1566–1573 *Nucleic Acids Research, 1996, Vol. 24, No. 8*

An NMR study of [d(CGCGAATTCGCG)]$_2$ containing an interstrand cross-link derived from a distamycin–pyrrole conjugate

Patricia A. Fagan, H. Peter Spielmann[+], Snorri Th. Sigurdsson[1,§], Stacia M. Rink[1], Paul B. Hopkins[1] and David E. Wemmer*

Structural Biology Division, Lawrence Berkeley National Laboratory and Department of Chemistry, University of California, Berkeley, CA 94720, USA and [1]Department of Chemistry, University of Washington, Seattle, WA 98195, USA

Received October 2, 1995; Revised and Accepted February 23, 1996

ABSTRACT

Minor groove binding compounds related to distamycin A bind DNA with high sequence selectivity, recognizing sites which contain various combinations of A·T and G·C base pairs. These molecules have the potential to deliver cross-linking agents to the minor groove of a target DNA sequence. We have studied the covalent DNA–DNA cross-linked complex of 2,3-bis(hydroxymethyl)pyrrole–distamycin and [d(CGCGAATTCGCG)]$_2$. The alkylating pyrrole design is based on the pharmacophore of mitomycin C and is similar in substructure to another important class of natural products, the oxidatively activated pyrrolizidine alkaloids. Ligand–DNA NOEs confirm that the tri(pyrrolecarboxamide) unit of the ligand is bound in the minor groove of the central A+T tract. Unexpectedly, it is shifted by 1 bp with respect to the distamycin A binding site on this DNA sequence. The cross-link bridges the 2-amino position of two guanine residues, G$_4$ and G$_{22}$. The C$_3$·G$_{22}$ and G$_4$·C$_{21}$ base pairs exhibit Watson–Crick base pairing, with some local distortion, as evidenced by unusual intensities observed for DNA–DNA NOE cross-peaks. The model is compared with a related structure of a cross-linked mitomycin C:DNA complex.

INTRODUCTION

Several effective anti-tumor substances function by alkylating DNA (1,2). Although the DNA sequence selectivity of these agents can be high, the length of the sequence recognized is invariably quite small. For example, mitomycin C, which has been used clinically to treat a variety of cancers, selectively cross-links guanines in the sequence CG to form a cross-link in the minor groove (3–6). It is possible that a molecule which could target a very restricted set of binding sites might be therapeutically more valuable. The cross-linking agent described herein was designed as an initial step toward the goal of achieving selectivity for longer DNA sequences.

Ligands based on the oligopeptide antibiotics distamycin and netropsin can be designed to bind specific sequences of DNA. These natural products are di- and tripeptides that bind with submicromolar affinity in the minor groove at sites containing at least four successive A·T base pairs (7–12). Investigations of these molecules in complexes with DNA by X-ray crystallography (11,13), NMR spectroscopy (9,14–17) and thermodynamic methods (18,19) reveal that hydrogen bonding, van der Waals interactions and electrostatics all contribute to their binding affinity and specificity.

Structural studies of complexes of netropsin and distamycin with various DNA oligomers led to the design of analogs which have been successful in recognizing specific mixed A+T- and G+C-containing sequences (20–26). Most recently an analog has been developed which binds an entirely G+C core site (22). This well-characterized class of molecules has the potential to target a variety of DNA sequences with predictable binding behavior and could deliver a cross-linking agent to the minor groove of the target DNA sequence.

2,3-bis(Hydroxymethyl)pyrrole, an inefficient DNA cross-linking agent alone, has been synthetically tethered to a distamycin analog to form a 2,3-bis(hydroxymethyl)pyrrole–distamycin conjugate (XL-Dst) (Fig. 1; 27). The bis(hydroxymethyl)pyrrole function mimics in part the functionality present in reductively activated mitomycins or oxidatively activated pyrrolizidine alkaloids (Fig. 2; 28). This substructure cross-links the minor groove amino group of guanines of the sequence d(CG), albeit slowly. The conjugation of this agent to distamycin results in a molecule shown to have a high efficiency of interstrand cross-linking in both a linearized plasmid and synthetic DNA oligomers (27), and specificity for sites bearing the distamycin binding sequence adjacent to a pyrrole cross-linking sequence. The distamycin portion of the ligand is important: the conjugate is 1000-fold more active than 2,3-bis(hydroxymethyl)-1-methylpyrrole alone (27).

Other DNA alkylating reagents have been conjugated to members of the distamycin/lexitropsin class of molecules. Most form DNA monoadducts (29–34), with some forming the generally more cytotoxic interstrand cross-links (26,35). Many of these alkylating agents alone have no sequence preference, but in

* To whom correspondence should be addressed

Present addresses: [+]Department of Biochemistry, University of Kentucky, Lexington, KY 40536, USA and [§]Max-Planck-Institut für Experimentelle Medizin, D-37075 Göttingen, Germany

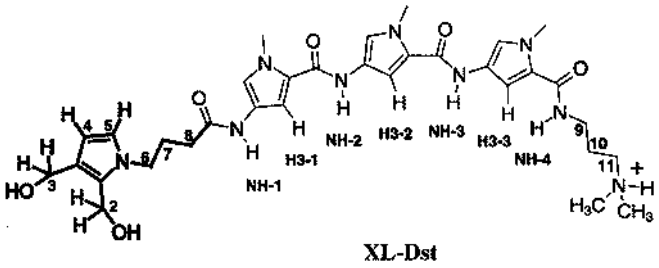

Figure 1. Structure of the 2,3-bis(hydroxymethyl)pyrrole–distamycin conjugate (XL-Dst). The cross-linking pyrrole unit is shown by heavy lines.

the conjugated molecule alkylate sequence specifically and in the minor groove. To our knowledge no detailed structural studies have been done on either covalent or non-covalent complexes of any of these compounds with DNA.

We have used NMR spectroscopy to determine the binding mode of a 2,3-bis(hydroxymethyl)pyrrole–distamycin conjugate covalently cross-linked to $[d(CGCGAATTCGCG)]_2$. We present a semi-quantitative model of the cross-linked DNA and discuss the ligand binding site. The model is compared with a related structure of a cross-linked mitomycin C:DNA complex.

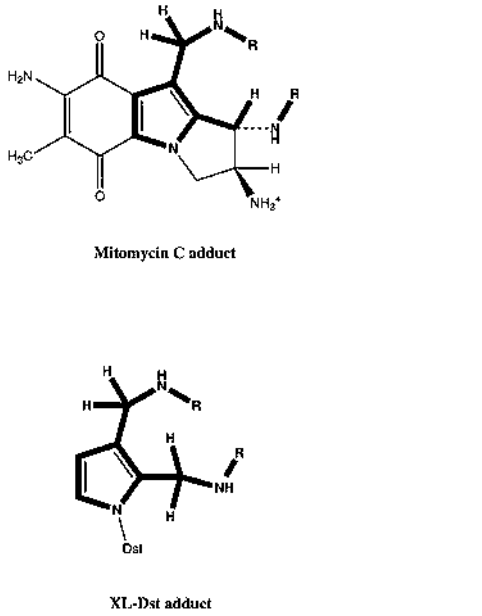

Figure 2. Comparison of the structures of the cross-linking agents mitomycin C and 2,3-bis(hydroxymethyl)pyrrole.

MATERIALS AND METHODS

Synthesis and purification of cross-linked oligonucleotides

XL-Dst was synthesized as described previously (36). The oligonucleotide d(CGCGAATTCGCG) was synthesized on an automated DNA synthesizer and purified by standard methods.

The following reaction was conducted to yield a sample of cross-linked DNA for NMR experiments. The purified oligomer $d(CGCGAATTCGCG)_2$ (2.3 µmol duplex) was dissolved in 35.2 ml of a solution of 50 mM HOAc/NaOAc buffer, pH 5.0, 100 mM NaCl and 5 mM $MgCl_2$ in a 50 ml plastic centrifuge tube. This solution was then treated with XL-Dst (7.5 µmol in a total volume of 0.5 ml CH_3OH, added in 50 µl aliquots, vigorously shaken after each addition) and the mixture was allowed to stand at ambient temperature for 4.5 days. The DNA-containing product was recovered by ethanol precipitation. The cross-linked product was separated from native DNA or alkylated single strands on a 20% denaturing polyacrylamide gel. The band with electrophoretic mobility roughly half that of native single strand was excised from the gel and the product recovered by a crush-and-soak procedure. The sample was desalted using a C18-RP Sep-Pak (Waters) to afford 137 OD_{260} of cross-linked duplex DNA.

The numbering scheme for the cross-linked oligomer is:

5′-C_1 G_2 C_3 **G_4** A_5 A_6 T_7 T_8 C_9 G_{10} C_{11} G_{12}-3′ (strand 1)
3′-G_{24} C_{23} **G_{22}** C_{21} T_{20} T_{19} A_{18} A_{17} G_{16} C_{15} G_{14} C_{13}-5′ (strand 2)

Bold lettering indicates the sites of cross-linking.

Sample preparation

NMR samples were prepared by dissolving the dried, cross-linked oligonucleotide duplex in 0.5 ml 10 mM potassium phosphate buffer, 50–500 mM NaCl, pH 7.0, and then lyophilizing to dryness. For experiments carried out in D_2O the solid was redissolved in 0.5 ml 99.96% D_2O (Cambridge Isotope Laboratories) and for experiments in H_2O a 90% H_2O/10% D_2O solution was used. The concentration of the NMR sample was 1.0 mM in duplex.

NMR experiments and resonance assignments

NMR experiments were performed on a 500 MHz GE Omega or a Bruker AMX-600 spectrometer. Spectra in H_2O were acquired using a 1-1 pulse sequence to suppress the water signal. 2D NOESY spectra in H_2O were acquired at 30°C with a spectral width of 13514 Hz, 64 scans/increment and a mixing time τ_M of 200 ms, and at 45°C with a spectral width of 13889 Hz, 64 scans/increment and τ_M 350 ms. 2D NOESY spectra in D_2O were acquired at 30 and 35°C, with a spectral width of 5000 Hz, 64 scans/increment and τ_M 200 ms. A 2D TOCSY in D_2O was acquired at 35°C, with a spectral width of 5000 Hz and τ_M 50 ms. All spectra were acquired in TPPI mode (37).

The data were processed with FELIX v. 2.30 (Hare Research) on a Silicon Graphics IRIS/4D workstation. Skewed sine bell functions were used for apodization of the free induction decays. A convolution method was applied to remove the intense H_2O resonance from the frequency spectrum. NOESY spectra of the ligand:DNA complex in D_2O and H_2O enabled assignment of the DNA and ligand resonances as described below (16,17,21,38,39).

Distance restraints

A model of XL-Dst cross-linked to $[d(CGCGAATTCGCG)]_2$ was built based on the NMR data. Watson–Crick hydrogen bond restraints were employed for all base pairs, based on the presence of slowly exchanging imino protons in the NMR spectrum. Ligand–DNA distance restraints were generated from the volume integrals of the cross-peaks in the D_2O NOESY spectrum acquired at a mixing time of 200 ms as described previously (20,21). The cross-peak volumes were classified semi-quantitatively into three categories: strong (1.8–2.8 Å), medium (1.8–3.5 Å) or weak (1.8–5.0 Å) relative to the volume integrals of the cytosine H5–H6 cross-peak volumes. Model refinement was performed with and without intramolecular DNA restraints in the region of the

cross-link. Some long distance restraints of 4.0–7.0 Å derived from several missing base aromatic–sugar H1′ NOESY cross-peaks were used.

Model refinement

The model of the [d(CGCGAATTCGCG)]$_2$ duplex was constructed using the Biopolymer module of Insight II (Biosym) from standard B-form DNA. A new potential was added to the AMBER force field for the adducted guanine N2 atoms. XL-Dst was assembled using the sketcher module of Insight II (Biosym) and assigned AMBER-type potentials. The partial charges on the atoms were obtained by a MOPAC calculation on the ligand. The ligand model was manually docked into the minor groove using Insight II.

Energy minimizations were performed using the Discover module of Insight II (employing the AMBER force field). Force constants of 200 (kcal/mol)/Å^2 and 50 (kcal/mol)/Å^2 were used for the experimentally derived Watson–Crick hydrogen bond restraints and ligand–DNA NOE restraints respectively. The cut-off distance for non-bonded interactions was set at 18 Å, with a switching distance of 2 Å. A distance-dependent dielectric of the form $\varepsilon = R$ was used to account for solvent effects. The energy of the complex was initially minimized using 100 steps of a steepest descents algorithm and 3000 steps of conjugate gradient minimization. The final model was created by subjecting the minimized starting model to 12 ps of restrained molecular dynamics (RMD) at 300 K, which was subsequently cooled to 100 K over 6 ps, followed by 5 ps of RMD at 100 K. The model was then energy minimized by conjugate gradient minimization to a final r.m.s. derivative of <0.05 (kcal/mol)/Å^2.

RESULTS

Sample preparation

The cross-linked dodecamer was prepared by direct admixture of two molar equivalents of XL-Dst with the synthetic dodecamer. Interstrand cross-linked DNA was separated from residual mono-adducts and unreacted DNA using preparative denaturing PAGE. The cross-linked material had an electrophoretic mobility roughly half that of the native single strands. The covalent connectivity of a closely related interstrand cross-link at the sequence CGAATT has previously been relatively well characterized (27,36) and that connectivity is assumed in the present case.

Titration of the AATT site with XL-Dst

NMR spectra were recorded at 25°C at a number of points in titration of the duplex [d(CGCGAATTCGCG)]$_2$ with XL-Dst (data not shown). The DNA resonances broaden significantly and shift upon addition of ligand, indicating that the complex is on the fast side of intermediate exchange on the NMR time scale. Upon addition of 1 M NaCl and increasing the temperature to 45°C the resonances sharpen slightly. However, due to the large linewidths we were unable to further characterize the non-covalent complex of XL-Dst with [d(CGCGAATTCGCG)]$_2$ in any detail.

In contrast, when the same DNA duplex is titrated with the parent compound distamycin distinct sets of resonances for the free and ligand-bound DNA exist (9), indicating a much slower dissociation rate for the unmodified ligand. This difference in rate probably reflects weaker binding of the modified ligand in the non-covalent DNA complex.

1D NMR of cross-linked DNA

We performed preliminary NMR studies on three versions of XL-Dst covalently cross-linked to d(CGCGAATTCGCG)$_2$. The compounds differed in the length of the $(CH_2)_n$ tether connecting the cross-linking pyrrole with the distamycin unit, where n = 2, 3 or 4. The 1D NMR spectra of the n = 2 and n = 4 cross-linked complexes show little dispersion and very broad imino proton resonances (data not shown). These data suggest weaker binding of the distamycin end of the molecule in these complexes. The NMR spectrum of the n = 3 cross-linked complex was chosen for further study because the resonances are relatively narrow with better chemical shift dispersion.

NMR resonance assignments of DNA

Upon formation of the covalent complex the symmetry of the DNA oligomer [d(CGCGAATTCGCG)]$_2$ is broken and the two DNA strands have non-degenerate resonances which can be assigned independently. Part of a NOESY spectrum of the XL-Dst:DNA complex in H_2O is presented in Figure 3, showing several ligand–DNA cross-peaks. Chemical shift assignments of the DNA base H6, H8, adenine H2 and thymine methyl proton resonances in the free and the cross-linked DNA were made from the D_2O NOESY spectra (see Table 1).

The adenosine H2 assignments were made based on cross-peaks from the H2 protons to intrastrand and interstrand deoxyribose H1′ protons in the D_2O NOESY. Cross-peaks from the H2 protons to thymidine imino protons in the H_2O NOESY confirm these assignments.

Several pieces of evidence define the location of the cross-link. The 2-amino protons on the two adducted guanines resonate at 8.89 p.p.m. and 8.08 p.p.m. (at 45°C) and give rise to a weak NOE cross-peak. NOE cross-peaks are also observed from the two adducted amino protons to their respective imino protons. The resonance at 8.08 p.p.m. also has a contact to the A_5 H2 and therefore is assigned to the G_4 2-amino proton, while the resonance at 8.89 p.p.m. must then be the G_{22} 2-amino proton.

NMR resonance assignments of the ligand

Intense NOE cross-peaks are observed from DNA adenine H2 protons to XL-Dst pyrrole H3 and amide NH protons. In identifying the location of the cross-link the orientation of the ligand with respect to the DNA is established, thereby facilitating the unambiguous assignment of ligand protons. Chemical shift assignments for the ligand are listed in Table 2. It was not possible to stereospecifically assign the methylene protons in the tether or in the $N(CH_3)_2$ tail.

The H4 and H5 protons on the cross-linking pyrrole are assigned to 6.37 and 6.70 p.p.m. (at 45°C) respectively. An NOE cross-peak was observed between the H4 resonance and the G_{22} C1′H resonance. The methylene protons on the cross-linking pyrrole (H2a, H2b, H3a and H3b) were assigned, although not stereospecifically, based on NOEs to the cross-linked guanine amino groups. NOEs are observed from the H3 protons to the ligand H4 and the DNA G_{22} N2H protons and from the ligand H2 protons to the DNA G_4 N2H proton.

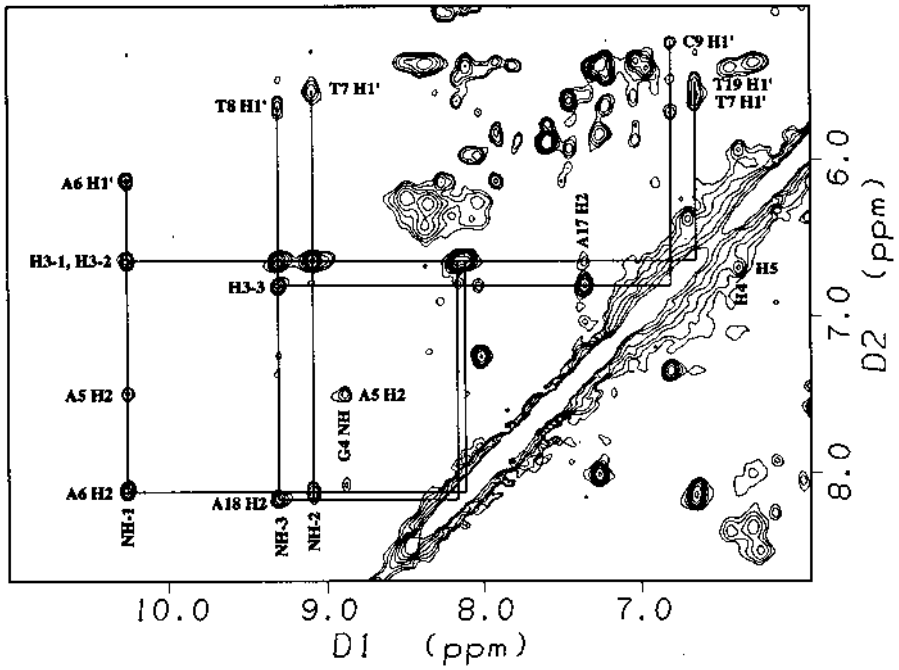

Figure 3. Expansion of the amide and aromatic to deoxyribose H1′ region of a NOESY spectrum of the XL-Dst:[d(CGCGAATTCGCG)]$_2$ complex (in H_2O, 45°C, τ_{mix} 350 ms). Various ligand–DNA and ligand–ligand NOE cross-peaks are labeled.

Table 1. Chemical shift assignments of the d(CGCGAATTCGCG)$_2$ duplex, free and in the XL-Dst:DNA complex[a]

	H6/H8			H1′			A H2/T CH_3		
	Free	Complex	Δδ	Free	Complex	Δδ	Free	Complex	Δδ
Strand 1									
C1	7.61	7.66	+0.05	5.79	5.77	–0.02			
G2	7.93	7.95	+0.02	5.88	5.95	+0.07			
C3	7.25	7.08	–0.17	5.63	5.84	+0.21			
G4	**7.82**	**7.87**	**+0.05**	**5.45**	**6.00**	**+0.55**			
A5	8.08	8.27	+0.19	5.98	6.22	+0.24	7.27	7.55	+0.28
A6	8.07	8.36	+0.29	6.14	6.20	+0.06	7.63	8.17	+0.54
T7	7.07	7.01	–0.06	5.88	5.63	–0.25	1.24	1.34	+0.10
T8	7.36	7.12	–0.24	6.08	5.70	–0.38	1.51	1.48	–0.03
C9	7.45	7.29	–0.16	5.66	5.30	–0.36			
G10	7.89	7.85	–0.04	5.85	5.94	+0.09			
C11	7.31	7.34	+0.03	5.78	5.79	+0.01			
G12	7.93	7.97	+0.04	6.15	6.19	+0.04			
Strand 2									
C13	7.61	7.64	+0.03	5.79	5.80	+0.01			
G14	7.93	7.98	+0.05	5.88	5.88	0.00			
C15	7.25	7.29	+0.04	5.63	5.74	+0.11			
G16	7.82	7.91	+0.09	5.45	5.43	–0.02			
A17	8.08	8.19	+0.11	5.98	6.05	+0.07	7.27	7.40	+0.13
A18	8.07	8.11	+0.04	6.14	5.77	–0.37	7.63	8.23	+0.60
T19	7.07	6.90	–0.17	5.88	5.54	–0.34	1.24	1.14	–0.10
T20	7.36	6.99	–0.37	6.08	5.35	–0.73	1.51	1.37	–0.14
C21	7.45	7.02	–0.43	5.66	5.41	–0.25			
G22	**7.89**	**7.87**	**–0.02**	**5.85**	**5.97**	**+0.12**			
C23	7.31	7.51	+0.20	5.78	6.10	+0.32			
G24	7.93	8.00	+0.07	6.15	6.17	+0.02			

[a]Chemical shifts are given in p.p.m. (± 0.01 p.p.m.). The residual HDO resonance is referenced to 4.65 p.p.m. (35°C) for both the free and the complexed DNA. Bold lettering indicates the sites of cross-linking.

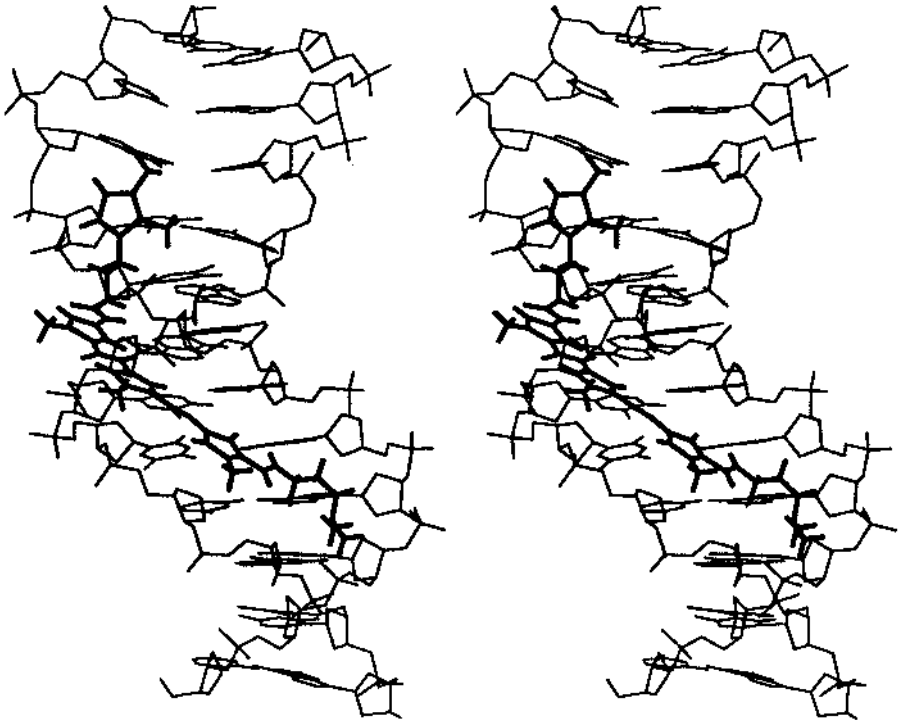

Figure 4. Stereo view of a molecular model of the XL-Dst:[d(CGCGAATTCGCG)]$_2$ complex obtained by energy minimization using semi-quantitative distance restraints derived from NOESY data. For clarity, hydrogen atoms are omitted for the DNA but not for the ligand molecule.

Table 2. Chemical shift assignments of XL-Dst in the interstrand cross-linked complex with [d(CGCGAATTCGCG)]$_2$[a,b]

Proton	Chemical shift
H2(a,b)[c]	3.75, 4.07
H3(a,b)[c]	3.80, 4.18
H4	6.37
H5	6.70
H6(a,b), H7(a,b), H8(a,b)[c,d]	2.26, 2.27, 2.45, 2.70, 3.73
NH-1	10.25
H3–1	6.65
NH-2	9.09
H3–2	6.66
NH-3	9.30
H3–3	6.80
NH-4	Not assigned
H9(a,b), H10(a,b), H11(a,b)[c]	Not assigned
N,N-$(CH_3)_2$	Not assigned

[a]Chemical shifts are given in p.p.m. (± 0.01 p.p.m.), with the residual HDO resonance referenced to 4.52 p.p.m. (45°C).
[b]Refer to Figure 1 for numbering system used.
[c]Resonances assigned as a group.
[d]One resonance in this group not assigned.

Several methylene proton resonances were assigned to the ligand protons in the tether, H6(a,b), H7(a,b) and H8(a,b), because they exhibit cross-peaks to NH-1 (the first amide of the distamycin fragment) and the cross-linking pyrrole H5. Some of the protons in this group have NOEs to the G_4 N2H and A_5 H2 protons, consistent with the location of the cross-link at G_4 and G_{22}.

Ligand–DNA contacts

NOE cross-peaks localize the ligand in the minor groove of the central AATT:AATT tract of the duplex (Table 3). Cross-peaks from DNA H2 and H1′ protons to ligand pyrrole H3 and amide NH protons indicate that the ligand pyrrole ring system spans the 5′-$A_5A_6T_7T_8$-3′ sequence (Fig. 3), in the minor groove. Several additional contacts are made between the XL-Dst protons and the DNA. These include NOEs from the cross-linking pyrrole H4 to G_{22} C1′H, from pyrrole H2 and H3 protons to G_4 N2H and G_{22} N2H respectively and from the ligand tether methylene protons to G_4 N2H.

Molecular modeling of the complex

A total of 16 ligand–DNA distance restraints derived from NOE data (see Table 3) and the two covalent ligand–DNA bonds were used to obtain the energy minimized model of the distamycin cross-linker complex with CGAATT:AATTCG (Fig. 4). There is some distortion of the DNA helix at the cross-linking site, which does not extend beyond the base pairs on either side of the cross-link. Although the dimethylammonium proton is likely to be involved in hydrogen bonding interactions with the DNA, specific acceptors cannot be resolved by the current data and model.

DISCUSSION

Site of cross-link

XL-Dst adducts guanine bases (36) and both N2 amino groups of deoxyguanosine residues at the sequence CG are necessary for efficient DNA–DNA interstrand cross-linking (27). XL-Dst has

previously been shown to form an interstranded cross-link with the sequence used in this study, d(CGCGAATTCGCG)$_2$, by denaturing gel electrophoresis and mass spectrometry (27).

Table 3. Ligand–DNA contacts for the XL-Dst:[d(CGCGAATTCGCG)]$_2$ interstrand cross-linked complex

XL-Dst	DNA	NOE classification[a]
NH-1	A5 H2	w
NH-1	A6 H1′	m
H3-1	A6 H2	s
H3-1	T20 H1′	m
H3-1	T7 H1′	m
NH-2	A6 H2	w
NH-2	T7 H1′	m
H3-2	A18 H2	s
H3-2	T19 H1′	w
H3-2	T8 H1′	w
NH-3	A18 H2	m
NH-3	T8 H1′	w
H3-3	A17 H2	s
H3-3	A18 H2	w
H3-3	C9 H1′	m
Cross-linking pyrrole H4	G22 H1′	m

[a]NOE classification: s, strong (1.8–2.8 Å); m, medium (1.8–3.5 Å); w, weak (1.8–5.0 Å).

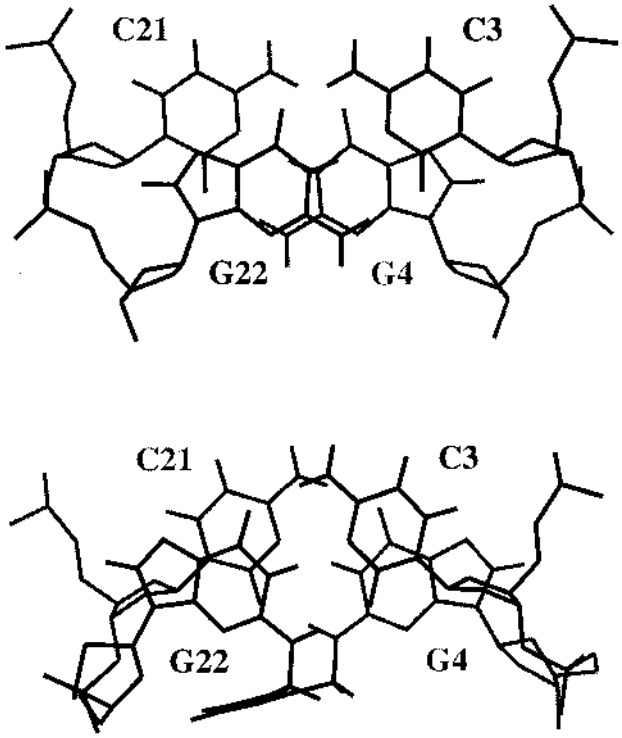

Figure 5. View down the helical axis of (**a**) a standard B-form DNA duplex of the sequence CG (built using Insight II software) and (**b**) the base pairs C_3·G_{22} and G_4·C_{21} in the XL-Dst:AATT cross-linked complex. Explicit hydrogen atoms are shown only for the aromatic bases and the cross-linking pyrrole unit.

The C_3·G_{22} and G_4·C_{21} base pairs are stabilized by the interstrand cross-link. The G_4 and G_{22} imino proton resonances exhibit slow exchange with solvent water, even at 45°C, where other imino proton resonances are broadened due to increased exchange with solvent. In the interstrand cross-linked complex of mitomycin C with [d(TACGTA)]$_2$ the imino proton resonances on the two cross-linked guanines are similarly unaffected by increasing temperature (6).

The NMR data confirm that the lesion occurs at the guanine N2 position. In unadducted DNA oligomers the guanine 2-amino proton resonances are broad due to the rotation of the amino group about the C2–N2 bond on a millisecond time scale. Covalent attachment of a ligand to N2 prevents this rotation, resulting in one narrow NMR resonance for the remaining amino proton. The two substituted amino proton resonances on G_4 and G_{22} have contacts to nearby ligand and DNA protons and intense NOE cross-peaks to the imino proton on the same guanine base. Similar patterns are observed in the cross-linked mitomycin C complex, in which mitomycin C forms adducts in the minor groove at N2 of guanines in the sequence CG (6). The chemical shifts of the substituted G_4 and G_{22} amino (8.89 and 8.08 p.p.m.) and imino (13.17 and 12.45 p.p.m.) protons in the XL-Dst:DNA cross-linked complex are similar to those observed in the mitomycin C complex (aminos at 9.36 and 8.87 p.p.m. and iminos at 13.12 and 12.60 p.p.m.).

Additional evidence that the cross-link is in the minor groove arises from NOEs from hydroxymethyl pyrrole protons and ligand tether protons to DNA nucleotides G_4, A_5 and G_{22}. We were unable to confirm scalar coupling between the ligand H2 protons and the G_4 amino proton and between the ligand H3 protons and the G_{22} amino proton.

Distortion of the DNA due to cross-link

The model of the XL-Dst:[d(CGCGAATTCGCG)]$_2$ complex is shown in Figure 4. Weak intraresidue NOE cross-peaks from base protons to H1′ protons indicate that the bases are all in *anti* geometry. Several extremely weak or missing cross-peaks are indicative of distortion in the two cross-linked base pairs, which extends to the base pairs on either side of the cross-link. In the refined model we observe a large negative buckle for the C_3·G_{22} base pair and a large positive buckle for the G_4·G_{21} base pair. Additionally, the model shows a decrease in helical twist between these two base pairs, indicating unwinding at the cross-link site.

The slide between the two base pairs at the cross-link site is unusual compared with standard B-DNA, so that C_3 and C_{21} are moved from the edges of the groove toward the center of the helix and G_{22} and G_4 are pushed away from the center of the helix (Fig. 5). As a result, the cross-linked guanine bases are positioned over the deoxyribose H2′ and H2″ of C_3 and C_{21}. The NMR data of the cross-linked complex agree with this model. The H2′ resonances of C_3 and C_{21} are dramatically upfield shifted, by 0.80 and 0.99 p.p.m. respectively, relative to the uncomplexed DNA. An upfield shift of this magnitude has not been observed in any complexes of distamycin and related analogs with DNA in our laboratory. We conclude that these unusual shifts are due to distortion of the DNA structure by the cross-link. The cytosine H2′ and H2″ resonances undergo an upfield ring current shift due to stacking over the aromatic guanine base, as seen in the model, similar to ring current shifts observed for protons which are stacked over aromatic amino acid side chains in protein NMR studies.

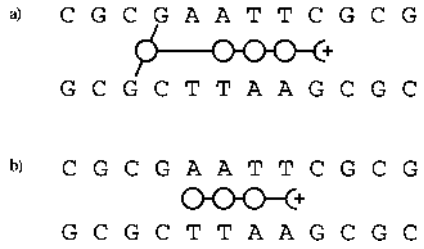

Figure 6. Schematic of (**a**) the XL-Dst:[d(CGCGAATTCGCG)]$_2$ cross-linked complex and (**b**) the non-covalent distamycin:[d(CGCGAATTCGCG)]$_2$ complex.

Overall binding site

An important feature of XL-Dst is that it requires both a distamycin binding DNA sequence and an adjacent pyrrole cross-linking sequence to cross-link DNA efficiently (27), resulting in a 6 bp binding site, CGAATT. Ligand–DNA NOEs unambiguously confirm formation of the XL-Dst:DNA complex at this site in the minor groove. The tether is fully extended, bridging the A_{17}·T_8 base pair. It is not clear whether the tether can tolerate a G·C base pair at this site. The amino group at the 2 position of guanine may sterically interfere with the ligand binding and/or reactivity at the cross-linking site. The $N(CH_3)_2$ tail extends across the G_9·C_{16} base pair, but no specific contacts are observed.

From the model, hydrogen bonds can be inferred between ligand amide protons and adenine N3 and thymine O2, as seen in other models of distamycin complexed to DNA (9,16,17). The tripyrrole unit contacts three adenine H2 protons in the sequence ATT, with the 'head' of the ligand (pyrrole ring number 1 in Fig. 1) nearest the adenine, pointing toward the 5′ direction of the strand. The non-covalent complex of distamycin with [d(CGCGAATTCG-CG)]$_2$ is depicted in Figure 6 (9). It appears that the covalent linkage in the cross-linked complex forces the tripyrrole unit to shift by 1 bp toward the 3′-end of the first strand, relative to the distamycin binding site. However, it is difficult to predict the preferred subsite, AAT or ATT, for non-covalent binding of XL-Dst to d(CGCGAAT-TCGCG)$_2$ by analogy with these complexes.

Comparison with mitomycin C

The design of the bis(hydroxymethyl)pyrrole unit was derived from the pharmacophore in mitomycin C (Fig. 2). There are many similarities between the structure discussed here and the structure of the cross-linked mitomycin C:DNA complex (6). In the XL-Dst cross-link we find base stacking throughout the helix, but observe several weak NOE cross-peaks between DNA base and sugar H1′ protons near the C_3·G_{22} and G_4·C_{21} base pairs. Patel and co-workers observed similar indications of base stacking and unusually weak NOE cross-peaks at the mitomycin C cross-link site (6).

The overall geometry of the modeled XL-Dst:DNA complex in this study is similar to the calculated structure of the cross-linked mitomycin C:DNA complex. In the XL-Dst model the cross-linking pyrrole ring interacts with bases C_{21} and G_{22} on one strand of DNA, rather than being positioned in the center of the minor groove. This geometry was proposed by Weidner *et al.*, based on molecular modeling (40), and appears to be a consequence of the covalent geometry of the cross-link, since energy minimizations arrive at this structure from different starting structures and without any ligand–DNA restraints. This structure is consistent with an observed NOESY cross-peak between the H4 proton on the pyrrole ring and the H1′ proton of G_{22}. Patel and co-workers came to the same conclusion, based on their NMR-restrained structure calculations and on unusual chemical shifts of the sugar H1′ and H2″ protons on the analogous strand in the region of the mitomycin C:DNA cross-link (6).

The minor groove width of GC regions has typically been measured to be ~5–7 Å (41,42). The non-planar five membered ring in mitomycin C may widen the minor groove to 9.2 or 10.8 Å, depending on its pucker (6). The bis(methoxy)pyrrole unit is planar. The minor groove width in the region of the XL-Dst cross-link (10–11 Å) is similar to that in the mitomycin C adduct.

Molecular design

The properties and dimensions of tethers are becoming increasingly important in ligand design. Tethers have been used to link a variety of reactive groups to lexitropsins (26,29–35). Two or more lexitropsins have been tethered together end-to-end (8,43), substantially increasing the length of the targeted site to 10 or more base pairs. The finding in this study that the XL-Dst tether bridges 1 bp, with the binding site for the tripyrrole unit shifted by 1 bp relative to the distamycin binding site, may be due in part to the properties of the tether. This emphasizes the need for further investigation and systematic characterization of different linkers and their minor groove binding properties in 1:1 and 2:1 ligand–DNA complexes.

A critical issue in the design of sequence-specific DNA ligands which contain a reactive group is the effect of the reactive group on the affinity and sequence specificity of the ligand. In the cases of XL-Dst and netropsin–diazene (44) each molecule contains a bulky hydrophobic group appended to one end of the molecule and the NMR exchange behavior indicates lowered non-covalent binding affinity relative to the parent molecules distamycin or netropsin. Since the non-covalent binding event is critical to the delivery of the reactive group to the proper target site, this highlights the need for studies examining how the presence of a reactive group changes the binding selectivity of such conjugated ligands.

CONCLUSION

We have characterized the DNA binding mode of a new rationally designed DNA cross-linking agent based on distamycin using 2D NMR spectroscopy. In this study a distamycin derivative has been chosen as a starting point for the rational design of a sequence-selective cross-linker. Distamycin and its analogs have shown flexible and predictable DNA binding selectivity when bound to the minor groove of DNA either singly or in an anti-parallel, side-by-side two drug complex. This system holds excellent promise for the eventual design of a compound which forms interstrand DNA cross-links with high sequence selectivity for many DNA sequences. Further studies on the binding properties and cross-linking mechanism of this and related molecules will prove useful in the development of effective anti-tumor and anti-viral agents.

ACKNOWLEDGEMENTS

This work was supported in part by National Institutes of Health grants GM-43219 (DEW) and GM-45804 (PBH), post-doctoral fellowship GM-14966 (HPS), through instrumentation grants from the US Department of Energy, DE FG05-86ER75281, and the National Science Foundation, DMB 86-09305 and BBS 87-20134,

and by the Director, Office of Biological and Environmental Research, General Sciences Division of the US Department of Energy under contract no. DE-AC03-76SF00098 (DEW).

REFERENCES

1 Iyer,V.N. and Szybalski,W. (1963) *Proc. Natl. Acad. Sci. USA*, **50**, 355–362.
2 Pinto,A.L. and Lippard,S.J. (1985) *Biochim. Biophys. Acta*, **780**, 167–180.
3 Teng,S.P., Woodson,S.A. and Crothers,D.M. (1989) *Biochemistry*, **28**, 3901–3907.
4 Millard,J.T., Weidner,M.F., Raucher,S. and Hopkins,P.B. (1990) *J. Am. Chem. Soc.*, **112**, 3637–3641.
5 Borowy-Borowski,H., Lipman,R. and Tomasz,M. (1990) *Biochemistry*, **29**, 2999–3006.
6 Norman,D., Live,D., Sastry,M., Lipman,R., Hingerty,B.E., Tomasz,M., Broyde,S. and Patel,D.J. (1990) *Biochemistry*, **29**, 2861–2875.
7 Zimmer,C. and Wähnert,U. (1986) *Prog. Biophys. Mol. Biol.*, **47**, 31–112.
8 Dervan,P.B. (1986) *Science*, **232**, 464–471.
9 Klevit,R.E., Wemmer,D.E. and Reid,B.R. (1986) *Biochemistry*, **25**, 3296–3303.
10 Wemmer,D.E., Fagan,P. and Pelton,J.G. (1990) In Pullman,B. and Jortner,J. (eds), *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions*. Kluwer Academic, Dordrecht, The Netherlands, pp. 95–102.
11 Coll,M., Aymami,J., van der Marel,G.A., van Boom,J.H., Rich,A. and Wang,A.H.-J. (1989) *Biochemistry*, **28**, 310–320.
12 Breslauer,K.J., Remeta,D.P., Chou,W.-Y., Ferrante,R., Curry,J., Zaunczkowski,D., Snyder,J.G. and Marky,L.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8922–8926.
13 Kopka,M.L., Yoon,C., Goodsell,D., Pjura,P. and Dickerson,R.E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1376–1380.
14 Patel,D.J. (1979) *Eur. J. Biochem.*, **99**, 369–378.
15 Patel,D.J. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6424–6428.
16 Pelton,J.G. and Wemmer,D.E. (1988) *Biochemistry*, **27**, 8088–8096.
17 Pelton,J.G. and Wemmer,D.E. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5723–5727.
18 Marky,L.A. and Breslauer,K.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4359–4363.
19 Dabrowiak,J.C., Goodisman,J. and Kissinger,K. (1990) *Biochemistry*, **29**, 6139–6145.
20 Mrksich,M., Wade,W.S., Dwyer,T.J., Geierstanger,B.H., Wemmer,D.E. and Dervan,P.B. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 7586–7590.
21 Dwyer,T.J., Geierstanger,B.H., Bathini,Y., Lown,J.W. and Wemmer,D.E. (1992) *J. Am. Chem. Soc.*, **114**, 5911–5919.
22 Geierstanger,B.H., Parks,M., Dervan,P. and Wemmer,D.E. (1994) *Science*, **266**, 646–650.
23 Lown,J., Krowicki,K., Bhat,U., Skorobogaty,A., Ward,B. and Dabrowiak,J. (1986) *Biochemistry*, **25**, 7408–7416.
24 Wade,W.S. and Dervan,P.B. (1987) *J. Am. Chem. Soc.*, **109**, 1574.
25 Geierstanger,B.H., Dwyer,T.J., Bathini,Y., Lown,J.W. and Wemmer,D.E. (1993) *J. Am. Chem. Soc.*, **115**, 4474–4482.
26 Lee,M., Rhodes,A.L., Wyatt,M.D., D'Incalci,M., Forrow,S. and Hartley,J.A. (1993) *J. Med. Chem.*, **36**, 863–870.
27 Sigurdsson,S.Th., Rink,S.M. and Hopkins,P.B. (1993) *J. Am. Chem. Soc.*, **115**, 12633–12634.
28 Woo,J., Sigurdsson,S.Th. and Hopkins,P.B. (1993) *J. Am. Chem. Soc.*, **115**, 3407–3415.
29 Broggini,M., Erba,E., Ponti,M., Ballinari,D., Geroni,C., Spreafico,F. and D'Incalci,M. (1991) *Cancer Res.*, **51**, 199–204.
30 Baker,B.F. and Dervan,P.B. (1985) *J. Am. Chem. Soc.*, **107**, 8266–8268.
31 Baker,B.F. and Dervan,P.B. (1989) *J. Am. Chem. Soc.*, **111**, 2700–2712.
32 Chen,F.-X., Zhang,Y., Church,K., Bodell,W.J. and Gold,B. (1993) *Carcinogenesis*, **14**, 935–940.
33 Church,K.M., Wurdeman,R.L., Zhang,Y., Chen,F.-X. and Gold,B. (1990) *Biochemistry*, **29**, 6827–6838.
34 Zhang,Y., Chen,F.-X., Mehta,P. and Gold,B. (1993) *Biochemistry*, **32**, 7954–7965.
35 Lee,M., Roldan,M.C., Haskell,M.K., McAdam,S.R. and Hartley,J.A. (1994) *J. Med. Chem.*, **37**, 1208–1213.
36 Sigurdsson,S.Th. and Hopkins,P.B. (1994) *Tetrahedron*, **50**, 12065–12084.
37 Marion,D. and WYthrich,K. (1983) *Biochem. Biophys. Res. Comm.*, **113**, 967–974.
38 Hare,D.R., Wemmer,D.E., Chou,S.-H. and Drobny,G. (1983) *J. Mol. Biol.*, **171**, 319–336.
39 Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York, NY.
40 Weidner,M.G., Sigurdsson,S.Th. and Hopkins,P.B. (1990) *Biochemistry*, **29**, 9225–9233.
41 Saenger,W. (1983) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York, NY.
42 Heinemann,U. and Alings,C. (1989) *J. Mol. Biol.*, **210**, 369–381.
43 Mrksich,M., Parks,M. and Dervan,P. (1994) *J. Am. Chem. Soc.*, **116**, 7983–7988.
44 Spielmann,H.P., Fagan,P.A., Bregant,T.M., Little,R.D. and Wemmer,D.E. (1995) *Nucleic Acids Res.*, **23**, 1576–1583.