NMR investigation of Hoogsteen base pairing in quinoxaline antibiotic – DNA complexes: comparison of 2:1 echinomycin, triostin A and [N-MeCys³,N-MeCys⁷] TANDEM complexes with DNA oligonucleotides

Kenneth J. Address⁺ and Juli Feigon^{*}

Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California, Los Angeles, CA 90024-1569, USA

Received May 4, 1994; Revised and Accepted September 13, 1994

ABSTRACT

Hoogsteen base pairs have been demonstrated to occur in base pairs adjacent to the CpG binding sites in complexes of triostin A and echinomycin with a variety of DNA oligonucleotides. To understand the relationship of these unusual base pairs to the sequence specificity of these quinoxaline antibiotics, the conformation of the base pairs flanking the YpR binding sites of the 2:1 drug–DNA complexes of triostin A with [d(ACGTACGT)]₂ and of the TpA specific [N-MeCys³, N-MeCys⁷] TANDEM with [d(ATACGTAT)]₂ have been studied by ¹H NMR spectroscopy. In both the 2:1 triostin A – DNA complex and the 2:1 [N-MeCys³, N-MeCys⁷] TANDEM – DNA complex, the terminal A·T base pairs are Hoogsteen base paired with the 5' adenine in the *syn* conformation. This indicates that both TpA specific and CpG specific quinoxaline antibiotics are capable of inducing Hoogsteen base pairs in DNA. However, in both 2:1 complexes, Hoogsteen base pairing is limited to the terminal base pairs. In the 2:1 triostin A complex, the internal adenines are *anti* and in the 2:1 [N-MeCys³, N-MeCys⁷] TANDEM – DNA complex, the internal guanines are *anti* regardless of pH, which indicates that the central base pairs of both complexes form Watson – Crick base pairs. This indicates that the sequence dependent nature of Hoogsteen base pairing is the same in TpA specific and CpG specific quinoxaline antibiotic – DNA complexes. We have calculated a low resolution three-dimensional structure of the 2:1 triostin A – [d(ACGTACGT)]₂ complex and compared it with other CpG specific quinoxaline antibiotic – DNA complexes. The role of stacking in the formation of Hoogsteen base pairs in these complexes is discussed.

INTRODUCTION

Triostin A (Scheme 1) is a quinoxaline antibiotic that binds sequence specifically to double stranded DNA as a bis-intercalator at NCGN sites. Triostin A inhibits the growth of gram-positive bacteria by inhibiting transcription and is a potent agent against the proliferation of tumor cells (1). The structure of triostin A and the related echinomycin bound to the hexamer [d(CGTACG)]₂ were solved by Rich and co-workers (2,3) (here and for subsequent sequences, the base pairs flanked by the two intercalated quinoxaline rings of each bound drug are underlined). For each complex, the crystal structure showed that two drugs bind to the DNA duplex as bis-intercalators around the two CpG binding sites. The internal A·T base pairs, adjacent to the quinoxaline rings, formed Hoogsteen base pairs with the adenines in the *syn* conformation. In the crystal structure of triostin A bound to [d(GCGTACGC)]₂, the terminal G·C base pairs, as well as the A·T base pairs, form Hoogsteen base pairs at low pH (4,5).

The discovery of Hoogsteen base pairs in the crystal structure fostered a series of footprinting and NMR studies aimed at addressing the existence of Hoogsteen base pairs in quinoxaline–antibiotic DNA complexes in solution and at determining the relevance of these unusual base pairs to the biological action of these drugs *in vivo*. Mendel and Dervan (6) showed that binding of echinomycin to long sequences of DNA rendered the purines, particularly adenines, of base pairs adjacent to the CpG binding sites hyperreactive to both DNase I and diethyl pyrocarbonate (DEPC) and concluded that Hoogsteen base pairing was the structural cause for the hyperreactivity. Their conclusions were subsequently challenged by Waring and co-workers, who showed that substituting the N7 with C7 of a purine base, which prohibits the formation of Hoogsteen base pairs, had little effect on the binding properties of triostin A or echinomycin and did not reduce the hyperreactivity of the bound DNA to osmium tetroxide (OsO₄), DEPC or DNase I (7–9).

^{*}To whom correspondence should be addressed

⁺Present address: Department of Chemistry and Biochemistry, University of Colorado, Campus Box 215, Boulder, CO 80309-0215, USA

Scheme 1. Chemical structure of triostin A ($R_1 = \text{CH}_3$, $R_2 = \text{CH}_3$) and [N-MeCys³, N-MeCys⁷] TANDEM ($R_1 = \text{CH}_3$, $R_2 = \text{H}$).

NMR studies of echinomycin complexed with the DNA tetramers [d(ACGT)]₂ and [d(TCGA)]₂ and with the DNA octamers [d(ACGTACGT)]₂ and [d(TCGATCGA)]₂ showed that Hoogsteen base pairs form in the A·T base pairs of the echinomycin complexes containing 5' adenines but not in the A·T base pairs of the echinomycin complexes containing 5' thymines (10–12). Similar studies also showed that G·C base pairs flanking the CpG binding sites of the echinomycin [d(GCGC)]₂ complex form Hoogsteen base pairs at pH 5.1, while the flanking G·C base pairs in the [d(CCGG)]₂ complex are Watson–Crick at the same pH (13). However, for the 2:1 echinomycin–[d(ACGTACGT)]₂ complex, the central A·T base pairs are stably Hoogsteen base paired at low temperature only (11,12). In the 2:1 echinomycin–[d(ACGTATACGT)]₂ complex, the central A·T base pairs do not form Hoogsteen base pairs under any conditions studied (14). In another NMR study, it was shown that binding of the CpG specific UK-65,662 to the tetramer [d(ACGT)]₂ induces Hoogsteen base pairing of the A·T base pairs while binding to the hexamer [d(GACGTC)]₂ does not induce Hoogsteen base pairing (15). Taken together, the NMR evidence has established that Hoogsteen base pairing occurs only for terminal base pairs containing purines on the 5' side of the CpG binding sites of these complexes, with the exception of the central A·T base pairs of the 2:1 echinomycin–[d(ACGTACGT)]₂ complex at low temperature.

Although the sequence dependent nature of Hoogsteen base pairing has been established, the relationship between Hoogsteen base pairing and the sequence specificity of these drugs is still unknown. All studies of Hoogsteen base pairing to date have focused solely on the CpG-specific complexes. In the solution structure of the TpA sequence specific derivative [N-MeCys³, N-MeCys⁷] TANDEM (CysMeTANDEM) (Scheme 1) bound to [d(GATATC)]₂, the base pairing scheme of the A·T base pairs adjacent to the TpA binding site was also Watson–Crick (16). However, it is unknown whether the base pairing scheme of these A·T base pairs adjacent to the [N-MeCys³, N-MeCys⁷] TANDEM binding site depends on the TpA specificity of the drug.

On the basis of theoretical studies, Gallego *et al.* (17) have proposed that stacking interactions between a quinoxaline ring and a Hoogsteen A·T base pair are very favorable. Stacking may be the principal cause for the formation of Hoogsteen base pairs in these drug–DNA complexes.

To understand the role of sequence specificity and stacking interactions in the formation of Hoogsteen base pairs in these

Scheme 2. A schematic illustrating the orientation of two drug molecules in the two binding sites of the 2:1 triostin A–[d(ACGTACGT)]₂ complex and the 2CysMeTANDEM–[d(ATACGTAT)]₂ complex. The adenines that adopt the *syn* glycosidic torsional angles in the two complexes are underlined.

quinoxaline antibiotic–DNA complexes, we have examined the conformation of the base pairs flanking the YpR binding site of the two complexes 2:1 triostin A–[d(ACGTACGT)]₂ and 2CysMeTANDEM–[d(ATACGTAT)]₂ by ¹H NMR spectroscopy. The orientation of the two drugs in the two binding sites of each complex is displayed in Scheme 2. We have calculated a low resolution three-dimensional structure of this 2:1 triostin A–[d(ACGTACGT)]₂ complex. The binding and structural features of the 2:1 triostin A complex are compared with those of the 1:1 triostin A–[d(GACGTC)]₂ complex (18), the 2:1 echinomycin–[d(ACGTACGT)]₂ complex (11,12) and the crystal structure of the 2:1 triostin A–[d(CGTACG)]₂ complex (2,3,5). A discussion of why Hoogsteen base pairing in these complexes occurs in some cases and not in others is presented.

MATERIALS AND METHODS

Sample preparation

d(A₁C₂G₃T₄A₅C₆G₇T₈) and d(A₁T₂A₃C₄G₅T₆A₇T₈) were synthesized on the 10 μmole scale on an Applied Biosystems 381A DNA synthesizer using phosphoramidite chemistry and purified as previously described (19). Two DNA samples were prepared by dissolving the DNA in 150 mM NaCl and adjusting the pH to 6.5 and 5.5 (uncorrected meter reading). The samples were then transferred to NMR tubes and dried down under a stream of N₂ gas dissolved in 400 μl of 99.996 %D D₂O (Cambridge Isotopes Laboratories, Cambridge, MA) or 400 μl 90% H₂O/10% D₂O.

Triostin A was a gift from Dr M.Shin, Shionogi Research Laboratories, Osaka, Japan. CysMeTANDEM was a gift from Dr Richard K.Olsen, Utah State University. The 2:1 triostin A–[d(ACGTACGT)]₂ complex was formed by adding two equivalents of triostin A dissolved in acetonitrile directly to the free DNA solution in the NMR tube as described previously (18). The 2:1 CysMeTANDEM–[d(ATACGTAT)]₂ complex was prepared as previously described (19).

NMR spectroscopy

All NMR experiments were done on a General Electric GN500 (500.119 MHz, ¹H) spectrometer. Chemical shifts were referenced to the chemical shift of water, which had been

Figure 1. Proton NMR spectra of the aromatic region of $[d(ACGTACGT)]_2$ -trioestin A as a function of added drug concentration at 25°C in D_2O . DNA:trioestin A ratios of (A) 1:0, (B) 1:1, and (C) 1:2 are shown. Dashed lines connect the free DNA resonances to the free DNA resonances in the spectra of the 1:1 complex; solid lines connect the resonances of the fully saturated complex with the resonances of the fully saturated complex in the spectrum of the 1:1 complex. The spectra were acquired with a sweep width of 5000 Hz in 4K complex points. Data were line broadened by 3Hz prior to Fourier transformation.

previously calibrated relative to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). Phase sensitive nuclear Overhauser effect (NOESY) (20) spectra in D_2O were obtained using the method of States *et al.* (21) with preirradiation of the HDO peak during the recycle delay. Phase sensitive NOESY spectra in H_2O were obtained by replacing the last 90° pulse with a 11 spin echo pulse sequence and phase cycling appropriately in order to suppress the large water resonance (22). The carrier was centered at the water resonance, and the delay τ was adjusted so that the excitation maximum was at ~12 p.p.m. ($\tau = 60 \mu\text{sec}$, $\Delta = 50 \text{ msec}$). HOHAHA spectra were acquired using MLEV 17 mixing sequence and two 1.5 msec trim pulses for the spin lock (23). All 2D NMR spectra were processed on a Personal Iris 4D/25 (Silicon Graphics Incorporated) using FELIX 1.1 (Hare Research). Descriptions of the acquisition and processing parameters for each experiment are given in the figure captions.

Determination of the structure of the 2:1 triostin A- $[d(ACGTACGT)]_2$ complex

NOE cross peak volumes of the non-exchangeable proton resonances from NOESY spectra acquired with mixing times of 75 and 100 ms at 25°C were obtained by cross peak integration using the FELIX 2.05 software (Biosym Technologies, San Diego, CA). NOE cross peak volumes of exchangeable proton resonances were obtained by cross peak integration of a NOESY spectrum of the sample in H_2O acquired with $\tau_m = 75 \text{ msec}$ and a 11 echo observe pulse at 10°C. These volumes were converted into distances by the isolated two spin approximation, using the

Figure 2. Imino proton spectra as a function of temperature of (A) free $[d(ACGTACGT)]_2$, (B) 2trioestin A- $[d(ACGTACGT)]_2$. Assignments of the imino resonances are indicated. Spectra were acquired with a sweep width of 10,000 Hz in 8K complex points. Spectra were obtained with the 11 spin echo pulse sequence (Sklénář and Bax, 1987) to suppress the H_2O resonance. The delay τ was adjusted to give maximum excitation at approximately 12 p.p.m., $\tau = 60 \mu\text{s}$, $\Delta = 50 \mu\text{sec}$. All spectra were line broadened by 3 Hz prior to Fourier transformation.

CH5-CH6 distance of 2.44 Å as a calibration. The criteria used to apply the upper and lower bounds for the distance restraints have been described previously (16).

468 distance restraints, representing 2×234 asymmetric distance restraints, were used to generate fifteen starting structures from the metric matrix distance geometry substructure embedding routine in XPLOR 3.1 (24). In the substructure embedding routine, 302 of the 782 atoms of the drug-DNA complex were embedded into Cartesian space. Of these fifteen, the five lowest energy structures were refined using restrained molecular dynamics followed by energy minimization. During the calculations, the geometry of all deoxyribose sugars was restrained to be S-type (near C2' endo) except C₂ and C₆, which were restrained to be N-type (near C3' endo). Each starting structure was subjected to 2000 cycles of energy minimization to relieve bond length and angle violations. During the minimization, the potential energy terms for the NOE distance and the dihedral angle restraints were included in the total energy function. The resulting structures were then subjected to restrained molecular dynamics. A total of 20 ps of Verlet dynamics was performed as previously described (16). This was followed by another 2000 cycles of energy minimization. All dynamics and minimization routines were performed with XPLOR 3.1 (24).

RESULTS

Binding of triostin A to $[d(ACGTACGT)]_2$

One-dimensional proton (1D) spectra of the aromatic region of the DNA octamer $[d(ACGTACGT)]_2$ at different triostin A concentration ratios in D_2O at 25°C are given in Figure 1. Upon addition of drug, new resonances belonging to the quinoxaline protons of bound drug appear with new DNA aromatic resonances belonging to the 2:1 drug-DNA complex. At a 1:1

Figure 3. (A) NOESY spectrum of the 2:1 triostin A– $[d(ACGTACGT)_2]$ complex $\tau_m = 150$ msec. Several important crosspeaks are boxed: the H8–H1' crosspeak for A₁ and A₅ (A), the crosspeaks for Val₁CH₃ to A₁H8 and Val₂CH₃ to A₅H2 (B). The spectrum was acquired with a sweep width of 5000 Hz in both dimensions. 360 t_1 blocks of 2K points and 32 scans were acquired. The data in both dimension were apodized by a skewed sine-bell squared function phase shifted by 60°. 360 points were apodized in both dimensions. Data in t_1 were zero-filled to 1K points prior to Fourier transformation. (B) Two schematics showing the interactions of the valine methyl protons with the aromatic protons of an adenine adopting an *anti* (right) and a *syn* (left) glycosidic torsional angle in a quinoxaline antibiotic–DNA complex. In each schematic, both interproton distances between the ValCH₃ protons and the adenine aromatic protons are given.

ratio of triostin A to $[d(ACGTACGT)_2]$, shown in Figure 1B, all the resonances correspond to either the 2:1 complex shown in Figure 1C, indicated by solid lines, or the free DNA shown in Figure 1A, indicated by dashed lines. Similar results are observed when either $[d(ACGTACGT)_2]$ or $[d(ACGTATACGT)_2]$ is titrated with echinomycin (11,12,14). This shows that triostin A and echinomycin bind similarly in a positive cooperative manner, i.e., either no drug or two drugs bind to the DNA. However, echinomycin forms a tighter binding complex with $[d(ACGTACGT)_2]$ than triostin A. All of the resonances of the free DNA resonances disappear at echinomycin:DNA ratios of 2:1 and greater (11, 12), while weak signals from the free DNA are still observed at 1:1 triostin A:DNA ratios (Figure 1C).

Comparison of the integrated intensities of the resonances of the 2:1 triostin A– $[d(ACGTACGT)_2]$ complex with those of the free DNA resonances indicates that only approximately 95% of the available CpG binding sites are bound by triostin A. In a 1:1 mixture of triostin A with the DNA hexamer $[d(GACGTC)_2]$, only 70% of the available CpG binding sites are bound by the drug (18). This indicates that the binding of triostin A to $[d(ACGTACGT)_2]$ is substantially tighter than to $[d(GACGTC)_2]$.

Imino proton spectra of the 2:1 triostin A– $[d(ACGTACGT)_2]$ complex at different temperatures

Figure 2 shows the 1-dimensional proton NMR spectra of $[d(ACGTACGT)_2]$ (panel A) and the 2:1 triostin A– $[d(ACGTACGT)_2]$ complex (hereafter referred to as the 2:1 triostin A complex) in H₂O (panel B) at four different temperatures. The chemical shifts of imino protons of the

complex, shown in B, are shifted upfield ~ 1 p.p.m. relative to the free DNA, shown in A. This upfield shift indicates intercalative binding of the DNA by triostin A (25). Weak signals of imino resonances of uncomplexed DNA are also detected (Figure 2B). The resonances of G₃, G₇ and T₄ imino of the free DNA are sharp while the T₈ imino is broad because of its faster exchange rate with H₂O (Figure 2A). In the complex, the T₈ resonance becomes slightly sharper relative to those of the free DNA. The T₄ resonances of the complex are broader than those of the free DNA. The linewidth at half height of the T₄ imino peak of the complex is approximately three times greater than the linewidth at half height of the T₄ imino peak of the free DNA. Comparison of the linewidth at half height of the A₁H8 aromatic resonance in both the free DNA and the complex indicates that drug binding contributes to only about 15% of the broadening of the T₄ imino of the complex. Thus, the remaining contribution to broadening of the T₄ imino resonance is due to the T₄ iminos protons of the complex exchanging more rapidly with H₂O than those of the free DNA (Figure 1A). A similar effect on the T₄ and T₈ imino resonances is observed in the H₂O spectrum of the 2:1 echinomycin– $[d(ACGTACGT)_2]$ complex (hereafter referred to as the 2:1 echinomycin complex) (11,12).

Hoogsteen versus Watson–Crick base pairing in the 2:1 triostin A complex

A NOESY spectrum of the 2:1 triostin A complex in D₂O at 25°C is shown in Figure 3A. Many of the cross peaks in this spectrum are similar to those in the D₂O and H₂O NOESY spectra of the 2:1 echinomycin complex. Therefore, the assignments of most resonances of the 2:1 triostin A and the 2:1 echinomycin

Figure 4. A stereo view of one of the five refined structures of the 2:1 triostin A-[d(ACGTACGT)]₂ complex. The figure shows two molecules of triostin A, drawn with black bonds and white atoms, sitting in each of the CpG binding sites of the DNA octamer, drawn with white bonds and white atoms.

complexes were similar. The A₁H8 resonance has a cross peak to the Val methyl resonance of triostin A (Figure 3A, Box B). In addition, the A₁H8-A₁H1' cross peak (Figure 3A, Box A) is much more intense than the other H8/H6 to H1' cross peaks. When A is *syn*, the Val methyl is much closer to the AH8 proton than the AH2 proton (Figure 3B). This is because the AH8 of the *syn* A is positioned in the minor groove of the drug-DNA complex. The strong A₁H8 to A₁H1' and the A₁H8 to Val methyl cross peaks indicate that A₁ is *syn*, characteristic of Hoogsteen base pairing between the A₁ and the T₈ bases. Both cross peaks are observed in the D₂O NOESY spectrum of the 2:1 echinomycin complex, in which the terminal A₁T₈ base pairs are Hoogsteen base paired (12). The base pairing scheme of the A₁T₈ base pair was corroborated by T₈ imino to A₁H8 cross peak in a NOESY IN H₂O at 1°C (data not shown). By contrast, the A₅H8 resonance does not have a crosspeak to the Val methyl. Instead, the A₅H2 has a cross peak to the Val methyl (Figure 3A, Box B). Furthermore, the A₅H8-A₅H1' cross peak (Figure 3A, Box A) is much less intense than the A₁H8-A₁H1' cross peak. In the NOESY spectrum of the triostin A-[d(GACGTC)]₂ complex, in which the A₂ is *anti* (discussed in greater detail below), there is a similar cross peak between the A₂H2 and Val methyl protons (18). This indicates that the A₅ base is *anti* (Figure 3B), characteristic of Watson-Crick base pairing. By contrast, the A₅ residue in the 2:1 echinomycin complex, which contains a cross peak between the A₅H8 and the Val Me, is *syn* (12). Complete H' assignments are given in (28).

Structure calculation of the 2:1 triostin A complex

Fifteen starting structures of the 2:1 triostin A-[d(ACGTACGT)]₂ complex were generated by substructure embedding of 468 interproton distances with the distance geometry subroutine in the program XPLOR (24). Five of these fifteen structures

Figure 5. A stacked plot of the base-H1' region of the NOESY spectrum ($\tau_m = 150$ msec) of [d(ATACGTAT)]₂-2CysMeTANDEM, 2 mM complex, 150 mM NaCl, pH 6.5 (A) and 5.5 (B). Assignments of the A₁, C₄, and G₅ base-H1' and CH5-CH6 crosspeaks are indicated. The spectrum was acquired with a sweep width of 5000 Hz in both dimensions, 300 t_1 Values of 64 scans and 2K complex points. 300 points were apodized in both dimensions with a sinebell phase shifted by 45°. Data in t_1 were zero filled to 1K prior to Fourier transformation.

had no distance violations of greater than 0.5 Å and were refined by restrained molecular dynamics and energy minimization. No refined structure contained a distance violation greater than 0.4 Å. A stereo diagram of one of the five refined structures is shown in Figure 4. The DNA in the complex retains its dyad axis of symmetry. The average pairwise RMSD for these five refined structures is 1.79 ± 0.27 Å calculated using all the heavy atoms of the complex. The most defined portions in each symmetrical half of the molecule are the drug, the terminal A·T base pairs, and the two G·C base pairs of the binding site. The average RMSD for these portions of the molecule taken together is 1.18 ± 0.20 Å. The least defined portion of the structure are the two internal A·T base pairs due to fewer distance restraints. The average pairwise RMSD for these two base pairs is 1.51 ± 0.38 Å. As a result, the orientation of the top and bottom halves of the structure with respect to each other is not well defined.

Hoogsteen versus Watson-Crick base pairing in a CysMeTANDEM-DNA complex

Previous studies of Hoogsteen base pairing in quinoxaline antibiotic DNA complexes have focused primarily on the CpG specific triostin A and echinomycin complexes. Here we extend these investigations to the sequence specific nature of Hoogsteen base pairing in TpA specific complexes by examining a 2:1 drug-DNA complex of CysMeTANDEM with the [d(ATACGTAT)]₂. Figure 5A and B shows the stacked plots of the base-H1' regions of the NOESY spectra of the 2:1 CysMeTANDEM-[d(ATACGTAT)]₂ complex (hereafter referred to as the 2:1 CysMeTANDEM complex) in D₂O at pH

Figure 6. Ball and stick diagram of one-half of the 2:1 triostin A-[d(ACGTACGT)]₂ complex. The DNA is drawn with white bonds and white atoms. The A₁ residue is shown in the *syn* conformation, while the A₅ residue is shown in the *anti* conformation. The amino acids residues (drawn in black with white atoms) of the drug are labeled. The two types of intermolecular hydrogen bonds are indicated with dashed lines and the atoms that form these hydrogen bonds are shaded in black.

6.5 and 5.5. In each spectrum, the A₁H8-A₁H1' cross peak is much more intense than the other base-H1' cross peaks, including the G₅ residue, and of comparable intensity to the CH5-CH6 cross peak. Furthermore, the NOE cross peak patterns from the A₁H8 resonance in this complex are similar to the patterns from the A₁H8 proton of the 2:1 triostin A complex. This shows that the A₁ residue is *syn*, indicative of Hoogsteen base pairing between the A₁ and T₈ bases. The central C₄.G₅ base pairs in the 2:1 complex are stable Watson-Crick base pairs at pH 6.5. In order to see if protonation of the cytosine bases results in Hoogsteen base pairs in the central G·C base pairs, we lowered the pH to 5.5. Comparison of the base-H1' cross peak of G₅ at pH 6.5 and pH 5.5 indicate that the conformation and stability of the internal G·C base pairs do not change when the pH is decreased. That is, lowering the pH does not induce Hoogsteen base pairs to form in the C₄.G₅ base pairs.

DISCUSSION

Structural features of the 2:1 triostin A complex

The top portion of the 2:1 triostin A complex is shown in Figure 6. The view is into the minor groove of the DNA. As anticipated, triostin A binds in the minor groove of the DNA as a bis-intercalator around the two G·C base pairs. The upper quinoxaline ring is stacked directly under the *syn* A₁ base of the Hoogsteen A₁.T₈ base pairs and directly over the C₂.G₇ base pair. The lower quinoxaline ring is stacked directly over the *anti* A₅ base of the T₄.A₅ base pair and directly under the G₃.C₆ base pair. The most unwound portion of the complex is at the CpG

binding site, which has an average helical twist angle of -7.0° . The residue steps above the two G·C base pairs have average helical twist angles of $\sim 25^\circ$. Both CpG base pairs buckle inward by about 20° . Both A·T base pairs adjacent to the CpG binding site have an outward buckle.

Four intermolecular hydrogen bonds, two between the Ala NH and N3 of guanine and two between Ala CO and the guanine 2-amino groups, are observed. The average distance between the Ala CO and the 2-amino groups of guanine and between the Ala NH and the N3 of guanine are 2.0 ± 0.1 Å. Strong van der Waals interactions are observed between the Ala methyl of triostin A and the C1' and C2' of the cytosine sugar and the O4' of the guanine sugar. The Val N-methyl groups point directly into the minor groove of the DNA where they have strong van der Waals interactions with the 2-amino groups of guanine and O2 carbonyl of cytosine of each of the CpG binding sites of the DNA.

Comparison with the solution structure of the 1:1 triostin A-[d(GACGTC)]₂ complex

The structural features of the 2:1 complex of triostin A with [d(ACGTACGT)]₂ are very similar to those observed in the structure of the 1:1 triostin A-[d(GACGTC)]₂ complex (18). Unwinding and buckling of both CpG base pairs and the nature of the drug-DNA interactions are similar in both complexes. The average pairwise RMSD calculated for the heavy atoms of the drug, the CpGpT step between the 1:1 complex (residues 3 to 5) and the 2:1 complex (residues 2 to 4) is 1.2 ± 0.3 Å.

The most noticeable difference between the 1:1 and 2:1 triostin A complexes is in the conformation of the A·T base pairs adjacent to the CpG binding sites. There is also a difference in the binding affinity of triostin A for the single binding site and two binding site DNA oligonucleotides. The stronger binding of triostin A to [d(ACGTACGT)]₂ than to [d(GACGTC)]₂ is also further evidence for cooperative binding of triostin A to [d(ACGTACGT)]₂. In the structure of the triostin A-[d(GACGTC)]₂ complex, the A₂.T₅ base pairs are stable Watson-Crick base pairs. In the structure of the 2:1 triostin A complex, the terminal A₁.T₈ base pairs are stable Hoogsteen base pairs, while the internal A₅.T₄ base pairs appear to be Watson-Crick base pairs. Thus, the stacking interactions between the quinoxaline rings and the A·T base pairs adjacent to the binding sites of the two complexes are different. Since the structure of CpG step with the bound drug in both complexes is very similar, this suggests that cooperative binding of triostin A to [d(ACGTACGT)]₂ must be caused by more favorable stacking interactions between the quinoxaline rings and the A·T base pairs in the 2:1 triostin A complex than in the 1:1 triostin A complex.

Comparison with the crystal structure of the 2:1 triostin A-[d(CGTACG)]₂ complex

The solution structure of the 2:1 triostin A complex and the crystal structure of 2:1 triostin A-[d(CGTACG)]₂ complex have similar structural features with only two noticeable differences. First and foremost, in the crystal structure of the 2:1 triostin A-[d(CGTACG)]₂ complex, the central A·T base pairs are Hoogsteen base paired with the internal adenines in the *syn* conformation (2,3,5). This is the conformation of the terminal A·T base pairs but not the internal A·T base pairs in the solution structure of the 2:1 triostin A complex. Second, in the crystal structure of triostin A bound to [d(CGTACG)]₂, only one of the two possible intermolecular hydrogen bonds between the Ala CO

and the 2-amino groups of guanine is observed per CpG binding site, indicating that the two Val N-methyl groups are positioned differently in the symmetrical halves of the molecule (2,3,5). In each CpG binding site of the solution structure of the 2:1 triostin A complex, evidence for both intermolecular hydrogen bonds between the Ala CO and the 2-amino groups of guanine is observed and the two Val N-methyl groups are symmetrically equivalent. Recently, it was shown that elimination of one or both 2-amino groups of guanine from the CpG binding site of DNA, which is achieved by substituting an inosine for a guanine, abolishes the sequence specific binding of echinomycin to DNA (Marchand *et al.*, 1992). Recent NMR studies also showed that triostin A cannot bind sequence specifically to DNA containing a CpI step (18). Therefore, both intermolecular hydrogen bonds between the Ala carbonyls and the guanine 2-amino groups in each CpG binding site are formed in the sequence specific binding of triostin A to $[d(ACGTACGT)]_2$ in solution.

In all crystal structures of echinomycin and triostin A complexes, the alanine CO closer to the central A·T base pairs does not form an intermolecular hydrogen bond with the guanine 2-amino group (Wang *et al.*, 1984; Ughetto *et al.*, 1985). In the crystal structure of the 2:1 triostin A- $[d(GCGTACGC)]_2$ complex, the intermolecular hydrogen bond between the Ala CO and the guanine 2-amino groups occurs on the alanyl residue that is closer to the terminal G·C base pair (Quigley *et al.*, 1986; Wang *et al.*, 1986). It is possible that in all these crystal structures, internal Hoogsteen A·T can exist only because the second intermolecular hydrogen bond between the Ala CO and the guanine 2-amino groups is broken. This could be caused by crystal packing forces.

Comparison with the NMR studies of the 2:1 echinomycin complex

The binding and structural characteristics of the 2:1 triostin A and echinomycin complexes are very similar. Both triostin A and echinomycin bind to the $[d(ACGTACGT)]_2$ in a cooperative manner (11,12). However, echinomycin binds more tightly to $[d(ACGTACGT)]_2$ than triostin A. 100% complex formation occurs at an echinomycin to DNA ratio of 2:1, while only 95% complex formation occurs at the same ratio of triostin A to DNA. This is consistent with the footprinting data that indicates that echinomycin binds more tightly to CpG steps than triostin A (Portugal *et al.*, 1988).

Both triostin A and echinomycin cause the terminal A·T base pairs of $[d(ACGTACGT)]_2$ to adopt the stable Hoogsteen base pairs. Even though the internal A·T base pairs of the two complexes are unstable, the conformation of these base pairs is different in the two complexes. In the 2:1 echinomycin complex, the A₅ is *syn* at 25°C and the T₄.A₅ base pairs form Hoogsteen base pairs at low temperature (12). In the 2:1 triostin A complex, A₅ is *anti* at low temperature. Therefore, the central A·T base pairs are not Hoogsteen base pairs even at low temperature.

The relationship between Hoogsteen base pairing and sequence specificity

The terminal A·T base pairs of 2:1 triostin A complex and 2:1 CysMeTANDEM complex adopt stable Hoogsteen base pairs (Scheme 3), similar to the terminal A·T base pairs of the 2:1 echinomycin complex and the echinomycin- $[d(ACGT)]_2$ complex (Gao and Patel 1988; Gilbert *et al.*, 1989; Gilbert and Feigon, 1991). Comparison of the NOE cross peaks involving the terminal AH8 proton of the 2:1 triostin A, echinomycin and

CysMeTANDEM complexes indicates that the stacking interactions between the quinoxaline rings and the terminal A·T base pairs are similar in all three complexes. Therefore, Hoogsteen base pairing occurs independent of the sequence specificity of these quinoxaline antibiotics.

In both TpA specific complexes and CpG specific complexes, the sequence-dependent nature of Hoogsteen base pairing is the same. The terminal base pairs in the 2:1 echinomycin, triostin A and CysMeTANDEM complexes occur in Hoogsteen form only at 5' terminal base at purines. The only exception to this is the central A·T base pairs in the 2:1 echinomycin complex, which adopt the Hoogsteen form at low temperature (11,12).

The role of stacking interactions in the formation of Hoogsteen base pairs

Two theoretical studies have been undertaken to understand the role of stacking in the formation of Hoogsteen base pairs in triostin A and echinomycin complexes. (17,26). In the later study, Gallego *et al.* (17) determined that the quinoxaline rings stack more favorably with Hoogsteen A·T base pairs than with Watson-Crick A·T base pairs. This is because the dipole moment of a Hoogsteen A·T base pair and the stacked quinoxaline ring align antiparallel to each other, which produces a favorable dipole-dipole interaction. The dipole moment of a Watson-Crick A·T base pair is fairly weak, which produces a much weaker stacking interaction with the quinoxaline ring. This would explain why the terminal A·T base pairs of the 2:1 complexes of echinomycin and triostin A with $[d(ACGTACGT)]_2$ and of the 2:1 complex of CysMeTANDEM with $[d(ATACGTAT)]_2$ are Hoogsteen base-paired. This would also explain why triostin A binds more tightly to $[d(ACGTACGT)]_2$ than to $[d(GACGTC)]_2$.

However, this does not explain why the A₂.T₅ base pairs remain stable Watson-Crick base pairs in the 1:1 complex. One possible reason is that if these base pairs were Hoogsteen base paired, their dipole moments would be parallel to the dipole moments of the terminal G·C base pairs. This would result in unfavorable interactions. Hence, the stabilization provided by favorable stacking interactions between the Hoogsteen A·T base pairs and the quinoxaline rings would be offset by unfavorable interactions between the Hoogsteen A·T base pairs and the Watson-Crick G·C base pairs. This would also explain why the bound derivative UK-65,662 induces Hoogsteen base pairs in $[d(ACGT)]_2$ but not in $[d(GACGTC)]_2$ (15).

The calculations of Gallego *et al.* (17) predict that the internal A·T base pairs of the 2:1 triostin A and echinomycin complexes should also adopt Hoogsteen base pairs. The reason is that the dipole moment of the internal Hoogsteen A·T base pairs, in addition to being antiparallel to the stacked quinoxaline rings, are antiparallel to each other. This enhances the stacking interaction between the central A·T base pairs in the 2:1 triostin A and echinomycin complexes. This enhancement, according to Gallego *et al.* (17), contributes to the binding cooperativity observed in both complexes. However, Hoogsteen base pairs for these internal A·T do not occur for the 2:1 triostin A complex and occur only for the 2:1 echinomycin $[d(ACGTACGT)]_2$ complex at low temperature.

Binding of echinomycin and triostin A to $[d(ACGTACGT)]_2$ causes the T₄ iminos to exchange more quickly with H₂O (Figure 2). This suggests that triostin A and echinomycin binding causes an increase in the rate of opening of the central A·T base pairs from the rate of opening of these base pairs in the free DNA.

As a result, both triostin A and echinomycin binding should lower the activation energy enough to allow the central A·T base pairs to undergo the Watson–Crick to Hoogsteen transition. However, the fact that the internal adenine of the 2:1 echinomycin complex is *syn* and the same adenine in the 2:1 triostin A complex is *anti*, indicates that echinomycin binding to [d(ACGTACGT)]₂ permits this *anti* to *syn* transition while triostin A binding does not. This could be due to triostin A having a disulfide cross bridge and echinomycin having a thioacetal cross bridge. A comparison of the crystal structure of the 2:1 echinomycin–[d(CGTACG)]₂ to the crystal structure of the 2:1 triostin A–[d(CGTACG)]₂ complex reveals that the difference in the chemical structure of the cross bridge causes the position of the bound echinomycin in the complex to be slightly different from the position of the bound triostin A in the complex (2). As a result, the stacking and drug–DNA van der Waals interactions in the 2:1 echinomycin complex are slightly different from the stacking and van der Waals interactions in the 2:1 triostin A complex. This difference in stacking interactions may lead to the difference in the binding constant for the two drugs.

Relevance to chemical and enzymatic footprinting studies

Several laboratories have demonstrated that DNA bound by both echinomycin and triostin A is rendered hypersensitive to DNase I and cleavage reagents, especially at runs of adenines on the 3' side of the CpG binding sites (6,7). As indicated by our results and the results of other NMR studies, Hoogsteen base pairs do not seem to form in this situation (11–13,15). Modification of the N7 with C7 of purines bases, which abolishes the ability of the DNA to form Hoogsteen base pairs, had no effect on the DNA binding properties of echinomycin (9). Therefore, Hoogsteen base pairing is most likely not the cause of triostin A or echinomycin induced hypersensitivity and therefore probably does not play a role in the biological action of these drugs *in vivo*.

It has been proposed that unwinding of the helix by echinomycin and triostin A binding is the principal structural cause for the footprinting hypersensitivity. However, Waring and co-workers pointed out that unwinding cannot completely explain the increase in hypersensitivity because monofunctional intercalators such as ethidium do not affect DEPC mediated cleavage patterns of DNA (7). However, the failure of ethidium to induce hypersensitivity in a similar manner as triostin A and echinomycin could be related to the lack of DNA binding specificity of ethidium. It was shown that binding of actinomycin, an antibiotic that binds sequence specifically to DNA at GpC sites by intercalation, can induce DNase I hyperreactivity in certain sequences flanking the GpC binding site (27). The failure of ethidium to induce this hypersensitivity could also be due to the rapid kinetics of DNA binding for the monointercalator compared to echinomycin and triostin A. Since buckling of the bases is the only other principal DNA structural change that occurs as a result of drug binding, another explanation for the triostin A and echinomycin induced hypersensitivity could be the buckling of the bases adjacent to the CpG binding sites. It is possible that a combination of buckling of the bases and unwinding of the helix of the DNA could produce the patterns observed in the footprinting of DNA bound by triostin A and echinomycin.

ACKNOWLEDGEMENTS

We wish to thank Peter Schultze for comments on the manuscript. This work was supported by grants from NIH (R01 GM 37254)

and NSF Presidential Young Investigator Award with matching funds from AmGen Inc., Monsanto Co., and Sterling Winthrop Drug Inc. to J.F.

REFERENCES

1. Katagiri, K., Yoshida, T. and Yoshida, K. (1974) In J. W. Corcoran and F. E. Hahn (ed.), *Antibiotics III: Mechanism of Action of Antimicrobial and Antitumour Agents*, Springer, Berlin, Vol. pp. 234–251.
2. Wang, A.H.-J., Ughetto, G., Quigley, G.J., Hakoshima, T., van der Marel, G.A., van Boom, J.H. and Rich, A. (1984) *Science*, **225**, 1115–1121.
3. Ughetto, G., Wang, A.H.-J., Quigley, G.J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1985) *Nucleic Acids Res.*, **13**, 2305–2323.
4. Quigley, G.J., Ughetto, G., van der Marel, G.A., van Boom, J.H., Wang, A.H.-J. and Rich, A. (1986) *Science*, **232**, 1255–1258.
5. Wang, A.H.-J., Ughetto, G., Quigley, G.J. and Rich, A. (1986) *J. Biomol. Struct. Dyn.*, **4**, 319–342.
6. Mendel, D. and Dervan, P.B. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 910–914.
7. Portugal, J., Fox, K.R., McLean, M.J., Richenberg, J.L. and Waring, M.J. (1988) *Nucleic Acids Res.*, **16**, 3655–3670.
8. McLean, M.J. and Waring, M.J. (1988) *J. Mol. Recognition*, **1**, 138–151.
9. Sayers, E.W. and Waring, M.J. (1993) *Biochemistry*, **32**, 9094–9107.
10. Gao, X. and Patel, D.J. (1988) *Biochemistry*, **27**, 1744–1751.
11. Gilbert, D.E., van der Marel, G.A., van Boom, J.H. and Feigon, J. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 3006–3010.
12. Gilbert, D.E. and Feigon, J. (1991) *Biochemistry*, **30**, 2483–2494.
13. Gao, X. and Patel, D.J. (1989) *Q. Rev. Biophys.*, **22**, 93–138.
14. Gilbert, D.E. and Feigon, J. (1992) *Nucleic Acids Res.*, **20**, 2411–2410.
15. Searle, M.S. and Wickham, G. (1990) *FEBS Lett.*, **272**, 171–174.
16. Address, K.J., Sinsheimer, J.S. and Feigon, J. (1993) *Biochemistry*, **32**, 2498–2508.
17. Gallego, J., Ortiz, A.R. and Gago, F. (1993) *J. Med. Chem.*, **36**, 1548–1561.
18. Address, K.J. and Feigon, J. (1994) *Biochemistry*, **33**, 12397–12404.
19. Address, K.J., Gilbert, D.E., Olsen, R.K. and Feigon, J. (1992) *Biochemistry*, **31**, 339–350.
20. Kumar, A., Ernst, R.R. and Wüthrich, K. (1980) *Biochemistry Biophys. Res. Comm.*, **95**, 1–6.
21. States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Res.*, **48**, 286–292.
22. Sklenár, V. and Bax, A. (1987) *J. Magn. Reson.*, **74**, 469–479.
23. Bax, A. and Davis, D.G. (1985) *J. Magn. Res.*, **65**, 355–360.
24. Brünger, A.T. (1992) X-PLOR (Version 3.1) Manual. Yale University Press, New Haven and London.
25. Feigon, J., Denny, W.A., Leupin, W. and Kearns, D.R. (1984) *J. Med. Chem.*, **27**, 450–465.
26. Singh, U.C., Pattabiraman, N., Langridge, R. and Kollman, P.A. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 6402–6406.
27. Bishop, K.D., Borer, P.N., Huang, Y.-Q., and Lane, M.J. (1991) *Nucleic Acids Res.*, **19**, 871–874.
28. Address, K.J. (1994) Ph.D Thesis. University of California, Los Angeles.