Locating Monovalent Cations in the Grooves of B-DNA[†,‡]

Shelley B. Howerton, Chad C. Sines, Don VanDerveer, and Loren Dean Williams*

School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400

Received February 26, 2001; Revised Manuscript Received June 14, 2001

ABSTRACT: Here we demonstrate that monovalent cations can localize around B-DNA in geometrically regular, sequence-specific sites in oligonucleotide crystals. Positions of monovalent ions were determined from high-resolution X-ray diffraction of DNA crystals grown in the presence of thallium(I) cations (Tl^+). Tl^+ has previously been shown to be a useful K^+ mimic. Tl^+ positions determined by refinement of model to data are consistent with positions determined using isomorphous $F_{Tl} - F_K$ difference Fouriers and anomalous difference Fouriers. None of the observed Tl^+ sites surrounding CGCGAATTCGCG are fully occupied by Tl^+ ions. The most highly occupied sites, located within the G-tract major groove, have estimated occupancies ranging from 20% to 35%. The occupancies of the minor groove sites are estimated to be around 10%. The Tl^+ positions in general are not in direct proximity to phosphate groups. The A-tract major groove appears devoid of localized cations. The majority of the observed Tl^+ ions interact with a single duplex and so are not engaged in lattice interactions or crystal packing. The locations of the cation sites are dictated by coordination geometry, electronegative potential, avoidance of electropositive amino groups, and cation$-\pi$ interactions. It appears that partially dehydrated monovalent cations, hydrated divalent cations, and polyamines compete for a common binding region on the floor of the G-tract major groove.

Understanding nucleic acid folding, deformation, and conformational heterogeneity in X-ray structures requires that positions of localized counterions be known. At least a subset of localized magnesium ions have been observed in close proximity to RNA (*1*, *2*) and DNA (*3*, *4*). Magnesium ions are identified primarily by distinctive octahedral coordination geometry. By contrast, positions of sodium or potassium cations in crystal structures of nucleic acids are difficult to determine even with ultrahigh-resolution data (*5*), with a few notable exceptions (*6*, *7*). Sodium and potassium ions and water molecules have irregular coordination geometry.

Here we observe that, in B-DNA crystals, partially dehydrated monovalent cations can localize in sequence-specific, geometrically regular sites within the major and minor grooves (Figures 1 and 2). Monovalent cation sites are located primarily within the G-tract major groove and the A-tract minor groove. The locations of monovalent cations appear to be influenced by base and backbone functional groups and by electrostatic interactions.

An atomic resolution structure of the duplex [d(CGCGAATTCGCG)]$_2$ was obtained from a crystal grown in the presence of thallium (Tl^+), along with magnesium and spermine. The characteristics of Tl^+ exploited here are its strong scattering (*8*) and anomalous scattering (*9*) of X-rays, providing a sensitive detection system. Tl^+ has previously been shown to be a useful K^+ mimic. Tl^+ and K^+ (i) have similar ionic radii [$K^+ = 1.33$ Å, $Tl^+ = 1.49$ Å] and irregular coordination geometries (*10*), (ii) have similar enthalpies of hydration [$K^+ = -77$ kcal/mol, $Tl^+ = -78$ kcal/mol] (*11*), (iii) both stabilize DNA G-quartets (*12*), and (iv) are nearly interchangeable in sodium−potassium pumps (*13*) and in catalytic mechanisms of fructose-1,6-bisphosphatase (*14*) and pyruvate kinase (*15*). Doudna and co-workers used Tl^+ as a probe for K^+ in the *Tetrahymena* ribozyme P4−P6 domain (*7*) and located three positions. Caspar used $T1^+$ to locate six K^+ positions adjacent to insulin (*9*). With both the *Tetrahymena* ribozyme P4−P6 domain and insulin, Tl^+ ions make contacts primarily with uncharged macromolecular oxygen and nitrogen atoms and water molecules. These interactions are very similar to our observations with B-DNA here. The utility of Tl^+ is supported by comparison of the Tl^+ positions with those of rubidium (Rb^+). Like Tl^+, Rb^+ sites are observed within the A-tract minor groove (*16*) and the G-tract major groove of CGCGAATTCGCG (S. B. Howerton and L. D. Williams, unpublished). However, in interpreting the results described here, it must be considered that Tl^+ is not a biological cation. Soft ligands such as sulfur (which are absent in our system) will interact differently with Tl^+ than with K^+. In addition, Tl^+ coordination is characterized by more variable contact distances than K^+ coordination.

MATERIALS AND METHODS

Crystallization and Data Collection. Crystals were grown in sitting drops by vapor diffusion from a solution containing a 1.0 mM amount of the ammonium salt of reverse-phase HPLC-purified d(CGCGAATTCGCG) (Midland Certified

[†] This work was supported by the National Science Foundation (Grant MCB-9976498) and the American Cancer Society (Grant RPG-95-116-03-GMC).

[‡] Atomic coordinates and structure factors have been deposited in the Nucleic Acid Database (entry code BD0054).

* To whom correspondence should be addressed. E-mail: loren.williams@chemistry.gatech.edu. Phone: (404) 894-9752. Fax: (404) 894-7452.

10.1021/bi010391+ CCC: $20.00
Published on Web 08/02/2001

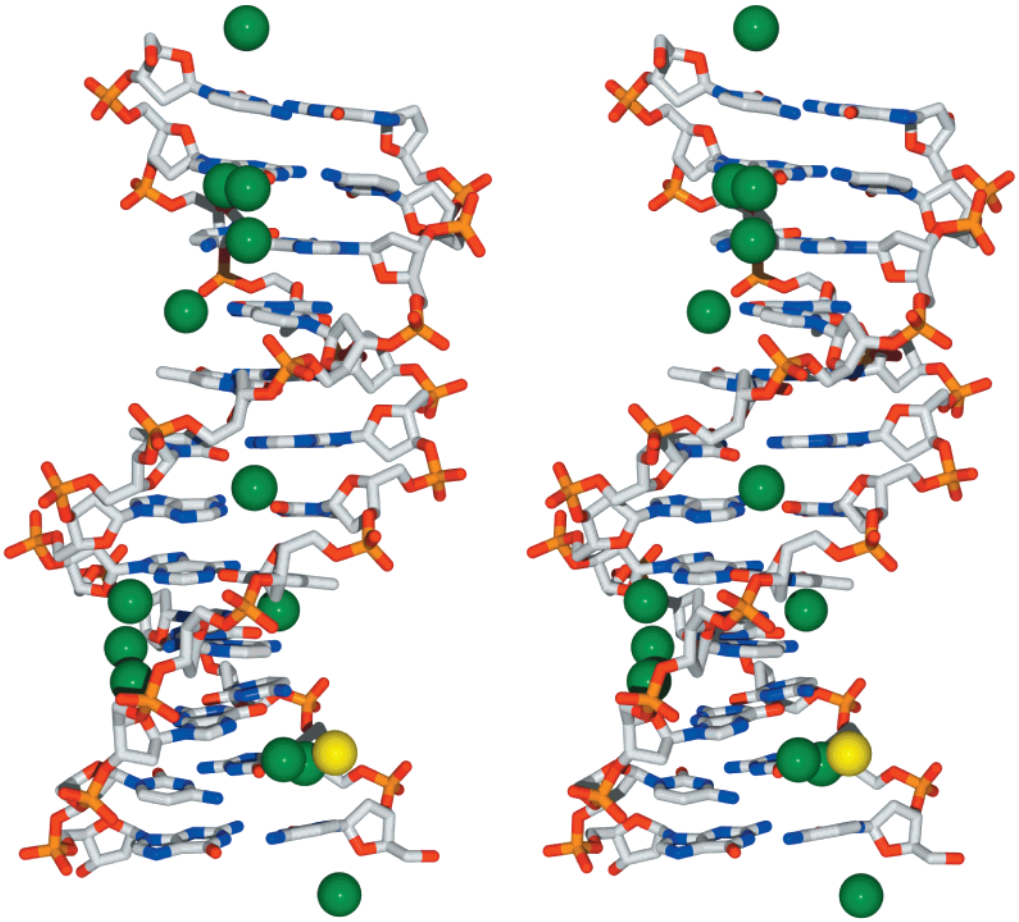

FIGURE 1: Stereoview of duplex CGCGAATTCGCG and associated ions. Tl^+ ions are depicted by green spheres. The magnesium ion is depicted by a yellow sphere. The DNA is shown in stick representation with standard CPK color coding of atoms. Water molecules have been omitted for clarity. The figure was rendered using Povray 3.1.

Reagent Co., Midland, TX), 19 mM thallium(I) acetate (pH 6.4), 5.2 mM magnesium acetate, 3.8% 2-methyl-2,4-pentanediol (MPD), and 8.9 mM spermine acetate. The crystallization solution was equilibrated against a reservoir of 35% MPD at 22 °C. A crystal grew to 0.7 × 0.5 × 0.5 mm^3 within a few weeks. Thus far, attempts to obtain crystals from solutions with increased concentrations of thallium(I) acetate yielded poor-quality crystals. X-ray diffraction data used for the refinement and the $F_{Tl} - T_K$ isomorphous difference calculations were collected at Brookhaven National Laboratory on beamline X26-C with an ADSC Quantum 4 CCD detector using 1.1 Å radiation. The crystal was maintained at −160 °C during data collection. A total of 290 894 reflections were indexed and integrated using MOSFLM 6.0 (*17*) and reduced to 21 760 unique reflections with SCALA (*18*). Unit cell dimensions are $a = 25.94$ Å, $b = 40.74$ Å, and $c = 66.20$ Å in space group $P2_12_12_1$. Data used for refinement included 20 997 unique reflections from 35 to 1.2 Å. Table 1 gives data collection and refinement statistics.

For calculation of the anomalous difference Fourier, a data set was collected from the same crystal, over 360° in ϕ at −180 °C using 1.54 Å Cu Kα radiation from an in-house Rigaku/MSC rotating anode generator with Osmic blue confocal mirrors and an R-AXIS4^{++} image plate detector. A total of 131 329 reflections to 1.55 Å resolution were reduced to a unique data set containing 10 320 reflections, preserving Bijvoet pairs using the dtprocess software in the program CrystalClear 1.3.0 ($R_{merge} = 0.048$, 99.4% complete, $\Delta F/F = 0.060$).

$F_{Tl} - F_K$ and Bijvoet Difference Fouriers. To calculate isomorphous $F_{Tl} - F_K$ difference Fouriers, high-resolution CGCGAATTCGCG/K^+ data [NDB (*19*) entry BD0041] described previously (*20*) were reindexed and reintegrated using MOSFLM 6.0, rereduced with SCALA (*18*), and anisotropically scaled by shell to the CGCGAATTCGCG/Tl^+ data using the Xmerge routine of XtalView (*21*). Reflections from 21 to 1.6 Å were scaled. The assumption of isomorphism of the DNA in the CGCGAATTCGCG/Tl^+ and CGCGAATTCGCG/K^+ crystals was tested in a variety of ways. The rms deviations between the final CGCGAATTCGCG/Tl^+ and CGCGAATTCGCG/K^+ models (all DNA atoms) of 0.32 Å indicate that the structures are isomorphous. The differences in unit cell parameters of $\Delta a = 1.3\%$, $\Delta b = 0.2\%$, and $\Delta c = 0.7\%$ are within generally accepted limits. Finally, the calculated and observed R_{iso} values are similar. The observed R_{iso} ($\sum|F_{Tl} - F_K|/|F_K|$) for common reflections between 21 and 1.6 Å resolution is 0.29 (average over all shells). This value is similar to the predicted R_{iso} calculated with the method of Crick and Magdoff (*22*). For the final model, with 13 Tl^+ ions at 17% average occupancy, the calculated R_{iso} equals 0.27.

$F_{Tl} - F_K$ difference electron density Fouriers were calculated using the scaled $F_{Tl} - F_K$ as coefficients and phases from the published CGCGAATTCGCG/K^+ model. Bijvoet difference ($F_+ - F_-$) electron density Fouriers were

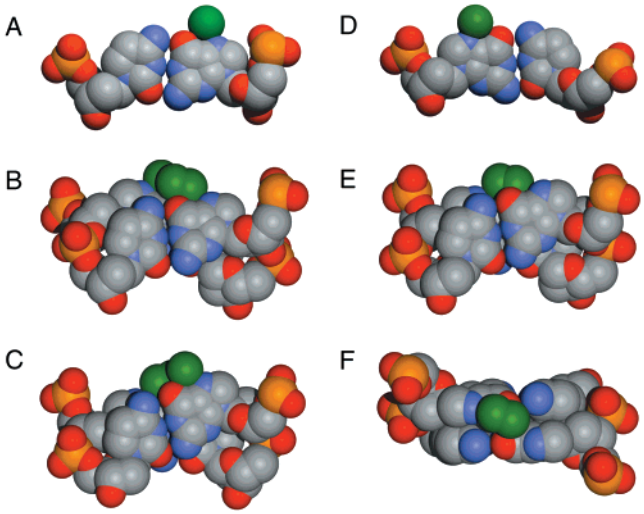

FIGURE 2: Major groove interactions of Tl^+ ions with guanine residues. DNA atoms, shown in space-filling models, are colored by the CPK standard. Positions of partially occupied Tl^+ are indicated by green spheres. (A) Tl^+ ion 2101 is located within the plane of G(2104), adjacent to the O6 and the N7 positions. (B) Tl^+ ions 2102 and 2113 are located within the plane of G(2002). Tl^+ ion 2102 is adjacent to the O6 and N7 positions while Tl^+ ion 2113 is adjacent to the O6 position. Tl^+ ion 2110 is visible behind Tl^+ ions 2102 and 2113. (C) Tl^+ ion 2110 is adjacent to the O6 and N7 positions of G(2110). This view contains the same atoms as in panel B but is rotated 180° about a vertical line. (D) Tl^+ ion 2103 is located in the plane of G(1004) and is adjacent to the O6 and N7 positions. (E) Tl^+ ions 2107 and 2108 are superimposed on the major groove $Mg(H_2O)_6^{2+}$, which is not shown. Tl^+ ion 2108 is adjacent to the N7 and O6 positions of G(1002). Tl^+ ion 2107 is located between the planes of two base pairs and is adjacent to the O6 positions of both G(2010) and G(1002). (F) Same atoms as in panel E rotated by 90° about a horizontal line.

calculated using $F_+ - F_-$ as coefficients and the same CGCGAATTCGCG/K^+ phases. Pairs of reflections with $|F|/\sigma(F) < 1.0$ for either Bijvoet were excluded. The map was calculated using data to 2.0 Å resolution. The map showed peaks from Tl^+ ions and weaker peaks from the anomalous scattering of phosphorus atoms.

Difference Patterson Fouriers. Patterson Fouriers were calculated using $(F_+ - F_-)^2$ or scaled $(F_{Tl} - F_K)^2$ as coefficients. A search of the $(F_+ - F_-)^2$ Patterson using the program CNS (*23*) with data to 2 Å resolution yielded three heavy atom sites, which after enantiomeric correction correspond to the three most highly occupied Tl^+ sites in the final refined model. An $F_+ - F_-$ Fourier was calculated using phases from the three sites identified in the $(F_+ - F_-)^2$ Patterson search. Three additional Tl^+ sites corresponding to those in the refined model were clearly indicated in that map. In total six Tl^+ sites were identified from the $F_{Tl} - F_K$ data, without additional phase information.

Thirteen of the highest 15 peaks in the predicted $F_{Tl} - F_K$ difference Patterson map (calculated with the positions of the 13 Tl^+ ions located in the final refined model) correspond to peaks in the observed difference Patterson map. Seven of the eight most intense peaks lying on Harker planes in the predicted map correspond to peaks in the observed map. Peaks lying in Harker planes in the observed difference Patterson map were used to manually solve the position of the most highly occupied Tl^+ site (2101). Automated solution of the $F_{Tl} - F_K$ difference Patterson was not attempted.

Table 1: Crystallographic and Refinement Statistics

unit cell	
α, *β*, *γ* (deg)	90
a (Å)	25.94
b (Å)	40.74
c (Å)	66.20
space group	$P2_12_12_1$
temp of data collection (°C)	−110
no. of reflections	290894
no. of unique reflections	21760
completeness (%)/highest shell (%)	93/68
max resolution of obsd reflections (Å)	1.14
max resolution of highest shell used in refinement (Å)	1.20
resolution range (Å)	35−1.2
no. of reflections used in refinement	18897
no. of reflections used in test set	2100
rmsd of bonds from ideality (Å)	0.010
rmsd of angles from ideality (deg)	0.030
DNA (asymmetric unit)	$[d(CGCGAATTCGCG)]_2$
no. of DNA atoms	486
no. of water molecules, excluding Mg first shell	116 full, 18 partial
no. Mg ions plus coordinating water molecules.	7 partial
no. of Tl^+ ions/summed occupancy	13/2.26
no. of spermine atoms	0
R-free (%)	22.2
R-factor (%) excluding test set data	16.9
final *R*-factor (using all data, 20997 reflections)	17.3

Refinement. A starting model consisting of DNA coordinates from the high-resolution sodium form of $[d(CGCGAATTCGCG)]_2$ (NDB entry BDL084) (*24*) was annealed and refined against the CGCGAATTCGCG/Tl^+ data with the program CNS, using parameters of Berman and coworkers (*25*−*27*). Water molecules were added iteratively to peaks of corresponding sum and difference density followed by refinement and phase recalculation. Some water molecules were converted to partially occupied Tl^+ ions using fixed criteria as described in the Results section. The refinement was not biased by the $F_{Tl} - F_K$ or $F_+ - F_-$ difference electron density Fourier in that no attempt was made to place Tl^+ ions in positions indicated there. After convergence with CNS, refinement was continued using SHELX-97 (*28*). A diffuse solvent correction, SWAT, was applied to simulate disorder in the unassigned solvent region. DELU restraints were applied to bonded atoms to equalize each anisotropic vector component parallel to the bond. A similar restraint, SIMU, was applied to atoms that are nonbonded but near in space. The parameter file was modified to remove the restraint of C1′ atom coplanarity with the aromatic ring of the base to which it is attached (*5*). Displacement parameters were refined anisotropically. The *R*-free, goodness of fit, and thermal ellipsoids were monitored to avoid overrefinement. Thermal ellipsoids (Figure 3) were computed with ORTEP-3 for Windows (*29*, *30*). Table 1S (Supporting Information) gives statistics describing the refinement progression.

RESULTS

The final CGCGAATTCGCG/Tl^+ model contains 13 partially occupied monovalent cation sites and one partially occupied magnesium ion but lacks spermine. The number of observed monovalent cation sites in this structure and

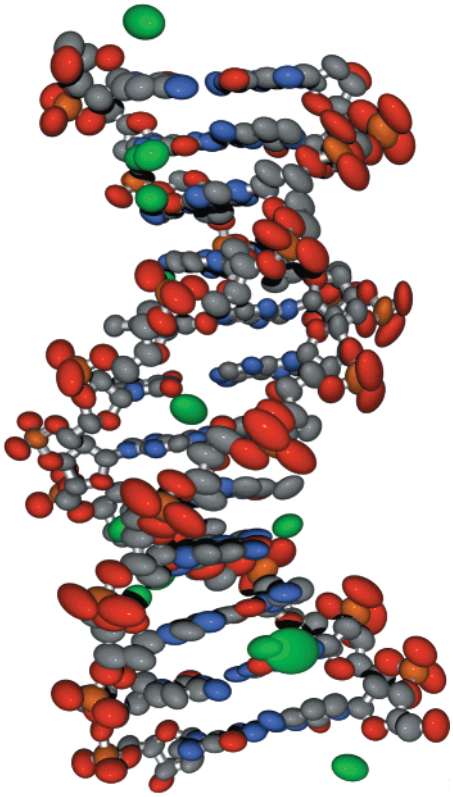

FIGURE 3: Thermal ellipsoids of $[d(CGCGAATTCGCG)]_2$ and Tl^+ ions at 50% probability. Tl^+ ions are in green. Water molecules and $Mg(H_2O)_6^{2+}$ were omitted for clarity. The plot was calculated with Ortep-3 for Windows (*29*) and rendered with Povray 3.1.

some of their locations differ significantly from previous observations within DNA crystal structures. Therefore, extreme care was taken during refinement and construction of the model and during structure validation. Several steps were taken to maximize the accuracy of the model and to avoid bias during the refinement. (i) The resolution and quality of the data were maximized by collection at a synchrotron source and by signal averaging. The intensity of each unique reflection was measured over 10 times on average. (ii) During the refinement, each solvent peak observed in the $F_o - F_c$ and $2F_o - F_c$ Fourier electron density maps was initially assigned as a water molecule, after which all atomic positions and thermal factors were refined and phases were calculated. Only if significant residual $F_o - F_c$ density, defined by criteria described below, was observed on top of the $2F_o - F_c$ density of the refined water molecule was the scattering at that site increased by conversion to a Tl^+ ion. This "water-first" approach avoided bias of the final model toward Tl^+ ions over water molecules. (iii) All Tl^+ assignments were made during the CNS stage of the refinement, when the ratio of observables to parameters was high. (iv) Tl^+ assignments were made by fixed explicit criteria, see below. (v) At the completion of the CNS stage of the refinement, each Tl^+ assignment was individually reconfirmed by an annealed omit map. Each Tl^+ site was converted back to a 100% occupied water molecule before simulated annealing, refinement, phase recalculation, and redetermination of residual $F_o - F_c$ peak intensity. (vi) The Tl^+ positions in the refined CGCGAATTCGCG/Tl^+ structure were determined to be consistent with the isomorphous $F_{Tl} - F_K$ difference Fourier and the anomalous $F_+ - F_-$ difference Fourier. These difference Fouriers were calculated with phases from a previous CGCGAATTCGCG/K^+ structure and so are not biased by positions of Tl^+ positions in the refined CGCGAATTCGCG/Tl^+ structure. (vii) The final refined model was determined to be consistent with difference Patterson maps computed with either $(F_+ - F_-)^2$ or $(F_{Tl} - F_K)^2$.

Refinement. Thirteen partially occupied Tl^+ ions were added to the model during the refinement. The minimum criteria for conversion of a water molecule to a Tl^+ ion during the refinement was a 3.5σ peak of residual $F_o - F_c$ Fourier electron density superimposed on the $2F_o - F_c$ Fourier density of a refined water molecule (Table 2; the DNA residues are 1001−1012 for one strand and 2001−2012 for the other. Tl^+ ions are residues 2101−2113). These residual difference peaks indicate that water molecules at a subset of sites do not scatter adequately, providing a poor fit of model to data (Figure 4).

The criteria employed here for identifying Tl^+ ions is conservative relative to generally accepted practices. For example, in performing a very careful refinement and attempting to avoid bias during refinement, Goodsell et al. used a 3.0σ criterion for identifying species with 10 electrons (100% occupied water molecules) in $F_o - F_c$ Fourier electron density maps (*31*). Our criterion for addition of a Tl^+ ion (3.5σ) is more stringent for species ranging from 8 electrons (10% occupied Tl^+ ions) to 28 electrons (35% occupied Tl^+ ions). The most intense residual $F_o - F_c$ Fourier electron density peak observed here was 11.7σ.

As the refinement progressed, occupancies of the Tl^+ ions were adjusted by inspection of electron density maps and thermal factors, by automated occupancy refinement, and by consideration of reasonable stereochemistry (two atoms occupying the same volume were restrained to occupancies that sum to unity at most). An overall *B*-factor correction was applied, but individual *B*-factors and occupancies were not refined simultaneously. Due to the nonorthogonality of thermal factors and fractional occupancies (*32*), these parameters require additional verification. Tl^+ occupancies determined from refinement to the fully reduced synchrotron data are generally consistent with those determined from SHELX-97 refinement to the nonmerged Bivjoet data collected on the home X-ray source.

Assignment of Tl^+ ions located within the major groove (for example, residual difference density shown in Figure 4A) was performed before assignments of other sites, which were conducted when the CNS refinement was otherwise close to convergence. For example, after the major groove Tl^+ ion assignments were completed, a difference peak of 3.9σ surrounded a water molecule at the center of the A-tract minor groove (Figure 4B). This water molecule was changed to a 10%/90% Tl^+/water hybrid (Tl^+ ion 2106), effectively increasing the number of electrons at that site from 10 to 17 and eliminating the difference peak.

Isomorphous Difference Electron Density. The $F_{Tl} - F_K$ electron density Fourier map shows spherical and elongated peaks of various intensities in the solvent region (Table 2, Figure 5). The $F_{Tl} - F_K$ map does not incorporate information or assumptions about the positions or scattering of Tl^+ ions. The peaks observed in the $F_{Tl} - F_K$ map can have several possible origins. They might arise from localized Tl^+ ions that scatter X-rays in the CGCGAATTCGCG/Tl^+ crystal

Table 2: Difference Peak Intensities and Locations

Tl no.	occupancy (%)	*B*-factor (Å^2)	$F_o - F_c$ peak height[a] (σ)	$F_{Tl} - F_K$ peak height[b] (σ)/rank	$\lvert F_+\rvert - \lvert F_-\rvert$ peak height[c] (σ)/rank	D^d (Å)	D^e (Å)	location
2101	0.34	17	11.7	21/1	20/1	0.06	0.68	G-tract major groove
2102	0.29	24	7.5	11/2	17/3	0.12	0.68	G-tract major groove
2103	0.20	24	4.6	10/3	19/2	0.18	0.38	G-tract major groove
2104	0.20	28	4.2	4.1/11	5.0/8	0.20	0.36	A-tract minor groove
2105	0.18	26	4.4	4.6/9	5.3/6	0.29	0.28	A-tract/G-tract junction minor groove
2106	0.10	36	3.9	no	4.0/10	na	0.33	A-tract minor groove
2107	0.10	26	4.5/7.8[f]	4.8/6	5.3/5	0.25	0.36	G-tract major groove[g]
2108	0.10	63	>3.5[f]	3/na[h]	2/na[h]	na	na	G-tract major groove[g]
2109	0.12	22	4.5/9.8[f]	4.8/7	6.0/4	0.22	0.61	A-tract/G-tract junction minor groove
2110	0.15	25	7.4[f]	7.1/4	5.1/7	0.13	0.61	G-tract major groove
2111	0.18	25	3.8	4.7/8	3.8/12	0.17	0.15	cation$-\pi$
2112	0.15	42	5.5[f]	no	no	na	na	cation$-\pi$
2113	0.16	26	3.5/6.0[f]	6/na[h]	7/na[h]	na	na	G-tract major groove

[a] Residual $F_o - F_c$ Fourier electron density superimposed on the $2F_o - F_c$ Fourier density of the fully occupied water molecule at this position. Phases were contributed by a partially refined CGCGAATTCGCG/Tl^+ model. [b] A $F_{Tl} - F_K$ Fourier electron density map with phases contributed by the refined CGCGAATTCGCG/K^+ model absent all solvent. [c] A CGCGAATTCGCG/Tl^+ $F_+ - F_-$ anomalous Fourier electron density map with phases contributed by the refined CGCGAATTCGCG/K^+ model absent all solvent. [d] Distance between the top of the $F_{Tl} - F_K$ peak and the position of the refined Tl^+ ion. [e] Distance between the top of $F_+ - F_-$ peak and the position of refined Tl^+ ion. [f] Residual $F_o - F_c$ Fourier electron density superimposed on the $2F_o - F_c$ (sum) Fourier density of the partially occupied water molecule. [g] Located within first hydration shell of the major groove magnesium ion. [h] These sites are within density, in shoulders of other peaks, but are not resolved from adjacent peaks. no, not observed. na, not applicable.

but not in the CGCGAATTCGCG/K^+ crystal. Additional contributions might arise from differences in positions in highly localized water molecules. Differential water location does appear to contribute one peak (the fifth most intense peak in the $F_{Tl} - F_K$ map). This peak appears to arise from a highly ordered water molecule (water 3003, thermal factor 14 Å^2) that is contained in the CGCGAATTCGCG/Tl^+ structure but not in the CGCGAATTCGCG/K^+ structure. The $F_+ - F_-$ map (below) suggests that a Tl^+ ion does not reside at this site. This water molecule is located in the minor groove of the A-tract, 2.6 Å from the O2 of cytosine 2009 and 3.0 Å from the O4′ of guanine 2010. The 10th most intense peak in the $F_{Tl} - F_K$ map could not be assigned, although it is 1.6 Å from one Tl^+ site and 2.8 Å away from another.

Anomalous Difference Electron Density Fourier. The features of the $F_+ - F_-$ map are very similar to those of the $F_{Tl} - F_K$ map (Table 2, Figure 5). Peaks in the $F_+ - F_-$ map arise from differences between Bijvoet pairs due to anomalous scattering by heavy atoms. The $F_+ - F_-$ map does not incorporate assumptions about the positions of Tl^+ ions and would not be affected by positions of water molecules. Several peaks observed in the $F_+ - F_-$ map do not fall on Tl^+ sites in the refined structure. The peak ranked 9th in intensity (4.0σ) is 1.9 Å from the C4 atom of Thy 1008. Peak 11 (3.8σ) is 1.4 Å from P of Gua 1004. Peak 13 (3.7σ) is adjacent to the N7 and O6 of Gua 1012 at the terminus of the G-tract major groove. It is suggestive that the location of this site adjacent to a guanine base is analogous to other G-tract Tl^+ positions.

Summary. Verification of the Tl^+ positions in the refined model was performed with $F_{Tl} - F_K$ and $F_+ - F_-$ Fourier maps (Table 2, Figure 5). The refinement did not incorporate information from either the $F_+ - F_-$ or $F_{Tl} - F_K$ maps. The refined structure and the $F_+ - F_-$ and $F_{Tl} - F_K$ maps are internally consistent. Eleven of the 13 Tl^+ sites determined by the refinement are confirmed by both the $F_+ - F_-$ and $F_{Tl} - F_K$ maps. One site, Tl^+ 2106 (in the minor groove at the ApT step), is observed in the $F_+ - F_-$ map but not the $F_{Tl} - F_K$ map. One site (Tl^+ 2112) is not observed in either the $F_+ - F_-$ map or the $F_{Tl} - F_K$ map and so must be considered a highly tentative assignment.

Some of the ion positions identified here, such as those within the A-tract minor groove, are located nearly on top of previously identified "water molecules" in other X-ray derived models of CGCGAATTCGCG. These ions appear to occupy "hybrid solvent sites". One of these positions (Tl^+ 2105) is nearly superimposed on a water molecule (W47) in the sodium form of CGCGAATTCGCG, although no solvent molecule was identified at this site in the K^+ form of CGCGAATTCGCG. The adjacent Tl^+ position (2109), which does not appear to occupy a hybrid solvent site, engages in the lone amino group−cation interaction in the refined model.

DISCUSSION

The effects of sequence on DNA conformation have been explained by two limiting models, as recently summarized (*33*−*35*). In one class of models, called base-clash models (using the nomenclature of McConnell and Beveridge), the sequence dependence of conformation is intrinsic, arising from short-range nonelectrostatic forces between bases. Sequence-specific conformation arises from the energetics of base−base stacking (*3*) and propeller twisting (*36*). In a second class of models, called electrostatic models, the sequence dependence of conformation is extrinsic, arising from interactions between cations, base and backbone functional groups, and solvent. In electrostatic models, functional groups of DNA bases and backbone influence the positions of cations and solvent. Cation positions influence DNA conformation through electrostatic interactions, causing sequence-dependent variation in groove width and axial bending (*24*, *33*, *37*−*43*). It is now generally accepted that the divalent cations observable by X-ray diffraction impact B-DNA conformation in a sequence-specific manner (*20*, *40*, *42*, *44*, *45*). But the issue is not fully resolved for monovalent cations.

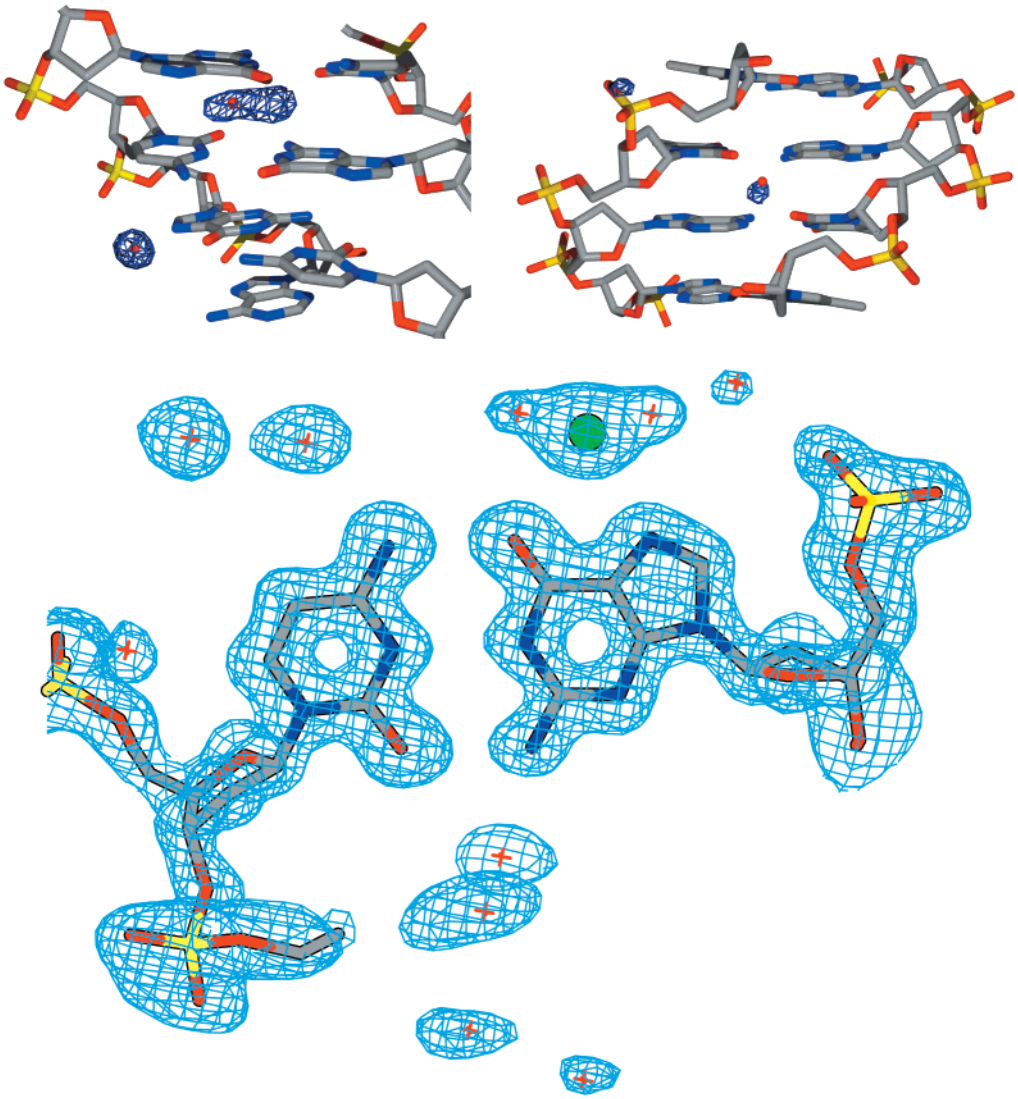

FIGURE 4: (A, top left) View into the major groove of [d(CGCGAATTCGCG)]$_2$ showing residual $F_o - F_c$ Fourier electron density peaks (blue net) superimposed onto two water molecules (red spheres). The positions and thermal factors of these water molecules were refined, and their scattering contribution was included in the phases when computing this map. These residual electron density peaks were the most intense in the $F_o - F_c$ Fourier map (contoured at 6σ here). Other solvent molecules are omitted for clarity. These two water molecules were changed to 30% Tl^+ ions in subsequent models. Both peaks are located in the major groove and are associated with guanine residues. (B, top right) View into the minor groove of d[CGCGAATTCGCG]$_2$, showing residual $F_o - F_c$ Fourier electron density superimposed onto a water molecule modeled at the primary hydration layer in the minor groove of the 5′ ApT 3′ step. The position and thermal factor of this water molecule were refined, and its scattering contribution was included in the phases when computing this map (contoured at 3.5σ here). This water molecule was converted to a 10% Tl^+ ion/90% water molecule in subsequent models. (C, bottom) Final $2F_o - F_c$ Fourier electron density (light blue net, 1.7σ) surrounding bases G(16) and C(9) and Tl^+ ion 2101. Water molecules are indicated by red crosses. A Tl^+ ion is indicated by a green sphere. The figures were created with Swiss-PDB Viewer (*63*) and rendered with Povray 3.1.

We have used the strong scattering and anomalous scattering of X-rays by Tl^+, a K^+ mimic, to determine that monovalent cations can localize in sequence-specific sites adjacent to B-DNA bases. Anomalous scattering has recently been used to identify monovalent cation positions adjacent to A-DNA (*46*). Here 13 Tl^+ ion positions are contained in the final refined model of duplex CGCGAATTCGCG (Figure 1, Table 3). The Tl^+ positions are located predominantly within the grooves and in general are not in direct proximity to phosphate groups. The majority of the observed Tl^+ ions interact with a single duplex and so are not engaged in lattice interactions or crystal packing. None of the observed Tl^+ sites surrounding CGCGAATTCGCG are fully occupied by Tl^+ ions. The most highly occupied sites, located within the G-tract major groove, have estimated occupancies ranging from 20% to 35%. The occupancies of the minor groove sites are estimated to be around 10%.

Cations in the G-Tract Major Groove. The greatest concentration of localized Tl^+ ions, in terms of both the number of positions and apparent occupancies, is within the major grooves of the G-tracts. All six O6 atoms and five out of six N7 atoms of nonterminal guanines are within 3.4 Å of a Tl^+ ion position. Each guanine, except those on the termini, is adjacent to at least one Tl^+ position. The Tl^+ positions adjacent to guanines lie predominantly within the planes of the guanine bases. One Tl^+ position, falling between guanine planes, is chelated by two cross-strand O6 atoms at a GpC step.

Selective partitioning of Tl^+ into the major groove of the G-tracts appears to be driven by (i) the appropriate geometry

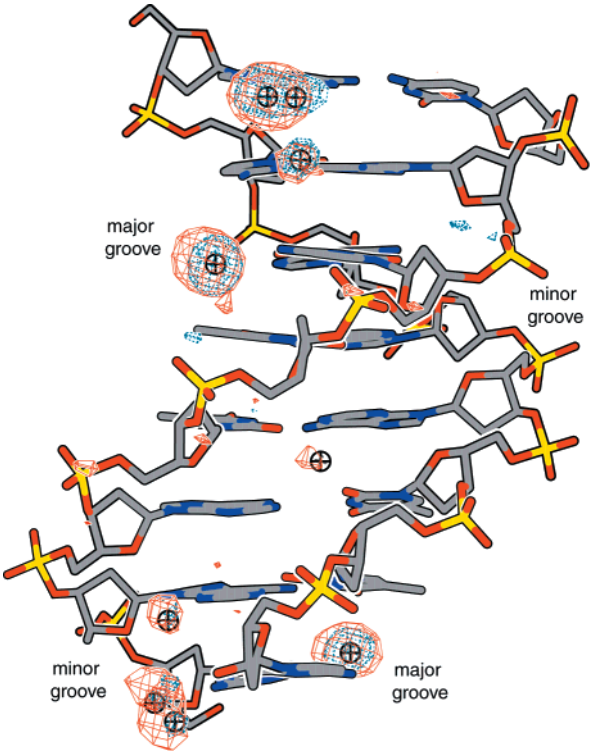

FIGURE 5: Three types of experiments suggest that Tl^+ ions localize in the grooves of duplex CGCGAATTCGCG. Tl^+ positions determined by refinement are indicated by crossed circles. Tl^+ positions determined by isomorphous $F_{Tl} - F_K$ Fourier are indicated by the dashed blue net. Tl^+ positions determined by anomalous $F_+ - F_-$ Fourier are indicated by the solid red net. The DNA is represented in stick format. Carbon atoms are in gray, nitrogen atoms are in blue, oxygen atoms are in red, and phosphorus atoms are in yellow. Tl^+ positions that are not in van der Waals contact with the DNA atoms shown here are not included. Eight of 12 base pairs are shown.

of basic functional groups for bidentate chelation, as by the O6 and N7 of a single guanine or the two cross-strand O6 atoms of a GpC step, (ii) relatively facile partial dehydration of Tl^+, allowing direct metal coordination by guanine functional groups, (iii) the high electronegative potential of the floor of the G-tract major groove, as originally noted by Lavery and Pullman (*47*), and (iv) a repulsion of cations from the A-tract major groove by electropositive amino groups (*42*). The importance of these factors in localizing the monovalent cation within the G-tract major groove is consistent with the results of MD simulations (*38*, *43*, *48*, *49*).

Partially dehydrated monovalent cations, hydrated divalent cations, and polyamines compete for a common binding region on the floor of the G-tract major groove. A $Mg(H_2O)_6^{2+}$ ion is contained in one of the G-tract major grooves of all high-resolution X-ray structures of CGCGAATTCGCG, including sodium (*24*, *50*), potassium (*20*), rubidium (*16*), cesium (*51*), monovalent-minus (*36*), and thallium (here) forms. That same region, within the G-tract major groove, is occupied by both $Mg(H_2O)_6^{2+}$ and Tl^+ ions in the structure described here. The appearance of superimposition is caused by ensemble averaging; the occupancies of the major groove $Mg(H_2O)_6^{2+}$ and the overlayed Tl^+ ions sum to 100%. Similar monovalent/divalent competition for the G-tract major groove is evident for Rb^+ and $Mg(H_2O)_6^{2+}$ (S. B. Howerton and L. D. Williams, unpublished). A partial

Table 3: Contacts of Tl^+ Ions with Nitrogen and Oxygen Atoms of CGCGAATTCGCG

Tl no.	DNA base	atom	distance (Å)	region
2101	G(2004)	O6	2.95	G-tract
	G(2004)	N7	2.75	major
2102	G(2002)	O6	2.63	groove
	G(2002)	N7	2.75	
	G(1010)	O6	3.06	
2103	G(1004)	O6	2.93	
	G(1004)	N7	2.27	
2107	G(1002)	O6	2.43	
	G(2010)	O6	2.86	
2108	G(1002)	O6	2.47	
	G(1002)	N7	2.93	
2110	G(1010)	O6	2.66	
	G(1010)	N7	2.57	
2113	G(2002)	O6	2.78	
	G(1010)	O6	3.06	
2105	A(1005)	O4′	3.01	A-tract/
	G(1004)	N3	2.65	G-tract
	G(1012)[a]	O3′	3.01	junction,
2109	C(2009)	O2	3.18	minor
	G(1004)	N2	3.18	groove
	G(2010)	O4′	2.77	
	G(2010)	N3	3.09	
2104	A(1005)	N3	2.74	A-tract,
	A(1006)	O4′	3.11	minor
2106	T(1007)	O2	2.67	groove
	T(2007)	O2	2.54	
	T(2008)	O4′	3.20	
2111	C(1001)	O4′	2.84	cation−π
	C(2003)[a]	O2P	2.77	
	G(2002)[a]	O3′	3.36	
2112	C(1003)[a]	O2P	2.42	

[a] Indicates lattice interaction with an adjacent duplex.

spermine molecule is observed in the one G-tract major groove in some high-resolution structures of CGCGAATTCGCG (*20*). An ordered spermine molecule is not observable in the G-tract major groove of CGCGAATTCGCG/Tl^+ maps, where it appears that the spermine molecule has been displaced by Tl^+ ions. In sum, a nonspecific ion-binding region on the floor of the G-tract major groove has been observed to be occupied by $Mg(H_2O)_6^{2+}$, monovalent cations, and spermine. Similar competition of monovalent and divalent cations for a common site was previously reported by Doudna and co-workers, who observed a Tl^+ ion in a site that also binds $Mg(H_2O)_6^{2+}$ in the *Tetrahymena* ribozyme P4−P6 domain (*7*). The $Mg(H_2O)_6^{2+}$ in the major groove of CGCGAATTCGCG is thought to be the origin of observed deformations such as axial bending. But the observation here, that the $Mg(H_2O)_6^{2+}$ is partially occupied, in competition with Tl^+, suggests that at least some of the conformational deformation previously attributed to $Mg(H_2O)_6^{2+}$ can arise from monovalent cations.

Cations in the Minor Groove. Four Tl^+ ion positions are located within the minor groove (Table 3). These positions are coordinated by the O2 atoms of pyrimidines, the N3 atoms of purines, and the O4′ atoms of deoxyriboses. Two adjacent minor groove Tl^+ positions are located near one of the G-tract/A-tract junctions, essentially one step beyond one end of the "spine of hydration" (*52*). Two minor groove Tl^+ positions (2104 and 2106) are nearly superimposed on previously assigned spine sites near the floor of the A-tract minor groove. One of these Tl^+ positions (2106) is located at the ApT step of CGCGAATTCGCG. Observation of Tl^+ sites within the A-tract minor groove is consistent with partial

occupancy at the minor groove ApT step of CGCGAATTCGCG by Na^+ (*24*), K^+ (*42*), Rb^+ (*16*), and Cs^+ (*51*).

Cations at the Duplex Termini. Both terminal cytosines in the CGCGCAATTCGC/Tl^+ model are stacked on Tl^+ sites. These Tl^+ positions, which engage in lattice interactions, are consistent with a previously noted tendency of cations to interact favorably with cytosine bases via cation$-\pi$ interactions (*53*). One of these cation$-\pi$ sites could not be confirmed by isomorphous difference or anomalous scattering and therefore should be considered tentative.

Relationships between Cation Positions and DNA Conformation. The cation dependencies DNA bending (*54*−*56*) and minor groove width (*20*, *43*, *44*, *57*) suggest mechanistic rolls for cations in B-DNA deformation and conformational heterogeneity. The observed directions of bending, A-tracts toward the minor groove (*58*) and G-tracts toward the major groove (*59*, *60*), are consistent with electrostatic models of DNA deformation. In solution DNA bends toward regions where localized cations are observed in crystals. One could more fully test electrostatic models of DNA deformation by determining the positions of most or all of the cations surrounding the DNA. The structure of CGCGAATTCGCG/Tl^+ descibed here does not account for all surrounding cations and does not approach charge neutrality. Each of the Tl^+ positions identified is less than 50% occupied. Additional cation positions may be revealed with further experimentation. Alternatively, a subset of cations might be sufficiently delocalized so as to be unobservable by crystallographic methods. Delocalized cations are not necessarily confined to remote disordered regions and are not necessarily unimportant in local deformation. Recent MD simulations (*43*, *57*) suggest that the minor groove narrows when monovalent cations are located either on the floor of the minor groove or higher up, in the lip of the groove (between opposing phosphate groups). Local shielding narrows the groove when cations are contained in a volume roughly between opposing phosphate groups. This minor groove lip region contains relatively disordered solvent. Therefore, monovalent cations in that region may be important to conformation but might not be readily detected by X-ray diffraction.

Conclusion. In sum, the CGCGAATTCGCG/Tl^+ data support the hypothesis that partially dehydrated monovalent cations are coordinated by oxygen and nitrogen positions of DNA bases. The primary region of Tl^+ localization is within the G-tract major groove. Tl^+ is also observed to penetrate the spine of hydration and to localize near the floor of the A-tract minor groove. NMR experiments support the proposal that monovalent cations distribute preferentially in the A-tract minor groove (*41*, *61*). Most recently, Stellwagen and coworkers measured free solution mobilities of DNA oligomers with and without A-tracts (*62*) and concluded that preferential counterion interaction occurs in A-tract DNA, especially those containing A_nT_n tracts. It remains to be experimentally confirmed whether ions preferentially interact with G-tract major grooves in solution, as suggested by X-ray diffraction (here) and MD simulations (*48*).

The number of monovalent cation positions evident in the Tl^+ form of CGCGAATTCGCG substantially exceeds previous observations using less strongly scattering Rb^+ (one ∼50% occupied position) (*16*) or Cs^+ (four ∼20% occupied positions) (*51*). Differences in monovalent cation distributions in various models are likely to arise both from real differences and from experimental uncertainty. The specific physical properties of the various cations would contribute to variations in their localization. In addition, positions of competing ions would influence monovalent cation positions. It has been shown that variability in positions of divalent cations and polyamines correlates with differences in crystallization conditions (*20*). Some of the apparent differences in cation distributions probably arise from experimental signal/noise ($Tl^+ > Cs^+ > Rb^+ > K^+ > Na^+$), data quality, and interpretation. With strong scattering and strong anomalous signal, it is not surprising that more sites are observed with Tl^+ than with other ions.

ACKNOWLEDGMENT

We thank Pascal Auffinger, Nicholas Hud, David Beveridge, Kevin McConnell, Angus Wilkinson, Dieter Schneider, and Jonathon Chaires for helpful discussions. This research was carried out in part at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the U.S. Department of Energy Division of Materials Sciences and Division of Chemical Sciences under Contract DE-AC02-98CH10886. Beamline X26-C is supported in part by the Georgia Research Alliance.

SUPPORTING INFORMATION AVAILABLE

One table that gives *R*-factor, *R*-free, and goodness of fit for the refinement progression. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Quigley, G. J., Teeter, M. M., and Rich, A. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 64−68.
2. Cate, J. H., Hanna, R. L., and Doudna, J. A. (1997) *Nat. Struct. Biol. 4*, 553−558.
3. Grzeskowiak, K., Yanagi, K., Prive, G. G., and Dickerson, R. E. (1991) *J. Biol. Chem. 266*, 8861−8883.
4. Gessner, R. V., Quigley, G. J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., and Rich, A. (1985) *Biochemistry 24*, 237−240.
5. Soler-Lopez, M., Malinina, L., and Subirana, J. A. (2000) *J. Biol. Chem. 275*, 23034−23044.
6. Wang, A. H.-J., Ughetto, G., Quigley, G. J., and Rich, A. (1987) *Biochemistry 26*, 1162−1163.
7. Basu, S., Rambo, R. P., Strauss-Soukup, J., Cate, J. H., Ferre-D'Amare, A. R., Strobel, S. A., and Doudna, J. A. (1998) *Nat. Struct. Biol. 5*, 986−992.
8. Gursky, O., Li, Y., Badger, J., and Caspar, D. L. (1992) *Biophys. J. 61*, 604−611.
9. Badger, J., Li, Y., and Caspar, D. L. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1224−1228.
10. Brown, I. D. (1988) *Acta Crystallogr. B44*, 545−553.
11. Wulfsberg, G. (1991) *Principles of Descriptive Inorganic Chemistry*, University Science Books, Sausalito, CA.
12. Basu, S., Szewczak, A. A., Cocco, M., and Strobel, S. A. (2000) *J. Am. Chem. Soc. 122*, 3240−3241.
13. Pedersen, P. A., Nielsen, J. M., Rasmussen, J. H., and Jorgensen, P. L. (1998) *Biochemistry 37*, 17818−17827.
14. Villeret, V., Huang, S., Fromm, H. J., and Lipscomb, W. N. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 8916−8920.
15. Loria, J. P., and Nowak, T. (1998) *Biochemistry 37*, 6967−6974.
16. Tereshko, V., Minasov, G., and Egli, M. (1999) *J. Am. Chem. Soc. 121*, 3590−3595.
17. Powell, H. R. (1999) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 55*, 1690−1695.
18. (1994) *Acta Crystallogr., Sect. D: Biol. Crystallogr. D50*, 760−763.

19. Berman, H. M., Zardecki, C., and Westbrook, J. (1998) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 1095−1104.
20. Sines, C. C., McFail-Isom, L., Howerton, S. B., VanDerveer, D., and Williams, L. D. (2000) *J. Am. Chem. Soc. 122*, 11048−11056.
21. McRee, D. E. (1999) *Practical Protein Crystallography*, 2nd ed., Academic Press, New York.
22. Crick, F. H. C., and Magdoff, B. S. (1956) *Acta Crystallogr. 9*, 901−908.
23. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.
24. Shui, X., McFail-Isom, L., Hu, G. G., and Williams, L. D. (1998) *Biochemistry 37*, 8341−8355.
25. Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W. K., and Berman, H. M. (1996) *J. Am. Chem. Soc. 118*, 519−529.
26. Clowney, L., Jain, S. C., Srinivasan, A. R., Westbrook, J., Olson, W. K., and Berman, H. M. (1996) *J. Am. Chem. Soc. 118*, 509−518.
27. Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T., and Berman, H. M. (1996) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 52*, 57−64.
28. Sheldrick, G. M. (1997) *SHELX-97*, Gottingen University, Germany.
29. Farrugia, L. J. (1997) *J. Appl. Crystallogr. 30*, 565.
30. Johnson, C. K. (1976) *Ortep Ii*, Oak Ridge National Labratory, Report Rnl-5138, Oak Ridge, TN.
31. Goodsell, D. S., Kopka, M. L., and Dickerson, R. E. (1995) *Biochemistry 34*, 4983−4993.
32. Ladd, M. F. C., and Palmer, R. A. (1985) *Structure Determination by X-ray Diffraction*, 2nd ed., Plenum Press, New York.
33. McConnell, K. J., and Beveridge, D. L. (2000) *J. Mol. Biol. 304*, 803−820.
34. McFail-Isom, L., Sines, C., and Williams, L. D. (1999) *Current Opin. Struct. Biol. 9*, 298−304.
35. Hud, N. V., and Polak, M. (2001) *Curr. Opin. Struct. Biol.* (in press).
36. Chiu, T. K., Kaczor-Grzeskowiak, M., and Dickerson, R. E. (1999) *J. Mol. Biol. 292*, 589−608.
37. Rouzina, I., and Bloomfield, V. A. (1998) *Biophys. J. 74*, 3152−3164.
38. Young, M. A., Jayaram, B., and Beveridge, D. L. (1997) *J. Am. Chem. Soc. 119*, 59−69.
39. Strauss, J. K., and Maher, L. J. (1994) *Science 266*, 1829−1834.
40. Hud, N. V., and Feigon, J. (1997) *J. Am. Chem. Soc. 119*, 5756−5757.
41. Hud, N. V., Sklenar, V., and Feigon, J. (1999) *J. Mol. Biol. 286*, 651−660.
42. Shui, X., Sines, C., McFail-Isom, L., VanDerveer, D., and Williams, L. D. (1998) *Biochemistry 37*, 16877−16887.
43. Hamelberg, D., McFail-Isom, L., Williams, L. D., and Wilson, W. D. (2000) *J. Am. Chem. Soc. 122*, 10513−10520.
44. Minasov, G., Tereshko, V., and Egli, M. (1999) *J. Mol. Biol. 291*, 83−99.
45. Chiu, T. K., and Dickerson, R. E. (2000) *J. Mol. Biol. 301*, 915−945.
46. Tereshko, V., Wilds, C. J., Minasov, G., Prakash, T. P., Maier, M. A., Howard, A., Wawrzak, Z., Manoharan, M., and Egli, M. (2001) *Nucleic Acids Res. 29*, 1208−1215.
47. Lavery, R., and Pullman, B. (1985) *J. Biomol. Struct. Dynm. 2*, 1021−1032.
48. Auffinger, P., and Westhof, E. (2000) *J. Mol. Biol. 300*, 1113−1131.
49. Feig, M., and Pettitt, B. M. (1999) *Biophys. J. 77*, 1769−1781.
50. Tereshko, V., Minasov, G., and Egli, M. (1999) *J. Am. Chem. Soc. 121*, 470−471.
51. Woods, K., McFail-Isom, L., Sines, C. C., Howerton, S. B., Stephens, R. K., and Williams, L. D. (2000) *J. Am. Chem. Soc. 122*, 1546−1547.
52. Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535−556.
53. McFail-Isom, L., Shui, X., and Williams, L. D. (1998) *Biochemistry 37*, 17105−17110.
54. Diekmann, S., and Wang, J. C. (1985) *J. Mol. Biol. 186*, 1−11.
55. Laundon, C. H., and Griffith, J. D. (1987) *Biochemistry 26*, 3759−3762.
56. Brukner, I., Susic, S., Dlakic, M., Savic, A., and Pongor, S. (1994) *J. Mol. Biol. 236*, 26−32.
57. Hamelberg, D., Williams, L. D., and Wilson, W. D. (2001) *J. Am. Chem. Soc.* (in press).
58. Zinkel, S. S., and Crothers, D. M. (1987) *Nature 328*, 178−181.
59. Dlakic, M., and Harrington, R. E. (1995) *J. Biol. Chem. 270*, 29945−29952.
60. Milton, D. L., Casper, M. L., Wills, N. M., and Gesteland, R. F. (1990) *Nucleic Acids Res. 18*, 817−820.
61. Denisov, V. P., and Halle, B. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 629−633.
62. Stellwagen, N. C., Magnusdottir, S., Gelfi, C., and Righetti, P. G. (2001) *J. Mol. Biol.*, 1025−1033.
63. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis 18*, 2714−2723.

BI010391+