Shelley B. Howerton
Akankasha Nagpal
Loren Dean Williams
School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400

Received 9 October 2002; accepted 9 October 2002

Surprising Roles of Electrostatic Interactions in DNA–Ligand Complexes*

Abstract: *The positions of cations in x-ray structures are modulated by sequence, conformation, and ligand interactions. The goal here is to use x-ray diffraction to help resolve structural and thermodynamic roles of specifically localized cations in DNA–anthracycline complexes. We describe a 1.34 Å resolution structure of a $CGATCG_2$–$adriamycin_2$ complex obtained from crystals grown in the presence of thallium (I) ions. Tl^+ can substitute for biological monovalent cations, but is readily detected by distinctive x-ray scattering, obviating analysis of subtle differences in coordination geometry and x-ray scattering of water, sodium, potassium, and ammonium. Six localized Tl^+ sites are observable adjacent to each $CGATCG_2$–$adriamycin_2$ complex. Each of these localized monovalent cations are found within the G-tract major groove of the intercalated DNA–drug complex. Adriamycin appears to be designed by nature to interact favorably with the electrostatic landscape of DNA, and to conserve the distribution of localized cationic charge. Localized inorganic cations in the major groove are conserved upon binding of adriamycin. In the minor groove, inorganic cations are substituted by a cationic functional group of adriamycin. This partitioning of cationic charge by adriamycin into the major groove of CG base pairs and the minor groove of AT base pairs may be a general feature of sequence-specific DNA–small molecule interactions and a potentially useful important factor in ligand design.* © 2003 Wiley Periodicals, Inc. Biopolymers 69: 87–99, 2003

Keywords: *major groove; minor groove; intercalation; daunomycin; adriamycin; alkaline metal; thallium; x-ray diffraction; anomalous scattering*

INTRODUCTION

In x-ray structures of several DNA–anthracycline complexes, sodium ions mediate interactions between the oxygen atoms of the intercalated chromophore and the N7 atom of an adjacent guanosine[1,2] (Figure 1). These monovalent cations appear to influence stability, conformation, and sequence specificity. However, x-ray structures of other DNA–anthracycline complexes lack localized cations.[3,4] The goals of the work

Correspondence to: Loren Dean Williams; email: loren.williams@chemistry.gatech.edu
Contract grant sponsor: National Science Foundation
Contract grant number: MCB-9056300
*Atomic coordinates and structure factors have been deposited in the Nucleic Acid Database (NDB) (in progress)
Biopolymers, Vol. 69, 87–99 (2003)

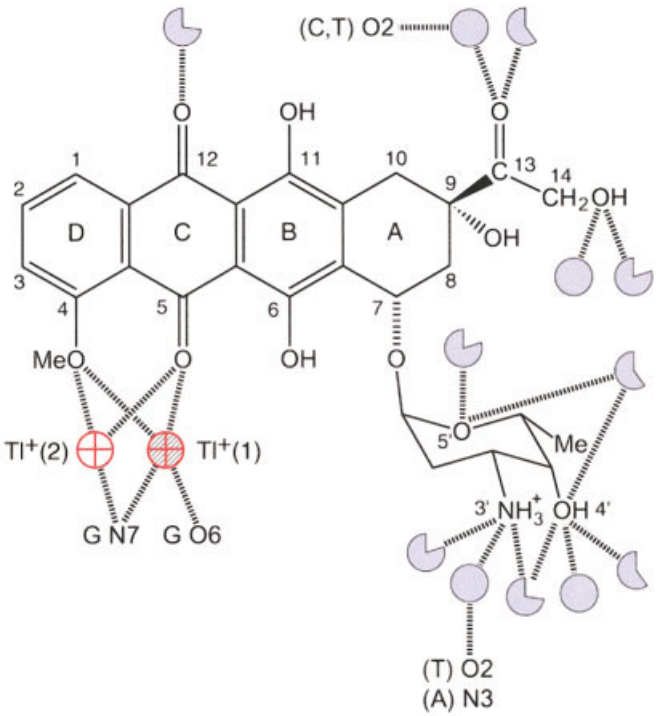

FIGURE 1 Chemical structure of adriamycin. Daunomycin is similar to adriamycin but lacks the 14-hydroxyl group. The circles and pie-shaped elements represent water molecules. A shaded full circle indicates that a water molecule interacts with the associated oxygen or nitrogen in 100% of the surveyed structures. The pie-shaped elements indicate quantitatively the fraction of structures in which a water molecule interacts with the oxygen or nitrogen atom. The circles that contain plus signs indicate sites monovalent cations determined here. The nonshaded cation is nearly identical to the location of sodium ions in previous structures.

described here are to use x-ray diffraction to resolve these discrepancies, and more generally, to help determine roles of localized cations and electrostatic forces in structure, thermodynamics, and sequence specificity of DNA–ligand complexes. Our results suggest that favorable interactions of adriamycin with the sequence-specific electrostatic landscape of DNA are universal characteristics of DNA–small molecule interactions and may be useful in sequence-specific ligand design.

Anthracyclines bind preferentially at pyrimidine–purine steps[5] (an example is shown in Figure 2) and cause DNA damage in vivo by interfering with the action of topoisomerase II.[6] The DNA binding energetics of daunomycin were experimentally decomposed by Chaires and co-workers into functional-group-specific contributions[7]. The daunosamine contributes a large and favorable binding free energy (2 kcal mol^{-1}). The 3′-amino group contributes 0.7 kcal mol^{-1} above the polyelectrolyte contribution of the positive charge. The hydroxyl groups at the 9 and 14 positions contribute approximately 1 kcal mol^{-1}. Chaires also identified water as an important thermodynamic participant, observing an uptake of water accompanying formation of DNA–anthracycline complexes.[8] The thermodynamic results are consistent with three-dimensional structure determinations of a series of DNA–anthracycline complexes.[1,2,9,10] The three-dimensional structures demonstrate that the daunosamine is in intimate contact with the DNA; hydrogen bonds and van der Waals contacts secure the daunosamine to the floor of the minor groove. The 9-hydroxy group forms multiple hydrogen bonds with a flanking guanine. A large number of solvent molecules are conserved in crystal structures of DNA–anthracycline complexes.[2] The role of cations and electrostatics is less well characterized.

Identifying localized cations by x-ray diffraction presents nontrivial analytical challenges. Sodium ions (Na^+), potassium ions (K^+), rubidium ions (Rb^+), cesium ions (Cs^+), ammonium ions, and water molecules, and even polyamines and divalent cations, compete for similar sites adjacent DNA or RNA.[11–13] Various species bind with partial and mixed occupancies in x-ray structures. Na^+ and ammonium scatter x-rays with nearly the same power as water or par-

tially occupied K^+. Each of these species has irregular and variable coordination. Therefore the solvent–ion environment can be difficult to fit unambiguously during structure refinement. For these reasons, the inferred variability in cation interactions in previous DNA–anthracycline complexes may be real, or may arise from difficulties in characterization of cation positions by x-ray diffraction. A well-developed K^+ substitute with a distinctive x-ray scattering fingerprint that obviates interpretation of subtle differences in coordination geometry and scattering power is provided by thallium (Tl^+). Tl^+ and K^+ (a) have similar ionic radii [K^+ = 1.33 Å, Tl^+ = 1.49 Å][14] and enthalpies of hydration [K^+ = −77 kcal mol^{-1}, Tl^+ = −78 kcal mol^{-1}].[15] Tl^+ can substitute for K^+ in the catalytic mechanisms of sodium–potassium pumps,[16] fructose-1-6-bisphosphatase,[17] and pyruvate kinase.[18] Tl^+ has been shown by NMR to stabilize guanine tetraplexes in a manner analogous to K^+ and ammonium.[19] We have used Tl^+ as a marker for sites of K^+ localization adjacent to B-DNA.[13] Doudna and co-workers used Tl^+ as a marker for sites of K^+ localization tetrahymena ribozyme P4–P6 domain.[20] Tl^+ was used by Caspar and co-workers to determine counterion positions adjacent to insulin,[21,22] and by Gill and Eisenberg to determine the location of ammonium ions in the binding pocket of glutamine synthetase.[23] Here we describe a 1.34 Å resolution structure of a $CGATCG_2$–$adriamycin_2$ complex obtained from crystals grown in the presence of Tl^+. Our goal is to use Tl^+ to determine locations of monovalent cations in DNA–anthracycline complexes. The RESULTs indicate that adriamycin interacts favorably with the asymmetric electrostatic landscape of DNA. Asymmetric cation localization is conserved and even enhanced upon binding of adriamycin.

MATERIALS AND METHODS

Crystallization and Data Collection

Crystals were grown by vapor diffusion from sitting drops of a solution initially containing 2.3 m*M* of the ammonium salt of reverse-phase high performance liquid chromatography (HPLC) purified d(CGATCG) (Midland Certified Reagent Company in Midland, TX), 27.5 m*M* thallium acetate (pH 6.4), 2.2 m*M* Mg^{2+} acetate, 4.5% 2-methyl-2,4-pentanediol (MPD), 1.3 m*M* spermine acetate, and 2.5 m*M* of the chloride salt of adriamycin (Sigma), equilibrated against a solution of 35% MPD. A tetragonal crystal grew to 0.4 × 0.2 × 0.2 mm^3 within a few weeks. Data were collected with 1.54 Å Cu K_α radiation on an in-house Rigaku/MSC RU-H3R rotating anode generator equipped with Osmic blue multilayer confocal optics and an RAXIS-IV^{++} image plate detector. The amount of 360° of reciprocal space data were collected at −180°C; 31,250 reflections to 1.26 Å resolution were reduced to a unique dataset containing 5559 reflections, preserving Bijvoet pairs using the dtprocess software in the program CrystalClear 1.3.0 (R^{merge} = 4.2%, 89.7% complete, $\Delta F/F = 0.083$). Unit cell dimensions are $a = b = 27.89$ Å, $c = 52.27$, $\alpha = \beta = \gamma = 90°$ in space group $P4_12_12$. Data used for refinement included 5034 reflections from 35 to 1.34 Å. (See Table I.)

Refinement

A starting model consisting of coordinates of a published $CGATCG_2$–$adriamycin_2$ complex (NDB entry DDF044)[4] was annealed and refined against the 1.34 Å $CGATCG_2$–$adriamycin_2$–Tl^+ data with the program CNS, using the parameters of Berman and co-workers.[24–26] The restraints for adriamycin were adapted from the small-molecule crystal structure of daunomycin.[27] The program O version 6.4B,[28] was used for viewing and manipulating molecular models and viewing electron density maps, which were clear, detailed and continuous (Figure 3). Water molecules were added iteratively to peaks of corresponding sum and difference density followed by refinement and phase calculation. A subset of water molecules were converted to Tl^+ ions during the refinement. The conversions were based on the presence of persistent $F_0 - F_c$ difference density superimposed on water molecules (Figure 3B). No attempt was made to bias the assignment of Tl^+ ions in the refinement toward previously published sodium ion assignments. Neither Mg^{2+} nor spermine was apparent in electron density maps.

Anomalous Difference Fouriers

$2F_0 - F_c$ peaks were assigned as Tl^+ ions only if Tl^+ occupancy is indicated in both $F_0 - F_c$ and Bijvoet difference ($F_+ - F_-$) Fourier maps. $F_+ - F_-$ maps were calculated using data to 1.4 Å and phases from the final refined $CGATCG_2$–$adriamycin_2$–Tl^+ model, minus water and ions. Pairs of reflections with $|F|/\sigma(F) < 1.0$ for either Bijvoet pair were excluded. The most intense peak (30σ) in the $F_+ - F_-$ map (Figure 3C) is 0.05 Å from the most highly occupied Tl^+ site determined from the refinement (Tl^+ 1). A shoulder (6σ) on the 30 σ peak is nearly superimposed on another Tl^+ site determined during the refinement (Tl^+ 2). The $F_+ - F_-$ electron density surrounding Tl^+ 1 and Tl^+ 2 is similar in size and shape to the $F_0 - F_c$ Fourier electron density. An additional $F_+ - F_-$ peak (4.3σ) is 0.18 Å from the third Tl^+ site determined in the refinement (Tl^+ 3). Additional peaks observed in the vicinity of Tl^+ 3 in the $F_+ - F_-$ map (Figure 3C) were not assigned as Tl^+ ions because of a lack of evidence for Tl^+ at these sites in $F_0 - F_c$ maps. The most intense, a peak of 4.5σ, is 0.43 Å from a water molecule (W81). Some remaining peaks in the $F_+ - F_-$ map correspond to anomalous signals from phosphorus atoms of the DNA, as observed previously.[29] The $F_+ - F_-$ Fourier map corre-

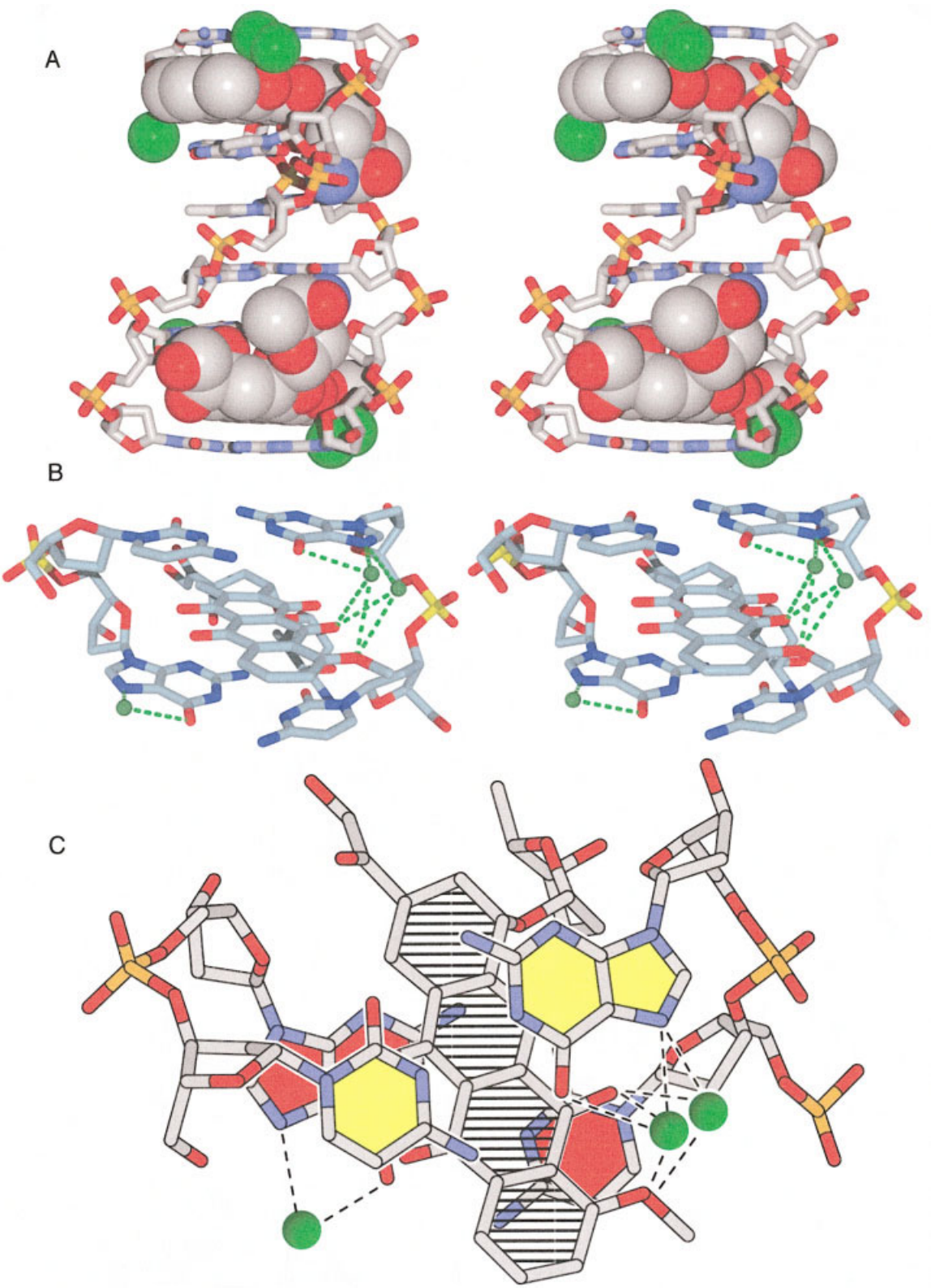

FIGURE 2

sponds well with the locations of anomalous scatterers in the final refined model. Each of the three Tl^+ sites are partially occupied in the final model. The occupancies were estimated during the refinement by manually minimizing the magnitudes of the $F_0 - F_c$ and $F_c - F_0$ peaks. Tl^+ occupancies were not automatically refined.

Superimpositions

Superimpositions were performed using the program ProFit.[30] Guanine residues adjacent to monovalent cations in NDB entries listed in Table II were superimposed on G(2) of $CGATCG_2$–$adriamycin_2$–Tl^+. All atoms of the purine ring system were used in the superimpositions. The average root mean square deviation (RMSD) was 0.04Å, and all were less than 0.07Å. The ions in the superimposition set were grouped according to their distances from O6 and N7 and the plane of the guanine base. To determine if the binding modes were statistically distinct, the mean position and 95% confidence intervals were calculated for each group. In pairwise comparisons of the four groups, the distance between the mean positions were greater than sum of the confidence intervals, indicating differences in the four groups were statistically significant.

Determination of DNA–anthracycline hydration patterns utilized sixteen complexes [Nucleic Acid Database (NDB) IDs: DD000,[31] DDF001,[1] DDF020,[2] DDF022,[32] DDF023,[33] DDF026,[34] DDF035,[9] DDF041,[35] DDF044,[4] DDF045,[4] DDF062,[36] DDFB24,[33] DDFB25,[34] DDFB70,[37] DDFP21,[3] and the structure described here]. For evaluation of major groove ligand interactions, the purine ring atoms of G(2) of three bis-intercalative complexes (NDB IDs: DD0018, DDD030, and DDDB46) was superimposed on G(2) of the current model.

Table I Crystallographic and Refinement Statistics

Unit cell	
α, β, γ	90°
a (Å)	27.89
b (Å)	27.89
c (Å)	52.27
Space group	$P4_12_12$
Temperature of data collection (°C)	−180
No. of reflections	31,250
No. of unique reflections	5559
No. of reflections used in refinement $[F/\sigma(F) > 2.0]$	4302
Max resolution of observed reflections (Å)	1.26
Average $I/\sigma(I)$	10.9
Completeness (%)/highest shell (%)	99.9/99.9
Max resolution of highest shell used in refinement (Å)	1.34
Resolution range (Å)	35.0–1.34
Completeness of data used in refinement (%)/highest shell (%)	85.5/60.1
No. of reflections used in test set	457
RMSD of bonds from ideality (Å)	0.02
RMSD of angles from ideality (°)	1.67
DNA (asymmetric unit)	d(CGATCG)
No. of DNA atoms	120
No. of adriamycin atoms	39
No. of water molecules	58
No. $T1^+$ ions/summed occupancy	3/0.9
No. Mg^{2+} ions	0
No. of spermine atoms	0
R free (%)	21.5
R factor (%) excluding test set data	19.3
Final R factor (using all data)	18.9

RESULTS

Tl^+ does not perturb the conformation of the DNA or the DNA–drug interactions in the $CGATCG_2$–$adriamycin_2$–Tl^+ complex. The conformation and DNA–drug interactions of the low temperature (LT) $CGATCG_2$–$adriamycin_2$–Tl^+ complex are similar to those in previous room temperature[2] (RT) and LT[4] structures of $CGATCG_2$–$adriamycin_2$ obtained from crystals grown with Na^+ as the primary monovalent cation. Each complex (biological unit) consists of $CGATCG_2$–$adriamycin_2$, a DNA hexamer duplex plus two intercalated adriamycin molecules (Figure 2), and contains two crystallographic asymmetric

FIGURE 2 (A) Stereoview of the $CGATCG_2$–$adriamycin_2$–Tl^+ complex. DNA atoms are shown in stick representation. Adriamycin atoms are displayed as space-filling representation. Tl^+ ions are green spheres. (B) Stereoview of the CpG intercalation site of the $CGATCG_2$–$adriamycin_2$–Tl^+ complex. DNA and adriamycin atoms are shown in stick representation and colored by CPK. Tl^+ ions are shown as green spheres. Interactions between Tl^+ ions and DNA/adriamycin functional groups are depicted as dashed lines. (C) View down the DNA helical axis showing the CpG intercalation step and the adriamycin molecule. The rings of the chromophore are shaded yellow. The rings of the terminal base pair are hatched with vertical lines. The planes of the bases in the internal CG base pair are shaded with solid red. The three Tl^+ ions are shown as green spheres. A and B were rendered with Povray 3.1. DNA and adriamycin atoms are colored by CPK (carbon, gray; oxygen, red; nitrogen, blue; phosphorous, yellow).

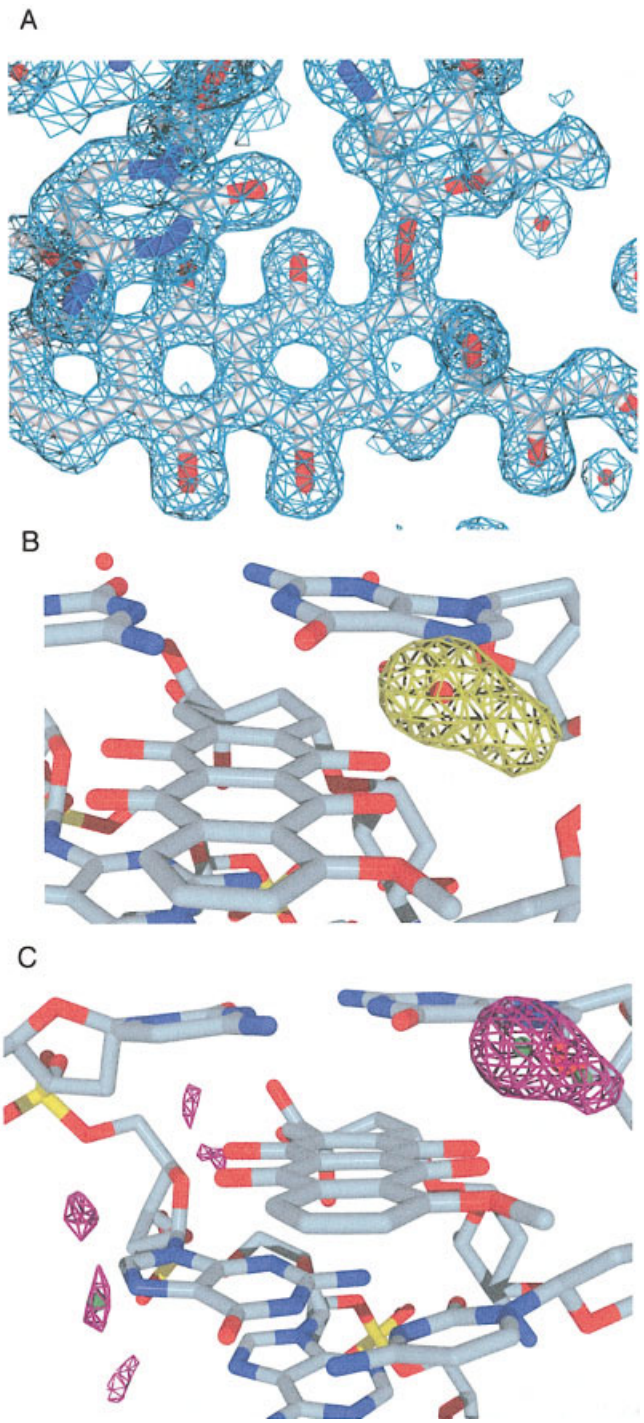

FIGURE 3 (A) $2F_0 - F_c$ Fourier electron density map contoured at 1σ. This view of adriamycin also shows an adjacent cytosine residing above the plane of the chromophore, and several water molecules. (B) Residual $F_0 - F_c$ Fourier electron density superimposed. This 5σ map was calculated with phases that included contributions from the water molecule within the difference density. That water molecule was subsequently converted to a Tl^+. (C) Anomalous $F_0 - F_-$ Fourier electron density Fourier map. The map is contoured at 3.5σ. Three Tl^+ ions are depicted as small green spheres. These figures were generated with SwissPDBViewer and rendered with Povray 3.1. DNA and adriamycin atoms are shown in stick representation. Water molecules are shown as red spheres.

Table II High Resolution Structures Containing Monovalent Cations Adjacent to Guanine Residue

NDB ID	Nucleic Acid Type	Resolution (Å)	Ion	Distance to N7	Distance to O6	Binding Mode
PR0052[49]	RNA: sarcin/ricin loop	1.97	K^+ 406	4.64	2.73	G O6
			K^+ 403	4.43	2.66	G O6
			K^+ 401	2.81	2.75	G N7/O6-bridge
			K^+ 404	2.65	2.85	G N7/O6-bridge
PR0021[50]	RNA: signal recognition particle	1.8	K^+ 703	5.82	2.77	G O6
			K^+ 701	3.93	2.68	G O6
			K^+ 702	3.90	2.72	G O6
PR0037[51]	RNA: signal recognition particle	1.52	K^+ 4001	3.96	2.75	G O6
			K^+ 4002	3.92	2.72	G O6
	Particle		K^+ 4003	3.21	2.80	G N7/O6-bridge
			K^+ 4003	3.00	3.8	G N7
AD0013[52]	A-DNA[a]	1.06	Cs^+ 22	3.50	3.56	DIN
AD0014[52]	A-DNA[b]	1.05	Cs^+ 22	3.54	3.45	DIN
AD0015[52]	A-DNA[c]	1.05	Rb^+ 22	3.38	3.27	DIN
AD0016[52]	A-DNA[d]	1.30	Rb^+ 22	3.79	3.49	DIN
AD0017[52]	A-DNA[e]	2.00	K^+ 22	3.64	3.33	DIN
AD0018[52]	A-DNA[f]	1.30	K^+ 22	3.83	3.50	DIN
AD0019[52]	A-DNA[g]	1.30	Na^+ 22	3.75	3.45	DIN
BD0054[13]	B-DNA	1.2	Tl^+ 2107	4.11	2.43	O6
			Tl^+ 2107	3.45	2.86	DIN
			Tl^+ 2103	2.27	2.93	G N7/O6-bridge
			Tl^+ 2101	2.75	2.95	G N7/O6-bridge
			Tl^+ 2108	2.93	2.47	G N7/O6-bridge
			Tl^+ 2102	2.93	2.63	G N7/O6-bridge
			Tl^+ 2102	4.82	3.06	G O6
			Tl^+ 2110	2.57	2.66	G N7/O6-bridge
			Tl^+ 2113	4.19	2.76	G O6
			Tl^+ 2113	3.79	3.07	G O6
UDB005[53]	DNA: dinucleotide	0.86	Na^+ 5	3.72	2.49	G O6
UDF025[54]	Z-DNA	1.92	Na^+ 13	3.73	2.58	G O6
DDF041[35]	DNA/anthracycline	1.5	Na^+ 17	2.56	3.51	G N7/DRU
			Na^+ 16	3.24	4.36	G N7/DRU
DDF020[2]	DNA/anthracycline	1.5	Na^+ 9	2.74	3.70	G N7/DRU
DDF001[1]	DNA/anthracycline	1.18	Na^+ 8	2.77	3.78	G N7/DRU
CGATCG–adriamycin–Tl^+	DNA/anthracycline	1.34	Tl^+ 1	2.75	2.90	G N7/O6-bridge
			Tl^+ 2	2.72	3.98	G N7/DRU
			Tl^+ 3	2.80	2.91	G N7/O6-bridge

[a] Modified with 2′-O-methyl-3′-methylenephosphonate at two residues per duplexes.
[b] Modified with 2′-O-methyl-[tri(oxyethyl).
[c] Modified with 2′-O-fluoroethyl.
[d] Modified with 2′-O-methyl-[tri(oxyethyl)].
[e] Modified with 2′-O-methyl-3′-methylenephosphonate.
[f] Modified with 2′-O-methyl-[tri(oxyethyl)].
[g] Modified with 2′-O-methyl-[tri(oxyethyl)].

units related by a twofold axis. The biological unit is numbered 5′ C(1)G(2) · · · G(6) 3′ and 5′ C(7)G(8) · · · · G(12) 3′. The chromophore is intercalated at the d(CpG) step, between base pairs C(1)–G(12) and G(2)–C(11) and by symmetry at the identical intercalation site at the other end of the complex. The aminosugar is located within the minor groove. The 3′-amino group interacts favorably with three oxygen atoms, the T(10) O2 (N—O distance = 3.0Å), the C(11) O2 (3.1Å), and C(11) O4′ (3.2Å). The 9-hydroxyl oxygen forms hydrogen bonds to both the N2 and N3 atoms of G(2). The adriamycin B–D ring systems (including their coplanar substituents) are in close van der Waals contact with the flanking base pairs. Fourteen atoms of B–D ring systems are within 3.4 Å of 23 DNA base atoms.

The combined structural results demonstrate that the anthracycline system is anchored within the minor groove, by a combination of direct drug–DNA interactions and water-mediated interactions. The anthra-

cycline is anchored at the other end, in the major groove, by water and ion-mediated interactions. The central portion of the intercalated chromophore engages in extensive stacking contacts with adjacent base pairs. The stacking interactions confer stability. An element of directionality to stacking interactions is suggested by the observation that the shortest subset of stacking contacts (<3.2Å) are between oxygen atoms and electron deficient carbon atoms. The anchoring interactions within the grooves, in possible conjunction with directional stacking interactions, restrict the intercalated chromophore to a well-defined and highly conserved position relative to the flanking base pairs.

Tl^+ Ions

Specific localization of Tl^+ ions is apparent from both $F_0 - F_c$ and $F_+ - F_-$ Fourier electron density maps calculated with $CGATCG_2$–adriamycin$_2$–Tl^+ data (Figure 3). The final model of $CGATCG_2$–adriamycin$_2$–Tl^+ contains six partially occupied Tl^+ sites (Figure 2). Three of these sites are crystallographically unique. All three localized Tl^+ sites are within the major groove. None are adjacent to phosphate groups or are involved in lattice interactions. Each site is directly coordinated by functional groups of guanine bases, and in some cases also by those of the intercalated adriamycin chromophore. Tl^+ sites are chelated by oxygen and basic endocyclic nitrogen atoms of DNA bases, but not by exocyclic amino nitrogen atoms. Each Tl^+ site is in contact with a guanine N7 position, and each guanine N7 position is in contact with a Tl^+ site. The localized Tl^+ ions are partially dehydrated. The water coordination numbers appear to be low, from 1 to 3. However, we must interpret the water interactions of Tl^+ with caution because water molecules scatter very weakly in comparison to Tl^+. A 30% occupied Tl^+ ion would be clearly visible in the electron density maps while a water molecule with the same occupancy would probably not be assignable.

The DNA and adriamycin combine to form a tetrachelator for the most highly occupied Tl^+ site (Tl^+ 1; about 50% occupancy, Figures 1 and 2). The Tl^+ 1 site is chelated by O6 (2.90Å) and N7 (2.75Å) atoms of G(6) and by the O4 (3.11 Å) and O5 (3.11 Å) atoms of adriamycin. This Tl^+ site is located in the plane of the guanine base. The adriamycin C4–O4–Tl^+ and C5–O5–Tl^+ angles are acute (103°). The Tl^+ 1 site is distinct from a previously identified Na^+ site[1,2] that is trichelated by the DNA–adriamycin complex. The Na^+ site is nearly superimposable a less highly occupied Tl^+ site (Tl^+ 2; about 20% occupancy). The Tl^+ 2 site was assigned to a Mg^{2+} in several structures,[33,34] although the coordination geometry in those structures is most consistent with a water molecule or a monovalent cation. The Tl^+ 2 site is 1.4 Å from the Tl^+ 1 site and unlike the Tl^+ 1 site does not interact with the O6 of G(6) (4.0 Å). Tl^+ 2 forms contacts with the N7 atom of G(6) (2.72 Å), and the O4 (3.14 Å) and O5 (3.03 Å) atoms of adriamycin. The occupancies of Tl^+ 2 and Tl^+ 1 must be exclusive. In a given complex either Tl^+ 2 or Tl^+ 1 is occupied by an ion; both sites would not be simultaneously occupied by ions. A third Tl^+ site (Tl^+ 3; about 20% occupied), is also located in the major groove, but on the opposite side of the adriamycin chromophore. Tl^+ 3 interacts with O6 (2.80 Å) and N7 (2.91 Å) atoms of G(2). The closest approach of Tl^+ 3 to adriamycin is 3.72 Å, from O12, suggesting at most a weak interaction between Tl^+ 3 and adriamycin. This partially dehydrated monovalent cation falls well within the volume occupied by a $Mg(H_2O)_6^{2+}$ observed in several previous DNA–anthracycline complexes.[35,37] Neither Mg^{2+} nor spermine is observable in the electron density of the $CGATCG_2$–adriamycin$_2$–Tl^+ complex. DNA–anthracycline complexes are characterized by variable spermine positions and occupancies.[2,32] If spermine is contained within the $CGATCG_2$–adriamycin$_2$–Tl^+ crystal, it is not sufficiently ordered to be observable and may lie within the disordered solvent channels between the duplexes.

Consensus Modes of Monovalent Cation Localization

A survey of the NDB[38] reveals that 103 localized monovalent cations (Na^+, K^+, Rb^+, Cs^+, or Tl^+) are contained in 37 nucleic acid entries of better than 2.0 Å resolution. These localized cations engage in a total of 376 contacts (149, values in parentheses exclude G-tetraplexes) of less than 3.5 Å with nitrogen and oxygen atoms of DNA or RNA. Most interactions of these localized monovalent cations, 87% (69%), are with bases, rather than backbone atoms. The majority of interactions, 89% (74%), are with oxygen atoms rather than nitrogen atoms. The O6 position of guanine is preferred, accounting for 64% (19%) of interactions. The sugar oxygen atoms (O2′, O3′, O4′ plus O5′) account for 7% (15%) of interactions. The O1P and O2P phosphate oxygen atoms similarly combine to engage in 7% (16%) of interactions. The O2 of T and U combine to account for 6% (9%) of interactions. These T O2 interactions are predominantly (86%) in B-form DNA. The only exceptions are in abbreviated A-form dinucleotide structures. The O4

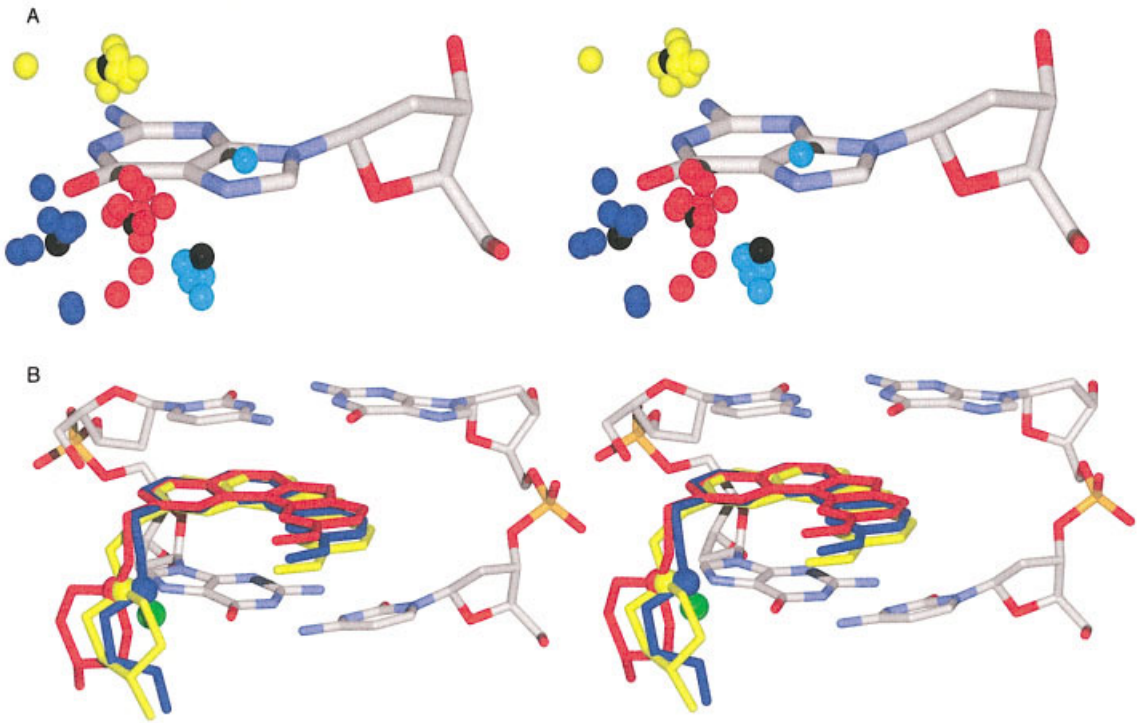

FIGURE 4 (A) Cations localize at consensus sites. Stereoview of ion positions obtained from superimposition of guanine residues that are adjacent to monovalent cations on their major groove edge (from NDB entries AD0013–AD0019, PR0052, PR0021, PR0037, UDB005, UDF025, DDF001, DDF020, DDF041, BD0054, and the current model). The DNA atoms shown are those of the reference guanine, G(2) of the current model, upon which other guanines have been superimposed. Inorganic monovalent cations are shown as colored spheres [blue, G O6 group; red, G O6/N7 bridging group; cyan, N7/DRU group; yellow DIN (dinucleotide) group, black, average positions of each group]. (B) DNA ligands locate cationic amino groups at consensus ion sites. Stereoview of three major groove binding/bis-intercalating DNA ligands (D232, DD0018; ditercalinium, DDD030; Flexi-di, DDDB46) intercalating CpG steps. The atoms of a single guanine of each complex were superimposed on G(2) of the current structure. D232 is red, ditercalinium is yellow, and Flexi-di is blue. The charged amino groups of each ligand are shown as spheres. The average ion position of the bridging O6/N7 group is shown as a green sphere. The DNA and ligands have been truncated for clarity. These figures were rendered with Povray 3.1. DNA atoms are shown in stick representation and colored CPK.

of T and U combine to account for 5% (11%) of interactions. These O4 interactions are exclusively in A-form DNA or RNA, with no occurrences in B-form DNA. The N7 of guanine is preferred among nitrogen atoms, engaging in 5% (12%) of interactions. Significant numbers of localized cations with contacts with other nitrogen atoms are not contained in the comparison set.

Localized monovalent cations adjacent to guanine are clustered into four distinct groups. Twenty-nine of the cations found in the NDB are located within 5.0 Å of both O6 and N7 of a given guanine residue (these criteria exclude cations in G-tetraplexes). Four of these ions are associated with an additional guanine base such that 33 cation contacts involve 31 guanine residues, in addition to the complex described here. A superimposition of these 31 guanine residues suggests that the 33 cation positions are distributed into four statistically distinct regions relative to the N7 and O6 atoms of guanine and the plane of the guanine base (Figure 4A). One group (G O6/N7-bridging cations) is composed of cations that are located close to the plane of the base and are less than 3.5 Å from both N7 and O6 of the same guanine. Most of the G O6/N7-bridging cations are adjacent to duplex B-DNA or A-RNA. Two of the Tl^+ ions of the structure described here are O6/N7-bridging cations. The second group (G O6 cations) is composed of ions that are located close to the plane of the base and are less than 3.5 Å from O6 but greater than 3.5 Å from N7. Most

of the G O6 cations are adjacent to duplex B-DNA or A-RNA. The third group (G N7/DRU cations, for drug, because these ions interact with a bound drug and the N7 of G) is composed of ions that are less than 3.5 Å from N7 and greater than 3.5 Å from O6. With the exception of one sodium ion that interacts with an adjacent phosphate group of RNA, all cations that contact N7 and not O6 of a given G are in contact with oxygen atoms of an intercalated anthracycline. One of the Tl^+ ions of the structure described here is a N7/DRU cation. A fourth group (DIN cations, for dinucleotide, because these ions interact with bases of a dinucleotide step) are in contact most generally with the O6, but in some cases with the N7, of a guanine (the primary guanine) and with a second exocyclic oxygen atom. The second exocyclic oxygen atom is either the O4 of a 3′ thymine or uracil residue of an A-form GpT or GpU step or the O6 of a 3′ cross-strand guanine residue of a B-form GpC step. DIN cations are displaced from the plane of primary guanine, in the 3′ direction.

Consensus Water Localization

As noted previously,[2] several water molecules appear to be conserved in DNA–anthracycline structures. Here, the analysis of water localization has been extended to include more recent high resolution DNA–anthracycline structures, with four crystal forms and variety of DNA sequences and anthracycline modifications. We have determined the average water population at consensus sites, summarized by pie representations in Figure 1. The average water population at a specific site is defined here by the total number of water molecules in the comparison set that interact with a given anthracycline functional group, normalized to the number of structures that contain that functional group. For example, adriamycin contains a hydroxyl group at the 14 position whereas daunomycin does not. There are five adriamycin structures in the comparison set. In those five structures a total of nine waters are within the distance threshold to give an average of 1.8 water molecules per 14-OH group (Figure 1). All DNA–anthracycline structures with resolution better than 1.8 Å were included in the comparison set, except NDB entry DDF035, which displays unreasonable geometry. Water sites were defined by proximity to oxygen or nitrogen atoms of the anthracycline. A site is defined by $r_{iw} < 3.4$Å, where "i" is an anthracycline atom and "w" is a water molecule. Some of the most highly conserved water molecules form bridges from anthracycline to DNA. A universally conserved bridging water molecule links the anthracycline O13 position to the O2 of a flanking pyrimidine. A second water bridge links the 3-amino group to the O2 (T) or N3 (A) position on the floor of the minor groove. A third bridging site involves the O4 and O5 positions. This set of bridging interactions is highly conserved, but the composition of the bridge appears to be variable. The bridge is composed of water molecules in some structures and ions in others. On average, carbonyl oxygens are more highly hydrated than hydroxyl oxygens. Hydroxyl groups 9, 11, and 6 are "buried" in the complex and do not interact with solvent.

Experimental Variability

Fully occupied Na^+ ions are contained at a single site in the major groove of several previous DNA–adriamycin and DNA–daunomycin complexes, coordinating O4 and O5 of daunomycin and N7 of guanosine.[1,2] Other DNA–anthracycline complexes lack monovalent cations. Here we observed three partially occupied monovalent cation sites. The variation might arise in part from differences in properties of Tl^+ and Na^+ ions, and from differences in crystallization conditions. An additional source of variation is probably associated with differences in experimental data. Tl^+ is distinguished from water by differential x-ray scattering. Na^+ ions and water molecules, in contrast, are nearly indistinguishable by x-ray scattering. The fully occupied Na^+ site was assigned to those models by geometric criteria. By that method it might be very difficult to characterize a mixed occupancy site.

DISCUSSION

The positions of cations in x-ray structures are modulated by sequence, conformation, and ligand interactions. Six localized Tl^+ sites are observable adjacent to each $CGATCG_2$–$adriamycin_2$–Tl^+ complex. Each of these localized monovalent cations is found within the G-tract major groove of the intercalated DNA–drug complex (Figure 2). Tl^+ ions interact with N7 and O6 atoms of guanine residues and with oxygen atoms of the intercalated adriamycin chromophore (Figure 1). Consistent with previous observations,[13,39] localized cations in the $CGATCG_2$–$adriamycin_2$–Tl^+ avoid exocyclic amino groups of DNA, which are relatively electropositive. Similarly, the localized cations adjacent to $CGATCG_2$–$adriamycin_2$–Tl^+ are not found to be adjacent to phosphate groups or to be involved in lattice interactions. The combined results, including recent work by ourselves and others,[12,40] the survey of the NDB described here, and the struc-

ture of $CGATCG_2$–$adriamycin_2$–Tl^+ described here, suggest general patterns of monovalent cation localization. Localized monovalent cations (a) partially dehydrate and interact directly with the functional groups of nucleic acid bases and ligands; (b) interact preferentially with oxygen atoms over nitrogen atoms; (c) interact with basic endocyclic nitrogen positions of DNA bases but not with exocyclic amino nitrogen atoms; (d) are most commonly found in the major groove; (e) when in the major groove of A-form or B-form duplexes associate preferentially with guanine; (f) when in the major groove of A-form duplexes, but not B-form duplexes, commonly associate with the O4 of thymine; (g) when in the minor groove of B-form duplexes associate preferentially with the O2 of thymine; (h) can interact simultaneously with ligand and DNA.

Asymmetric Electrostatic Landscape: The Major Groove

In B-form DNA, cation localization is asymmetric in that the major groove edge of G-tracts and the minor groove edge of A-tracts favor cation localization. Partially dehydrated monovalent cations,[13] hydrated divalent cations,[41,42] and polyamines[39,42] localize in a common region in the major groove of B-form CGCGAATTCGCG. The O6 and N7 of guanine residues chelate the ions directly (most common for monovalent cations) or via water molecules (most common for divalent cations). In at least a subset of dodecamer x-ray structures a $Mg(H_2O)_6^{2+}$ complex within the CGCGAATTCGCG major groove shares occupancy with monovalent cations.[13] Here we demonstrate that monovalent cation localization can be maintained within the major groove after specific types of DNA ligand binding and deformation, such as intercalation of an anthracycline.

Adriamycin appears to be designed by nature to interact favorably with, and possibly enhance, the asymmetric electrostatic landscape surrounding B-DNA. Prepositioned monovalent cations adjacent to unliganded B-DNA are retained, and interact favorably with the intercalated anthracycline and with DNA. The interactions involve various combinations of the N7 and O6 positions of guanine and the O4 and O5 positions of the anthracycline. The anthracycline chromophore acts in concert with the DNA in stabilizing localized cations, which in turn contributes to a preorganized binding environment for adriamycin.

The most highly occupied cation site in the structure described here (Tl^+ 1, Figures 2B and 2C) is tetracoordinated by the DNA–drug complex, by a guanine residue (N7 and O6) and an adriamycin molecule (O4 and O5). Thus the intercalation of adriamycin at a CpG step can apparently increase the monovalent cation coordination number from a maximum of three in unliganded B-DNA[13] to four. This increase in coordination number appears to be associated with increased Tl^+ occupancy. A third Tl^+ site (Tl^+ 3) in the $CGATCG_2$–$adriamycin_2$–Tl^+ structure occupies the region in the major groove where a $Mg(H_2O)_6^{2+}$ complex is located in several previous DNA–anthracycline structures.[35,37] The location of this shared monovalent and divalent cation site is consistent with the general pattern of cation localization in the major groove of G-tracts.[13]

Asymmetric Electrostatic Landscape: The Minor Groove

Monovalent cations localize in the minor groove of B-DNA A-tracts.[12,40] Some of these cations would be displaced by the cationic amino group of adriamycin. The position and DNA interactions of this amino group are similar to those of localized monovalent cations. Like monovalent cations in the minor groove of B-DNA, the amino group in the $CGATCG_2$–$adriamycin_2$ complex interacts with the O2 atoms of pyrimidines and the O4′ atoms of deoxyriboses. Therefore the localization of cationic charge is preserved in the minor groove of B-DNA, but by a substitution mechanism rather than by retention of prepositioned cations, as in the major groove (above).

Ligand Design

Favorable interaction of localized inorganic cations and cationic functional groups of adriamycin with the asymmetric electrostatic environment of DNA may be a general, but unrecognized, feature of DNA–small molecule interactions. Some previously characterized minor groove binders[43,44] and anthracyclines position cationic functional groups adjacent to the minor groove edges of AT base pairs, and so replace recently discovered localized inorganic cations.[12,40] Previously characterized major groove binders[45–48] position cationic functional groups adjacent to the major groove edges of CG base pairs, and so replace recently discovered localized inorganic cations (Figure 4B). In fact, favorable interaction of cationic charge at the N7/O6 bridging position provides a compelling structural rationale for the major groove selectivity of the bipypiridine linker of ditercalinium. Positioning of the linker in the major groove allows cationic amino groups to mimic inorganic cations and complement the electronegative functional groups of DNA bases. If the linker were switched to the minor

groove of the GC-tract recognition sequence, cationic amino groups of the linker would be unfavorably positioned near electropositive amino groups of DNA. Thus the asymmetric electrostatic landscape of DNA appears to drive groove selection.

To our knowledge, complementarily to the sequence-specific asymmetrical electronegative functional groups of DNA has not been explicitly or deliberately incorporated into the design of DNA-binding ligands. In one example of how this design strategy might be implemented, substitution of a $—CH_2—NH_3^+$ at the O4 position of adriamycin would place a cation in the O6/N7 bridging position of an adjacent guanine and should increase affinity and specificity.

The authors thank Drs. Nick Hud, Angus Wilkinson, and Allen Orville for helpful discussions.

REFERENCES

1. Wang, A. H.; Ughetto, G.; Quigley, G. J.; Rich, A. Biochemistry 1987, 26, 1152–1163.
2. Frederick, C. A.; Williams, L. D.; Ughetto, G.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Wang, A. H.-J. Biochemistry 1990, 29, 2538–2549.
3. Williams, L. D.; Egli, M.; Ughetto, G.; van der Marel, G. A.; van Boom, J. H.;, Quigley, G. J.; Wang, A. H.-J.; Rich, A.; Frederick, C. A. J Mol Biol 1990, 215, 313–320.
4. Lipscomb, L. A.; Peek, M. E.; Zhou, F. X.; Bertrand, J. A.; VanDerveer, D.; Williams, L. D. Biochemistry 1994, 33, 3649–3659.
5. Chaires, J. B.; Fox, K. R.; Herrara, J. E.; Britt, M.; Waring, M. J. Biochemistry 1987, 26, 8227–8236.
6. Capranico, G.; Kohn, K. W.; Pommier, Y. Nucleic Acids Res 1990, 18, 6611–6619.
7. Chaires, J. B.; Satyanarayana, S.; Suh, D.; Fokt, I.; Przewloka, T.; Priebe, W. Biochemistry 1996, 35, 2047–2053.
8. Qu, X.; Chaires, J. B. J Am Chem Soc 2001, 123, 1–7.
9. Leonard, G. A.; Brown, T.; Hunter, W. N. Eur J Biochem 1992, 204, 69–74.
10. Hu, G. G.; Shui, X.; Leng, F.; Priebe, W.; Chaires, J. B.; Williams, L. D. Biochemistry 1997, 36, 5940–5946.
11. Shui, X.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. Biochemistry 1998, 37, 8341–8355.
12. McFail-Isom, L.; Sines, C.; Williams, L. D. Curr Opin Struct Biol 1999, 9, 298–304.
13. Howerton, S. B.; Sines, C. C.; VanDerveer, D.; Williams, L. D. Biochemistry 2001, 40, 10023–10031.
14. Brown, I. D. Acta Crystallogr 1988, B44, 545–553.
15. Wulfsberg, G. Principles of Descriptive Inorganic Chemistry; University Science Books: Sausalito, CA, 1991.
16. Pedersen, P. A.; Nielsen, J. M.; Rasmussen, J. H.; Jorgensen, P. L. Biochemistry 1998, 37, 17818–17827.
17. Villeret, V.; Huang, S.; Fromm, H. J.; Lipscomb, W. N. Proc Natl Acad Sci USA 1995, 92, 8916–8920.
18. Loria, J. P.; Nowak, T. Biochemistry 1998, 37, 6967–6974.
19. Basu, S.; Szewczak, A. A.; Cocco, M.; Strobel, S. A. J Am Chem Soc 2000, 122, 3240–3241.
20. Basu, S.; Rambo, R. P.; Strauss-Soukup, J.; Cate, J. H.; Ferre-D'Amare, A. R.; Strobel, S. A.; Doudna, J. A. Nat Struct Biol 1998, 5, 986–992.
21. Badger, J.; Kapulsky, A.; Gursky, O.; Bhyravbhatla, B.; Caspar, D. L. Biophys J 1994, 66, 286–892.
22. Badger, J.; Li, Y.; Caspar, D. L. Proc Natl Acad Sci USA 1994, 91, 1224–1228.
23. Gill, H. S.; Eisenberg, D. Biochemistry 2001, 40, 1903–1912.
24. Clowney, L.; Jain, S. C.; Srinivasan, A. R.; Westbrook, J.; Olson, W. K.; Berman, H. M. J Am Chem Soc 1996, 118, 509–518.
25. Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S.-H.; Olson, W. K.; Berman, H. M. J Am Chem Soc 1996, 118, 519–529.
26. Parkinson, G.; Vojtechovsky, J.; Clowney, L.; Brunger, A. T.; Berman, H. M. Acta Crystallogr, Sect D: Biol Crystallogr 52, 57–64.
27. Neidle, S.; Taylor, G. Biochim Biophys Acta 1977, 479, 450–459.
28. Kleywegt, G. J.; Jones, T. A. Acta Crystallogr, Sect D: Biol Crystallogr 1999, 55, 941–944.
29. Dauter, Z.; Adamiak, D. A. Acta Crystallogr D, Biol Crystallogr 2001, 57, 990–995.
30. Martin, A. C. R. www.rubic.rdg.ac.uk/~andrew/bioinf.org/software/profit/index.html.
31. Schuerman, G. S.; Van Meervelt, L. J Am Chem Soc 122, 232–240.
32. Williams, L. D.; Frederick, C. A.; Ughetto, G.; Rich, A. Nucleic Acids Res 1990, 18, 5533–5541.
33. Wang, A. H.-J.; Gao, Y.-G.; Liaw, Y.-C.; Li, Y.-K. Biochemistry 1991, 30, 3812–3815.
34. Gao, Y. G.; Liaw, Y. C.; Li, Y. K.; van der Marel, G. A.; van Boom, J. H.; Wang, A. H. Proc Natl Acad Sci USA 88, 4845–4849.
35. Cirilli, M.; Bachechi, F.; Ughetto, G.; Colonna, F. P.; Capobianco, M. L. J Mol Biol 1993, 230, 878–889.
36. Berger, I.; Su, L.; Spitzner, J. R.; Kang, C.; Burke, T. G.; Rich, A. Nucleic Acids Res 1995, 23, 4488–4494.
37. Gao, Y. G.; Robinson, H.; Wijsman, E. R.; vander-Marel, G. A.; vanBoom, J. H.; Wang, A. H. J. J Am Chem Soc 1997, 119, 1496–1497.
38. Berman, H. M.; Zardecki, C.; Westbrook, J. Acta Crystallogr, Sect D: Biol Crystallogr 1998, 54, 1095–1104.
39. Shui, X.; Sines, C.; McFail-Isom, L.; VanDerveer, D.; Williams, L. D. Biochemistry 1998, 37, 16877–16887.
40. Hud, N. V.; Polak, M. Curr Opin Struct Biol 2001, 11, 293–301.

41. Minasov, G.; Tereshko, V.; Egli, M. J Mol Biol 1999, 291, 83–99.
42. Sines, C. C.; McFail-Isom, L.; Howerton, S. B.; VanDerveer, D.; Williams, L. D. J Am Chem Soc 2000, 122, 11048–11056.
43. Pjura, P. E.; Grzeskowiak, K.; Dickerson, R. E. J Mol Biol 1987, 197, 257–271.
44. Teng, M.; Usman, N.; Frederick, C. A.; Wang, A. H.-J. Nucleic Acids Res 1988, 16, 2671–2690.
45. Gao, Q.; Williams, L. D.; Egli, M.; Rabinovich, D.; Chen, S.-H.; Quigley, G. J.; Rich, A. Proc Natl Acad Sci USA 88, 2422–2426.
46. Williams, L. D.; Gao, Q. Biochemistry 1992, 31, 4315–4324.
47. Peek, M. E.; Lipscomb, L. A.; Bertrand, J. A.; Gao, Q.; Roques, B. P.; Garbay-Jaureguiberry, C.; Williams, L. D. Biochemistry 1994, 33, 3794–3800.
48. Peek, M. E.; Lipscomb, L. A.; Haseltine, J. Gao, Q.; Roques, B. P.; Garbay-Jaureguiberry, C.; Williams, L. D. J Bioorg Med Chem 1995, 3, 693–699.
49. Yang, X.; Gerczei, T.; Glover, L. T.; Correll, C. C. Nat Struct Biol 2001, 8, 968–973.
50. Batey, R. T.; Rambo, R. P.; Lucast, L.; Rha, B.; Doudna, J. A. Science 2000, 287, 1232–1239.
51. Batey, R. T.; Sagar, M. B.; Doudna, J. A. J Mol Biol 2001, 307, 229–246.
52. Tereshko, V.; Wilds, C. J.; Minasov, G.; Prakash, T. P.; Maier, M. A.; Howard, A.; Wawrzak, Z.; Manoharan, M.; Egli, M. Nucleic Acids Res 2001, 29, 1208–1215.
53. Coll, M.; Solans, X.; Font-Altaba, M.; Subirana, J. A. J Biomol Struct Dynam 1987, 4, 797–811.
54. Malinina, L.; Urpi, L.; Salas, X.; Huynh-Dinh, T.; Subirana, J. A. J Mol Biol 1994, 243, 484-493.