Monovalent Cations Sequester within the A-Tract Minor Groove of [d(CGCGAATTCGCG)]$_2$[†,§]

Kristen Kruger Woods, Lori McFail-Isom, Chad C. Sines, Shelley B. Howerton, Rasheeda K. Stephens, and Loren Dean Williams*

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

Received June 11, 1999

It has been proposed that sequence-specific and groove-specific interactions of cations cause DNA deformations. Monovalent cations appear to partition into the minor groove of A-tracts, as indicated by NMR,[1] molecular dynamics simulations,[2−4] and X-ray diffraction.[5−9] Divalent cations, with a greater tendency to orient and polarize water molecules[10] and to resist dehydration,[11] appear to partition into the major groove of G-tracts.[6,12] Electrostatic deformation[13−15] by mobile cations[16] provides a unified model that explains diverse phenomena such as A-tract bending,[17] G-tract bending,[18,19] groove-width variation,[20] and indirect readout.[21] This electrostatic model is consistent with observations that DNA bending in solution is cation dependent[18,22−25] and that A-tracts are bent in solution[26] but not in crystals,[27] where intermolecular interactions would wash out intramolecular effects.

An alternative nonelectrostatic view is supported by Dickerson and colleagues, who stress the importance of sequence-specific, direct base−base interactions. A representative expression of this mechanical, "heterocycle-centric", model was articulated recently; "The minor groove in A-tract B-DNA is narrow because of intrinsic properties of base-sequence, most notably the greater ease of propeller twisting of A−T base pairs...."[28] In this model, short-range, through-space interactions between bases modulate propeller twist, groove width, and axial trajectory.

A series of implicit assumptions underlie heterocycle-centric models. It is assumed that an invisibility of cations in electron density maps indicates delocalization and lack of specific energetic and structural roles. This assumption is incorrect. Ordered, structural cations are located in "hydration" regions adjacent to DNA. Their invisibility is in part an artifact of low resolution. For example with 1.4 Å resolution data we observed a magnesium ion bound in the major groove of one of the G-tracts of duplex CGCGAATTCGCG.[5] That magnesium ion has been conserved in all subsequent high-resolution CGCGAATTCGCG structures[6,7,28] but was invisible in dozens of previous low-resolution iterations. The uncloaking of the major groove magnesium ion demonstrates that unobserved cations are not necessarily dissolved, delocalized, and remote from the DNA, as has been proposed.[28] The invisibility of this divalent cation in previous electron density maps did not indicate a lack of structural significance. The major groove magnesium ion, observed or not, exerts strong electrostatic forces on the DNA, causing the "dodecamer bend". The location and interactions of this magnesium ion explain the dependence of G-tract bending on divalent cations.[18,24,25]

Evaluating the utility of electrostatic and heterocycle-centric models requires determining positions of the cations that surround nucleic acid structures in crystals. But a difficult analytical challenge is presented by the equivalence in the number of electrons in a sodium cation and a water molecule, by unpredictable coordination geometry at macromolecular surfaces, and by water/cation mixed occupancy. A 20% potassium−80% water hybrid will effectively display only 1.6 more electrons than a water molecule. The analytical challenges can be at least partially overcome with substitution by heavy monovalent cations such as cesium. Here we report the X-ray structure of the cesium form of CGCGAATTCGCG grown from a solution containing cesium, magnesium, and spermine. Detailed crystallization conditions and complete data collection and refinement statistics are given in Table 1S. The final model (resolution = 1.8 Å; *R*factor = 20.93; *R*free = 26.68) contains one fully hydrated magnesium ion and 141 additional water molecules. A partial spermine molecule was observed in the electron density maps but was not included in the model.

The electron density computed from the cesium form CGCGAATTCGCG data indicates occupancy of monovalent cations in the A-tract minor groove (Figure 1). Difference ($F_o - F_c$) Fourier electron density maps calculated with phases from the completely refined cesium form model (DNA, magnesium, water molecules, no cesium ions) reveal peaks at adjacent sites near the floor of the minor groove, within the primary layer of the fused hexagon motif.[6] This type of map gives positive peaks where the model contains inadequate scattering power. The maps indicate that water molecules located within the primary layer do not contain sufficient electron density to accurately fit the data. For models refined without monovalent cations, conversion of monovalent cations in the crystallization solution from sodium to cesium increases the difference density distributed throughout the primary layer.

We have performed a series of computations to further evaluate which hydration regions contain monovalent cations. We have

* Address correspondence to this author.

† This work was supported by the National Science Foundation (Grant MCB-9056300) and the American Cancer Society (Grant RPG-95-116-03-GMC).

§ Atomic coordinates and structure factors have been deposited in the PDB (entry code 1DOU and NDB entry code BD0029).

(1) Hud, N. V.; Sklenar, V.; Feigon, J. *J. Mol. Biol.* **1999**, *286*, 651−60.
(2) Young, M. A.; Beveridge, D. L. *J. Mol. Biol.* **1998**, *281*, 675−687.
(3) Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Am. Chem. Soc.* **1997**, *119*, 59−69.
(4) Feig, M.; Pettitt, B. M. *Biophys. J.* **1999**, *77*, 1769−81.
(5) Shui, X.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. *Biochemistry* **1998**, *37*, 8341−8355.
(6) Shui, X.; Sines, C.; McFail-Isom, L.; VanDerveer, D.; Williams, L. D. *Biochemistry* **1998**, *37*, 16877−16887.
(7) Tereshko, V.; Minasov, G.; Egli, M. *J. Am. Chem. Soc.* **1999**, *121*, 3590−3595.
(8) Bartenev, V. N.; Golovamov Eu, I.; Kapitonova, K. A.; Mokulskii, M. A.; Volkova, L. I.; Skuratovskii, I. Y. *J. Mol. Biol.* **1983**, *169*, 217−234.
(9) Rosenberg, J. M.; Seeman, N. C.; Kim, J. J. P.; Suddath, F. L.; Nicholas, H. B.; Rich, A. *Nature* **1973**, *243*, 150−154.
(10) Katz, A. K.; Glusker, J. P.; Markham, G. D.; Bock, C. W. *J. Phys. Chem. B* **1998**, *102*, 6342−6350.
(11) Collins, K. D. *Biophys. J.* **1997**, *72*, 65−76.
(12) Buckin, V. A.; Kankiya, B. I.; Rentzeperis, D.; Marky, L. A. *J. Am. Chem. Soc.* **1994**, *116*, 9423−9429.
(13) Manning, G. S.; Ebralidse, K. K.; Mirzabekov, A. D.; Rich, A. *J. Biomol. Struct. Dyn.* **1989**, *6*, 877−889.
(14) Mirzabekov, A. D.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 1118−1121.
(15) Strauss, J. K.; Maher, L. J. *Science* **1994**, *266*, 1829−1834.
(16) Rouzina, I.; Bloomfield, V. A. *Biophys. J.* **1998**, *74*, 3152−3164.
(17) Zinkel, S. S.; Crothers, D. M. *Nature* **1987**, *328*, 178−181.
(18) Brukner, I.; Susic, S.; Dlakic, M.; Savic, A.; Pongor, S. *J. Mol. Biol.* **1994**, *236*, 26−32.
(19) Milton, D. L.; Casper, M. L.; Wills, N. M.; Gesteland, R. F. *Nucleic Acids Res.* **1990**, *18*, 817−820.
(20) Burkhoff, A. M.; Tullius, T. D. *Cell* **1987**, *48*, 935−943.
(21) Koudelka, G. B.; Harrison, S. C.; Ptashne, M. *Nature* **1987**, *326*, 886−888.
(22) Diekmann, S.; Wang, J. C. *J. Mol. Biol.* **1985**, *186*, 1−11.
(23) Laundon, C. H.; Griffith, J. D. *Biochemistry* **1987**, *26*, 3759−3762.
(24) Dlakic, M.; Harrington, R. E. *J. Biol. Chem.* **1995**, *270*, 29945−52.
(25) Strauss, J. K.; Roberts, C.; Nelson, M. G.; Switzer, C.; Maher, L. J. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9515−9520.
(26) Koo, H.-S.; Drak, J.; Rice, J. A.; Crothers, D. M. *Biochemistry* **1990**, *29*, 4227−4234.
(27) Dlakic, M.; Park, K.; Griffith, J. D.; Harvey, S. C.; Harrington, R. E. *J. Biol. Chem.* **1996**, *271*, 17911−17919.
(28) Chiu, T. K.; Kaczor-Grzeskowiak, M.; Dickerson, R. E. *J. Mol. Biol.* **1999**, *292*, 589−608.

10.1021/ja9919579 CCC: $19.00
Published on Web 02/04/2000

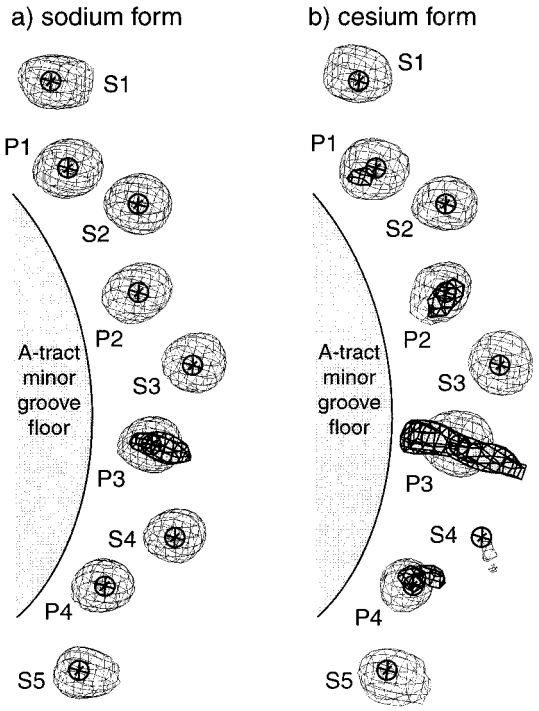

Figure 1. Electron density surrounding the primary (P) and secondary (S) solvent layers of the A-tract minor groove. Sum of the Fourier electron density ($2F_o - F_c$) contoured at 1.0σ is indicated by thin lines. The difference Fourier electron density ($F_o - F_c$) contoured at 2.5σ is indicated by thick lines. The phases were obtained from fully refined models that contain DNA, magnesium, and water, but lack monovalent cations. (a) The 1.4 Å sodium form of d[(CGCGAATTCGCG)]$_2$ (NDB entry BDL084). (b) The 1.8 Å cesium form of d[(CGCGAATTCGCG)]$_2$ (NDB entry BD0029).

used ΔRfree to indicate differences in fit of data to model upon variation of cesium ion occupancies and locations. As shown in Figure 2 the results confirm that cesium ions are sequestered within the A-tract minor groove of CGCGAATTCGCG. Here the hydration region has been subdivided into primary and secondary layers of the minor groove A-tract, major groove, phosphate, and bulk regions. The *R*free is a cross-validated estimate of goodness of fit of model to data that uses a reserved subset of the diffraction data, insulating it from the refinement.[29] The ΔRfree is the change in *R*free when the identity of a solvent site is independently incremented on the path from fully occupied water molecule to fully occupied cation. The ΔRfree information on specific hydration sites is contained in Table 2S.

Conclusion. The unambiguous confirmation here that monovalent cations and water molecules share sites within the minor groove of A-tracts supports previous indications of water−cation hybrids in sodium,[5] potassium,[6] and rubidium[7] forms of CGCGAATTCGCG. However the occupancies are not fixed; divalent cations and polyamines compete with monovalent cations in solution[30,31] while interacting preferentially at different sites.[1,5,32] Analogous competition and differential site preferences can be

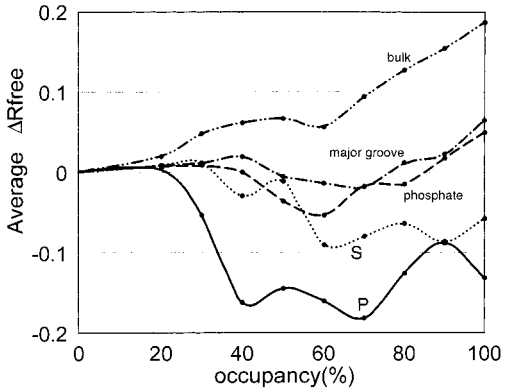

Figure 2. Variation in fit of the model to data upon addition of cesium to the model. Each point represents the average change in *R*free when the solvent sites in that category are independently switched from the water molecule to cesium with the specified occupancy. Solvent sites within the phosphate region are within 3.5 Å of phosphate oxygens and over 3.5 Å from any other DNA atom. Bulk waters are those 20 sites with lowest thermal factors located over 3.5 Å from any DNA atom. The primary layer (P sites) and secondary layers (S sites) are located within the minor groove as shown in Figure 1. The major groove is within 3.5 Å from any purine 6 or 7 position, or any pyrimidine O4 position. ΔRfree's have been obtained for each solvent site, at fractional cesium occupancies ranging from 0.2 to 1.0, by the following procedure: (1) refine a model to convergence with all solvent sites fully occupied by water molecules, to give model$_w$(the *R*free data subset is not refined against model$_w$), (2) calculate *R*free$_w$, (3) switch the identity of solvent site i from water molecule to 20% occupied cesium ion ($f_{i,j} = 0.2$), (4) re-refine thermal factors, to give model$_{i,j}$, (5) calculate *R*free$_{i,j}$, (6) calculate $\Delta R\text{free}_{i,j} = (R\text{free}_{i,j}) - (R\text{free}_w)$, (7) increment $f_{i,j} = 0.2$, to $f_{i,j+1,2,3} = 0.3, 0.4, 0.5, \ldots$ and return to step 4 (at $f_{i,9}$ go to step 8), (8) return to step 3 and switch to solvent site $i + 1$, then $i + 2$, ..., restarting with the original model after each cycle to avoid phase bias, to accumulate $\Delta R\text{free}_{mn}$, where *m* indicates any solvent site and *n* indicates occupancy between 0.2 and 1.0. For cesium occupancy ≥ 20%, scattering from water was neglected.

"observed" in crystals by comparing CGCGAATTCGCG structures obtained from crystals grown from solutions of relatively low[5,6] and high[33] concentrations of magnesium. Divalent cations outside the minor groove displace monovalent cations from within the minor groove. Helical parameters and minor groove width profile are dependent on positions and populations of monovalent cations (primarily within the groove) and divalent cations (primarily outside the groove) (Chad C. Sines and Loren Williams, unpublished). Current data do not allow us to be quantitative about occupancies, effects of specific properties of various monovalent cations,[34] or specific effects of various charge distributions on groove width. In the crystal described here, we estimate cesium occupancy within hybrid solvent sites in the A-tract minor groove to be from 10 to 40% at each site.

Supporting Information Available: A description of crystallization, data collection and reduction, and refinement of the cesium form of CGCGAATTCGCG is available free of charge via the Internet at http://pubs.acs.org.

JA9919579

(29) Brunger, A. T. *Nature* **1992**, *355*, 472−475.
(30) Bleam, M. L.; Anderson, C. F.; Record, M. T., Jr. *Biochemistry* **1983**, *22*, 5418−5425.
(31) Braunlin, W. H.; Nordenskiold, L.; Drakenberg, T. *Biopolymers* **1991**, *31*, 1343−1346.
(32) Hud, N. V.; Feigon, J. *J. Am. Chem. Soc.* **1997**, *119*, 5756−5757.
(33) Tereshko, V.; Minashov, G.; Egli, M. *J. Am. Chem. Soc.* **1999**, *121*, 470−471.
(34) Collins, K. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5553−5557.