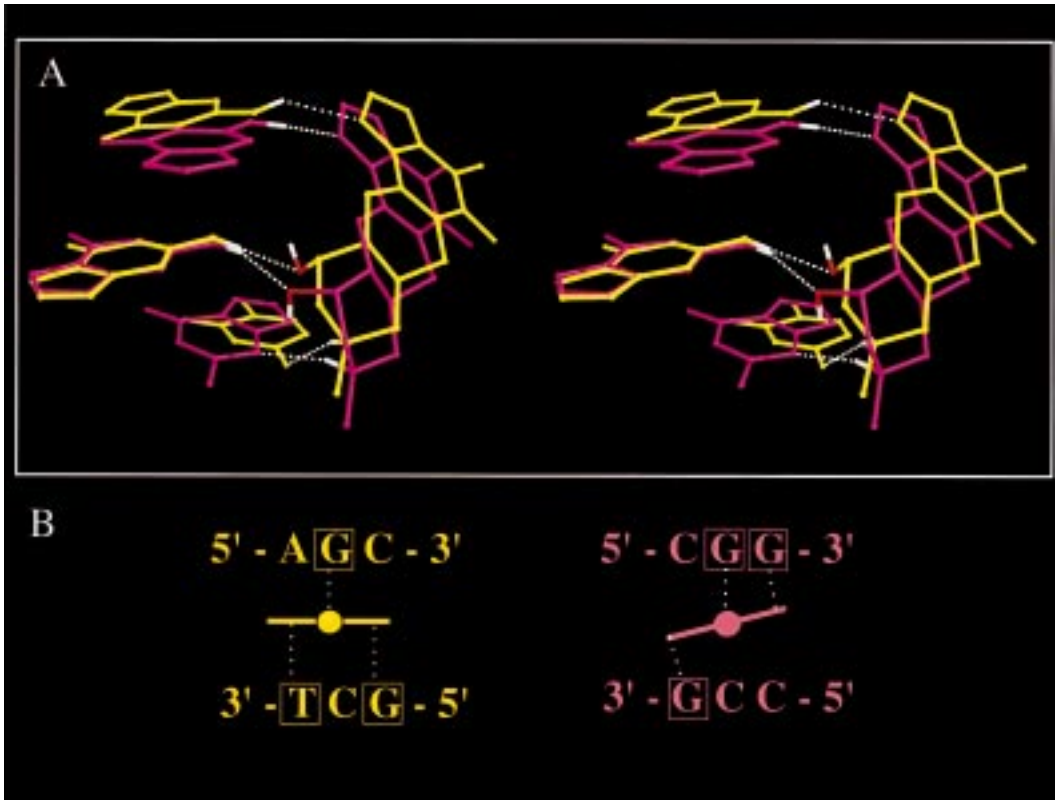

Figure 10. (A) Contrasting Et 736 B-subunit positions in binding associations with 12-mers containing the target sequences 5′-CGG (red) and 5′-AGC (yellow). Full duplex adducts were overlapped (matched) at the 6G base. Hydrogen bonds are indicated by white dotted lines. (B) Diagram showing the relative positions of strand one and two drug-to-DNA HB1, HB2, and HB4 hydrogen bonds for the two sequences.

The solvated and in vacuo MD results for the 12-mer containing the 5′-CGG target sequence parallel those of the previous sequence except for the DNA sites of HB1−HB4 association: 20G N3 (HB1), 7G N2H (HB2), 20G O3′ (HB3), and 6G N2H (HB4). The 7G N2H HB2 donor (Figure 10, upper hydrogen bond) orients the rigid A−B scaffold portion of the drug in the minor groove so that the HB1 donor (12NH) is opposite its target acceptor, 20G N3 (Figure 10, bottom hydrogen bond), the HB3 donor (C18-*O*H) is opposite its 20G O3′ acceptor, and the HB4 C21 hydroxyl group acceptor is opposite its 6G donor (Figure 10, middle hydrogen bond). The degree of hydrogen bond stability during the in vacuo and solvated MD simulations (Figure 11, lower panel; Table 2A) rivals that of the 5′-AGC binding complex.

D. Molecular Modeling of the Precovalent Binding Form of Et 736 with 5′-GGG and 5′-AGT. For the series of DNA target sites of decreasing sequence selectivity,[7] there is a corresponding reduction in the Et 736 hydrogen-bonding network stability. 5′-GGG, the next most reactive sequence following the four highly reactive sequences, was selected for modeling analysis. Although 5′-GGG presents an HB1 acceptor (20C O2) resembling that (20T O2) of the favored sequence, 5′-AGC, its HB2 donor (7GN2) originates from the strand opposite the HB2 donor source in 5′-AGC (i.e., 18G N2). Solvated MD results for this model structure were examined in order to compare the stability of its HB1−HB4 association relative to the stability of the favored sequences. The lower average hydrogen-bond frequency and the greater average donor−acceptor distance for HB1 (Table 2A) indicate a lower level of hydrogen bond stability than 5′-AGC.

The poorly reactive sequence, 5′-AGT, fails as a favored sequence presumably because it lacks either a 3′-side G or C. The resulting absence of the guanine amino group in the minor groove eliminates all prospects of HB2 hydrogen bonding. To evaluate the effect of this loss on the stability of the binding complex, solvated MD analysis was performed. The hydrogen-bond stability characterizing the favored sequences (Table 2A) was absent from this binding complex, which failed to generate uniformly positioned A−B scaffolds upon completion of the MD trajectories. Different MD conditions and starting structures yielded different aberrant A−B scaffold orientations, in contrast to the favored sequences, which retained comparable A−B scaffold positions despite varying MD conditions. For the 5′-AGT model, drug−DNA association was characterized by such unusual behavior as partial intercalation of the B subunit and migration of the drug in the minor groove away from the target guanine.

Molecular Modeling of the Et 736 and Et 743 Covalent DNA Adducts of Favored Sequences Shows That They Share Stable Hydrogen Bonding Networks. The Et 736−5′-AGC duplex adduct structure differs from the carbinolamine binding forms by the substitution of a covalent bond between 6G N2 and Et 736 C21 in roughly the same location as the HB4 hydrogen bond. Parallel MD studies for the covalent adduct show similar hydrogen-bond stabilities (Table 2B). MD simulation of the Et 736−5′-CGG target adduct shows a level of hydrogen-bond stability equivalent to that shown by both the 5′-CGG noncovalent binding model and the 5′-AGC covalent adduct model (Table 2B).

The complementary "fit" of the covalently bound Et 736 in the minor grooves of the 5′-AGC and 5′-CGG target sequences facilitated by the array of HB1−HB3 hydrogen bonds raises the question of whether Et 743 fits as well despite the presence of the additional HB5 hydrogen bond. To address this question, a solvated MD analysis was conducted on the Et 743 duplex adduct of the 5′-AGC 12-mer and the 5′-CGG alternative 12-mer sequence (Table 2C). On the basis of an analogous earlier study,[10] a hydrogen bond was positioned between the model's

Table 2. Drug-to-DNA (HB1−HB5) Hydrogen Bond Data from Molecular Dynamics Analysis of (A) Et 736 Carbinolamine Binding Form, (B) Et 736 N2 Adduct Form, and (C) Et 743 Adduct Form

A. Et 736: Binding Form

	Donor	Acceptor	Avg. Frequency		Avg. Distance/Angle	
			In Vacuo	Solvated	In Vacuo	Solvated
5′-AGC-3′						
HB1	25 N12	20T O2	1.00	1.00	2.92/15.55	3.02/16.90
HB2	18GN2	25Et O8	1.00	1.00	2.96/12.87	2.90/21.38
HB3	25Et O18	20T O3′	0.96	0.92	2.99/39.69	3.10/26.02
HB4	6G N2	25Et O21	1.00	0.98	2.90/19.17	2.79/29.84
5′-CGG-3′						
HB1	25Et N12	20G N3	0.99	0.94	3.17/13.80	3.21/14.62
HB2	7G N2	25Et O8	1.00	1.00		2.88/20.92
HB3	25Et O18	20T O3′	0.98	0.92	2.64/12.56	2.88/25.44
HB4	6G N2	25Et O21	0.99	0.99	2.77/45.06	2.85/28.54
			Solvated		Solvated	
5′-GGG-3′						
HB1	25Et N12	20C O2	0.94		3.15/24.56	
HB2	7G N2	25Et O8	0.98		2.99/20.07	
HB3	25Et O18	20C O3′	0.98		3.01/28.08	
HB4	6G N2	25Et O21	1.00		2.88/22.56	
5′-AGT-3′						
HB1	25 N12	20T O2	Not Observed			
HB3	25Et O18	20T O3′	Not Observed			
HB4	6G N2	25Et O21	0.99		2.88/30.25	

B. Et 736: Covalent Form

	Donor	Acceptor	Avg. Frequency	Avg. Distance/Angle
			Solvated	Solvated
5′-AGC-3′				
HB1	25 N12	20T O2	0.99	3.02/21.06
HB2	18G N2	25Et O8	1.00	2.84/18.62
HB3	25Et O18	20T O3′	0.96	3.03/31.0
5′-CGG-3′				
HB1	25Et N12	20G N3	0.99	3.17/19.17
HB2	7G N2	25Et O8	1.00	2.88/23.04
HB3	25Et O18	20T O3′	0.96	3.01/31.68

C. Et 743: Covalent Form

	Donor	Acceptor	Avg. Frequency	Avg. Distance/Angle
			Solvated	Solvated
5′-AGC-3′				
HB1	25 N12	20T O2	99.4	3.00/19.42
HB2	18G N2	25Et O8	1.00	2.91/19.91
HB3	25Et O18	20T O3′	97.4	3.03/23.58
HB5	25 HO6′	9T O1P	97.2	2.93/18.74
5′-CGG-3′				
HB1	25Et N12	20G N3	98.8	3.05/23.77
HB2	7G N2	25Et O8	99.2	2.92/20.50
HB3	25Et O18	20T O3′	95.2	2.98/29.01
HB5	25 HO6′	9T O1P	95.2	2.90/16.03

Average values for hydrogen bond presence/absence, distance, and angle are calculated from the 500 steps of the 100 psec molecular dynamics trajectory (0K to 600K to 300K temperature ramp). Hydrogen bond presence is determined according to criteria of maximum distance (3.00Å) and angle (60°).

Et 743 C6′-OH and the 12-mer duplex 9T O1P. In both adducts, the results confirm that a stable Et 743 adduct is produced with the full complement of HB1, HB2, HB3, and HB5 hydrogen bonds without any significant deformation of the duplex structure.

Discussion

The ecteinascidins are structurally unique among DNA-reactive drugs and appear to have promising clinical activity on the basis of the results of phase I clinical trials in the United

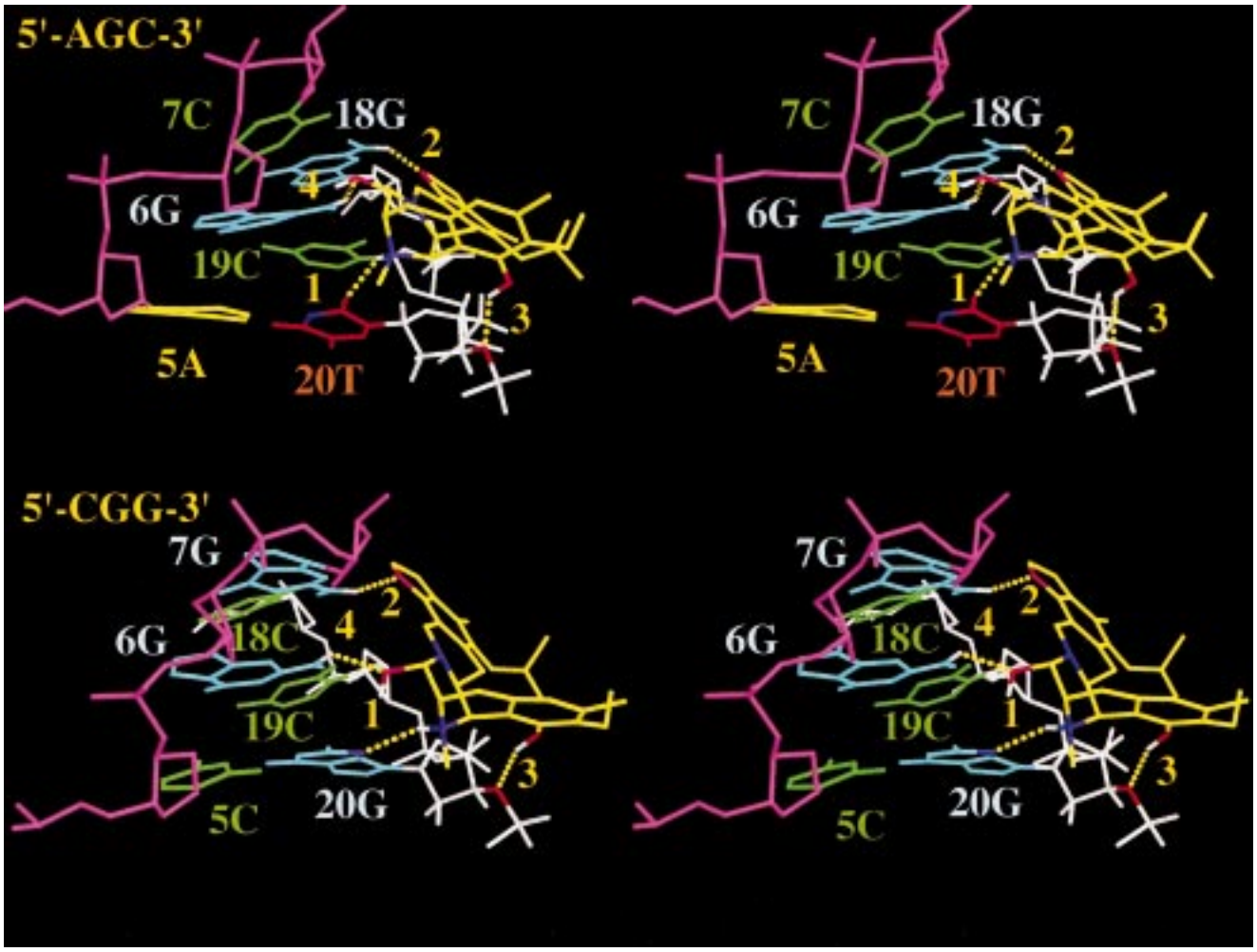

Figure 11. Stereoviews of DNA triplet regions 5′-AGC (upper panel) and 5′-CGG (lower panel) hydrogen bonding to the A−B-subunit scaffold (yellow) of Et 736. The four hydrogen bonds (1−4, yellow dotted lines) are HB1 (1), HB2 (2), HB3 (3), and HB4 (4).

States and Europe. Biosynthetically, they are more complex than saframycins and naphthyridinomycins due to the additional C-subunit moiety, which imparts a more wedgelike shape to the ecteinascidins. In an attempt to derive a rationale for the biological properties of the Et compounds, we have embarked upon structural and mechanistic studies of the interaction of Et 743 and Et 736 with DNA. In a previous study, we have unambiguously determined the covalent linkage between Et 743 and DNA and the conformational fit of Et 743 in the minor groove,[8] and most recently we have determined a mechanism for catalytic activation of the carbinolamine with N2 of guanine.[9] In this paper, we provide a proposal for the observed sequence selectivity of Et 743 and 736 that involves direct readout via coordinated hydrogen bonding. Before elaborating on this proposal we provide evidence that this mechanism is common to both Et 743 and Et 736.

Evidence for Structural Parallels between Et 736 and Et 743 and Their Interaction with DNA. Conformational Parallels. The results described in this contribution together with the previously published findings document structural parallels between the Et 736− and Et 743−5′-AGC 12-mer duplex adducts: (1) The two covalent adducts share a common site of drug protonation, N12, which forms a hydrogen bond with the 20T O2 acceptor on the noncovalently modified DNA strand. (2) Shared ^{1}H NOE patterns and chemical shift effects resulting from duplex alkylation suggest that the A−B subunit scaffolds of Et 736 and Et 743 occupy very similar sites within the minor groove and presumably generate identical hydrogen-bond associations. Evidence from both Et 736 and Et 743 duplex adducts indicate that the direct contact between drug and DNA bases encompasses only the two base pairs that flank the covalently modified guanine. For Et 743, the association is extended to include the C-subunit C6′-OH hydrogen bond (HB5) with the 12-mer backbone 9T phosphate oxygen (O1P). (3) ^{1}H−^{1}H NOESY cross-connectivity patterns for DNA that remain unbroken after drug alkylation (e.g., Figures 1 and 3, Supporting Information) suggest that the DNA duplex is not highly distorted in either the Et 736 or Et 743 duplex DNA adducts.

Hydrogen-Bond Network Parallels. With the exception of the DNA backbone hydrogen-bonding potential of the Et 743 C subunit, the evidence indicates that Et 736 and Et 743 share a common pattern of hydrogen-bond donor/acceptor associations. The absence of a C-subunit C6′-OH moiety in Et 736 precludes any HB5 interaction between its C subunit and the neighboring DNA backbone. Any proposal of a hydrogen-bonding role in the sequence recognition process must conform to the experimentally observed pattern detailed above.

Deriving a Structural Basis for the Proposed Rules That Govern the Three-Base (5′-XGY) Sequence Preferences of Et 736 and 743. A. Analysis of the Results of the Experimentally Determined Sequence Specificity Demonstrates a Base-Triplet Sequence Specificity Pattern for Et 743. The NMR and modeling results suggest that ecteinascidin sequence specificity can be defined in terms of the base-triplet target sequence. This sequence selectivity profile must agree with the experimental findings of an earlier oligonucleotide band shift

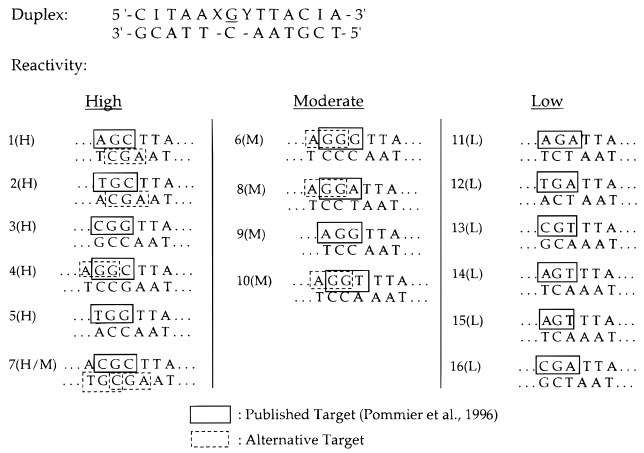

Figure 12. Et 743 adduct sequences examined by duplex band shift assay[7] are ranked in descending order from 1 to 16, according to their reactivity with 100 μM Et 743. Each duplex is identified by its classification as high (H), medium (M), or low (L) reactivity based on all potential target base triplets found in the duplex.

assay study of Et 743 adducts of different three-base sequence targets, 5′-XGY.[7] In this band shift assay study of 16 sample sequences, a total of five sequences surpassed the guanine alkylation activity of the reference sequence, 5′-GGG. It is clear from a ranking of these 16 sequences from high to low reactivity (Figure 12) that more potential triplet target sequences are present in some of the duplexes than identified in the original study. For example, alkylation of 5′-TGC in 5′-CITAATGCT-TACIA cannot be experimentally distinguished from alkylation of the 5′-AGC target sequence on the opposite strand of the duplex. The reinterpretation of these results (Figure 12) led to the proposal of a single set of rules governing drug preference: For the target sequence, 5′-XGY, the favored base to the 3′-side, Y, is either G or C. When Y is G, then a pyrimidine base (T or C) is favored for X. When Y is C, a purine (A or G) is favored for X. All high-reactivity sequences (Figure 12, duplexes 1−5) conform to these rules regarding both 3′- and 5′-side requirements, while moderately reactive sequences (duplexes 6 and 8−10) satisfy only the first requirement, that the 3′-side base (Y) be G or C. Low reactivity sequences (Figure 12, duplexes 11−16) fail to satisfy this first requirement. One sequence, duplex 7, contains, in addition to the target base-triplet identified in the original study (5′-TGC), overlapping high- and low-reactivity targets, 5′-AGC and 5′-CGT, respectively, on the opposite strand. It is the only duplex containing a high-reactivity target that ranked lower in reactivity than 5′-GGG (duplex 6), possibly resulting from the doubly overlapped nature of the three base-triplet targets. Despite this one anomalous result, the band shift assay rankings correlate with predicted reactivity based on the application of our proposed reactivity rules to the potential targets in the duplexes.

The Hydrogen-Bonding Properties of the Base Pair That Is to the 3′-Side of the Covalently Modified Guanine Differentiate Moderate- and High-Reactivity Sequences from Low-Reactivity Sequences. Sequence specificity data for the three-base sequence 5′-XG$_{COV}$Y demonstrate that either G or C is favored for Y (Figure 13); the presence of either A or T results in low reactivity (Figure 12). For either Y = G or C, an amino group HB2 hydrogen-bond donor occupies the center of the minor groove. This amino group donor, projecting from a guanine on either the covalently modified strand or the opposite strand, can target the Et 736 C8-O-C23 oxygen acceptor. Failure to satisfy this requirement is typical of all low-reactivity sequences (Figure 12).

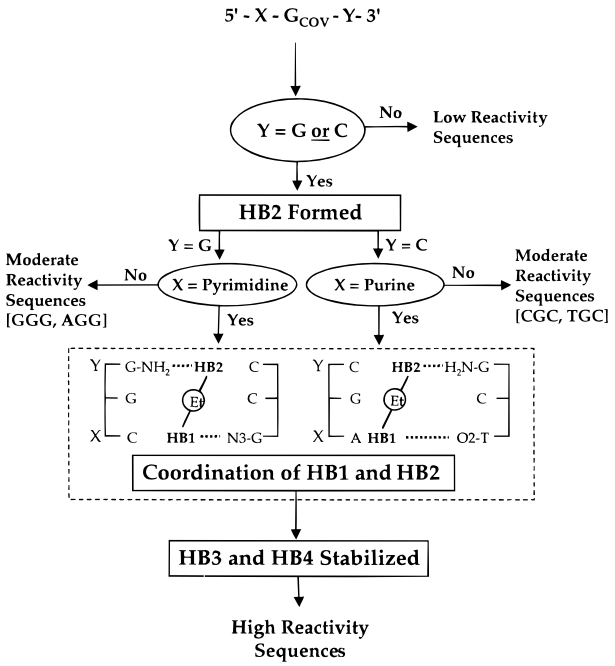

Figure 13. Sequence/hydrogen-bonding criteria: flowchart of reactivity. Flowchart analysis of the sequence specificity of the three-base sequence 5′-XG$_{COV}$Y shows that Y must be either G or C, and depending on which base occupies Y, the X choice must be of the opposite ring system type (e.g., Y = G and X = pyrimidine). The relationship between these sequence rules and DNA−drug hydrogen bonding is displayed in the flowchart, which ultimately leads to the goal of high reactivity.

Proper Orientation of the Ecteinascidin A−B-Subunit Scaffold within the Minor Groove Differentiates High- from Moderate-Reactivity Sequences. Identifying the favored base to the 5′-side of the covalently modified guanine of a high-reactivity sequence depends on whether the 3′-side base is either G or C. If Y = G in 5′-XG$_{COV}$Y, then the favored selection for X is a pyrimidine, either cytosine or thymine. For example, 5′-CGG is one of the favored sequences in the Pommier study,[7] and 5′-TGG is only slightly less reactive. If Y = C, then the favored X selection is a purine, either adenine or guanine. Consequently, both 5′-GGC and 5′-AGC are highly reactive. These four sequences, CGG, TGG, GGC, and AGC, are the most reactive of all the two-base permutations around the central G (Figure 12). Presumably, the significance of this pattern is that it maximizes the coordination of HB1 and HB2 hydrogen bonding, which, in combination with HB3, results in optimal positioning and stabilization of HB4 at the site of alkylation.

The precise nature of the three-base sequence specificity indicates that a combined read-out of the two flanking bases plus the central guanine governs reactivity. An explanation for the precision of the three-base sequence read-out follows from the combination of minor groove geometry and the positioning of HB1−HB4 drug donors/acceptors along the full length of the A−B scaffold. The essential determinant seems to be the coordination of HB1 and HB2 hydrogen bonding (Figure 13). Dual optimal hydrogen-bonding orientations result from a guanine HB2 donor that originates from either the covalently modified strand (e.g., 5′-CGG) or the opposite strand (e.g., 5′-AGC). These different drug positions in the minor groove correlate with different optimal HB1 arrangements. The 5′-CGG target sequence orients the drug such that HB1 association favors the N3 acceptor of a purine base, while the 5′-AGC target orients the drug in a way that directs the HB1 donor preferentially toward O2 of a pyrimidine base (Figures 12 and 13). Examina-

tion of model structures shows that the HB1- and HB2-governed drug positioning pivots the A−B scaffold about a central point approximating the site of HB4 hydrogen bonding.

In summary, the mostly rigid A−B-subunit scaffold binds to DNA through the C8-O-C23 oxygen acceptor at one end (HB2) and through the N12 hydrogen-bond donor (HB1) at the other end (Figures 12 and 13). In any of the four favored 5′-XGY sequences, this HB1 and HB2 coordination in combination with HB3 positions the carbinolamine C21-OH on top of the target 6G N2 donor, setting up another hydrogen bond, HB4. Hydrogen bonding through the C8-O-C23 oxygen (HB2) orients the drug differently depending on whether the amino group hydrogen-bond donor extends from the covalently modified strand or the opposite strand. Shifting the drug orientation at the methylenedioxy end of the B subunit in order to properly orient the guanine HN2 hydrogen bond affects the positioning of the axial HN12 substituent at the other end of the scaffold. Examination of model structures indicates that for the 5′-XGC sequences, 5′-GGC and 5′-AGC, wherein the guanine amino group originates from the unmodified strand, the HN12 donor is more suitably oriented to hydrogen bond with an O2 substituent on the pyrimidine base paired with X. In the Et 736− or Et 743−5′-AGC 12-mer duplex adduct, this is 20T O2 (Figures 12 and 13). For the 5′-XGG sequences, 5′-CGG and 5′-TGG, wherein the guanine amino group projects from the covalently modified strand, the HN12 donor is better positioned to hydrogen bond with N3 of the purine base paired with X (Figures 12 and 13). In either case, the reactive carbinolamine portion of the drug is optimally positioned relative to guanine N2, and C18-OH is positioned optimally relative to the backbone O3′ acceptor. This sequence-determined stable hydrogen-bonding complex is the only apparent candidate for the mechanism underlying the sequence specificity of Et 736. It is a mechanism that ensures the proper conformation for the previously proposed dehydration of the carbinolamine to iminium ion and subsequent formation of the covalent bond.[9]

A Comparison of Ecteinascidin Sequence Specificity with the Specificity of Other Carbinolamine-Based Minor Groove Alkylating Agents. The experimental and modeling data appear to lead to much better defined conclusions regarding ecteinascidin sequence specificity than those generated for the other carbinolamine-based agents, including saframycins (e.g., saframycin A, Figure 1), pyrrolo[1,4]benzodiazepines [P(1,4)B] (e.g., anthramycin, Figure 1), and naphthyridinomycins (e.g., naphthyridinomycin, Figure 1). The saframycins, which most closely resemble the ecteinascidins, yielded somewhat analogous results, but the quinonoid nature of saframycin A and B subunits must lower expectations regarding the degree of similarity. The examined saframycins were found in an extensive footprinting analysis to recognize primarily 5′-GGG and secondarily the 5′-GGPy series, especially the Py = C members, which were favored over Py = T. Exonuclease III stop assay independently confirmed the preference for 5′-GGG and 5′-GGC.[11]

Anthramycin and related P(1,4)B's, which have been more extensively examined than the ecteinascidins, show specificity that ranges from a most favored target, 5′-Pu-G-Pu-3′, to a least favored target, 5′-Py-G-Py-3′.[12] P(1,4)B interaction with DNA sequences is determined by a unique combination of helix and drug distortion energies and intermolecular binding energies. These intermolecular influences can range from predominantly hydrogen-bonding (e.g., anthramycin) to non-hydrogen-bonding interactions (e.g., tomaymycin) dependent on the DNA sequence. Even for tomaymycin, the relative importance of intermolecular binding energies and helix and drug distortion energies is sequence dependent.[12] Naphthyridinomycins have been the target of a modeling study that identified 5′-ATGCAT-3′ as the favored sequence based on net binding energy and DNA distortion energy calculations.[13]

For these other classes of carbinolamine-based agents, the factors governing sequence specificity are far less well characterized than those determining ecteinascidin specificity. A possible explanation for this is that in the case of the ecteinascidins base-directed hydrogen bonding seems to play an extraordinary role in determining specificity. This role is further magnified by the size of the ecteinascidin molecule, which permits this base-directed hydrogen bonding to be spread across a range of three base pairs.

Experimental Procedures

(a) Chemicals. Et 736 and Et 743 were gifts from PharmaMar. Reagents used to prepare the NMR buffer, sodium phosphate (99.99%) and sodium chloride (99.99%), were purchased from Aldrich. HPLC water and methanol were purchased from Baxter Scientific and Fisher, respectively.

(b) Oligonucleotide Preparation and Purification. Synthesis and purification methods used for the self-complementary 12-mer [d(CGTAAGCTTACG)$_2$] were previously described.[14]

(c) Adduct Preparation and Purification. Preparation of the Et 736 adduct involved the reaction of 20.0 mg of 5′-AGC 12-mer in buffer (800 μL of 100 mM KCl, 50 mM potassium phosphate, pH 6.9) with 2.0 mg of Et 736 in 150 μL of DMSO for 4 h at 25 °C. Unreacted drug was removed by centrifugation, and the DMSO was removed through repeated evaporation under vacuum using multiple additions of D_2O.

(d) Proton NMR Experiments. One- and two-dimensional 500 MHz ^{1}H and 202.44 MHz ^{31}P NMR data sets in H_2O- and D_2O-buffered solution were recorded on Bruker AMX 500 and Varian UNITYplus 500 FT NMR spectrometers. Proton chemical shifts of the ca. 6 mM buffered solution were recorded in parts per million (ppm) and referenced relative to external TSP (1 mg/mL) in D_2O (HOD signal was set to 4.751 ppm).

Phase-sensitive two-dimensional NOESY spectra (Bruker) were obtained at 27 °C (TPPI) for two mixing times, 100 and 200 ms, using a presaturation pulse to suppress the HOD signal. All spectra were acquired with 16 scans at each of 1024 t_1 values, spectral width of 10.002 ppm, and a repetition time of 10 s between scans. During data processing, a shifted squared sine bell function (shift = 90°) was used in both ω_1 and ω_2 dimensions. The FID in ω_1 was zero-filled to 2 K prior to Fourier transformation to give a 2K × 2K spectrum. Two-dimensional NOE spectra in 90% H_2O at 150 ms mixing time were recorded at 27 °C using the 1−1 echo read pulse sequence[15,16] with a 2.5 s pulse repetition time, a sweep width of 24.396 ppm, and a 90° pulse width of 28.75 ms.

Molecular Dynamics. The molecular dynamics hydrogen-bond analysis starting structures consisting of the Et 736 carbinolamine binding form complexed with DNA 12-mer duplexes and Et 736 covalent form attached at guanine N2 of 12-mer duplexes were generated by docking the drug in the minor groove, positioning C21 near the target guanine N2, and orienting the drug as previously documented.[8] Solvated rMD trajectories (100 ps) of the docked carbinolamine complex or covalent adduct were calculated using Watson−Crick hydrogen-bond angle and distance constraints[17,18] and distance constraints that restrained HB1, HB2, and HB3 donors and

(11) Rao, K. E.; Lown, J. W. *Biochemistry* **1992**, *31*, 12076−12082.
(12) Mountzouris, J. A.; Hurley, L. H. *Advances in DNA Sequence Specific Agents*; JAI Press Inc.: New York, 1992; Vol. 1, pp 263−392.
(13) Cox, M. B.; Arjunan, P.; Arora, S. K. *J. Antibiot.* **1991**, *44*, 885−894.
(14) Seaman, F. C.; Hurley, L. H. *Biochemistry* **1993**, *32*, 12577−12585.
(15) Sklenar, V.; Bax, A. *J. Magn. Reson.* **1987**, *74*, 469−479.
(16) Blake, P. R.; Summers, M. F. *J. Magn. Reson.* **1990**, *86*, 622−624.
(17) Schmitz, U.; Pearlman, D. A.; James, T. L. *J. Mol. Biol.* **1991**, *221*, 271−292.

acceptors within hydrogen-bonding distances. These latter constraints were derived from NOESY-based information documenting the proximity of Et 736 N12*H* to its HB1 acceptor, the HB2 guanine N2*H* donor to its drug acceptor, and the Et 736 HB3 C18-O*H* acceptor to its DNA backbone O3′ acceptor. Solvated rMD was performed on the 12-mer adducts by first positioning counterions (counterion charge = 1.0) 4.5 Å away from phosphorus and surrounding the minimized 12-mer adducts with 64 boxes of 216 Monte Carlo waters and limiting the number of water molecules to those whose oxygen atoms are within a 5.0 Å cutoff distance. Following Belly energy minimization for water molecules only, energy minimization to a maximum derivative of 0.01 Å was performed with distance restraints and hydrogen bond and angle restraints (maximum force constant = 10 kcal/mol·Å). Belly dynamics for water only (0−300 K; 10 ps; no restraints) was followed by rMD (solvated). rMD was conducted with no periodic boundary conditions, coupling with constant temperature heat bath, and SHAKE (removal of bond stretching freedom) option applied to all bonds. The rMD trajectory followed a temperature program beginning at 0 K and ramping to 600 K over a period of 60 ps. After 10 ps at 600 K, the program was ramped down to 300 K. The weights of the hydrogen-bond angle and distance restraints were modulated by multiplying the force constraints by a scaling factor. At 600 K, the restraint force constants reached their maximum values of 20 kcal/mol·Å^2 (distance restraints), 10 kcal/mol·Å^2 (hydrogen bond distance restraints), and 10 kcal/mol·rad^2 (hydrogen bond angle restraints) and were reduced to 0 during the ramping down to 300 K. The restart coordinates from these rMD trajectories were used as starting structures for a second solvated MD analysis used to evaluate hydrogen bonding. Molecular dynamics hydrogen-bond analysis was conducted using the MD conditions described above and the CARNAL module of AMBER 4.1.[19] After MD stabilization at 300 K (final 20 ps), trajectories from rMD analyses were examined for evidence of drug−DNA hydrogen bonds. Every 500 steps (0.5 ps) a coordinate set was generated for the Et 736−12-mer duplex binding complex or covalent adduct. For a given coordinate set, the CARNAL HBOND option specifies the H-bond distances and angles for each pair of donors and acceptors that meets a set of predefined criteria (distance = 3.4 Å, angle = 60°).

Acknowledgment. This research was supported by grants from the National Institutes of Health (CA-49751) and The Welch Foundation. We thank PharmaMar and Dr. Kenneth Rinehart for the gifts of Et 736 and Et 743 and Drs. Bob Moore and Richard Wheelhouse for helpful discussions. We also thank David Bishop for proofreading, editing, and preparing the final version of the manuscript.

Supporting Information Available: NOESY cross-connectivity walk data, ^{1}H chemical shift difference, phosphorus−proton correlation 2D spectrum, DQF-COSY cross-peak data, and structures of the Et 736 N12H and C8-O-C23 hydrogen bonds together with NOESY data (14 pages, print/PDF). See any current masthead page for ordering information and Web access instructions.

JA983091X

(18) Weisz, K.; Shafer, R. H.; Egan, W.; James, T. L. *Biochemistry* **1994**, *33*, 354−366.

(19) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Ferguson, D. M.; Seibel, G. L.; Singh, C.; Weiner, P. K.; Kollman, P. A. AMBER 4.1, University of California, San Francisco, 1995.