X-ray crystal structures of half the human papilloma virus E2 binding site: d(GACCGCGGTC)

James B. Finley and Ming Luo*

Center for Macromolecular Crystallography, Department of Microbiology, University of Alabama-Birmingham, Birmingham, AL 35294-0005, USA

Received June 2, 1998; Revised September 14, 1998; Accepted October 6, 1998 NDB accession nos AdJ0102 and AdJ0103

ABSTRACT

The X-ray crystal structure of the DNA decamer d(GACCGCGGTC), containing half the human papilloma virus E2 binding site, has been solved from two crystals grown at different ionic conditions (50 mM MgCl₂ and 50 mM spermine or 1.56 mM MgCl₂ and 1.56 mM spermine). Despite the variation in salt concentration, the two DNA structures are in a very similar, A-type DNA conformation, with helical axes curving towards the major groove. Although the salt concentrations do not effect the helical parameters or hydration to a large degree, there is a change in the overall helical curvature; 18° and 31° for the low and high salt structures, respectively. This curvature appears to be sequence specific and biologically relevant when compared with similar DNA structures, including the E2 binding site of a protein–DNA complex.

INTRODUCTION

Human papilloma viruses (HPVs) cause a broad spectrum of disease, the most prevalent being the ordinary wart (1,2). However, infections of genital and oral epithelia by HPVs cause epithelial lesions ranging from benign condylomata associated with types 6 and 11 to intraepithelial dysplasias associated with types 16, 18 and other closely related types.

The HPV genome has a domain termed the upstream regulatory region (URR) which contains the origin of replication, promoters, enhancers and other regions which may effect viral replication or transcription. One HPV regulatory sequence, present in multiple copies in the URR and throughout the genome, is the E2 binding site (E2-BS). The E2 protein, by binding to the E2-BS, directs viral replication and regulates transcription of the viral oncoproteins E6 and E7 (3–15).

The highly conserved E2-BS DNA sequence is ACCN₆GGT. The E2 protein forms a two-fold related homodimer and each monomer specifically recognizes the d(ACC) or d(GGT) sequence and can induce DNA bending (16,17). The specific d(ACC) sequence as well as the 6 nt spacer sequence are important for protein recognition of this site (18). The basic E2-BS of the HPV strains may be further generalized to ACCNg₆cGGT. Where ACC is the conserved specific recognition sequence, the lower case g and c are preferred sequences and N₆ tend to include three to four d(A) residues (19–21). A sequence alignment of several HPV E2-BSs is provided in Figure 1. The DNA sequence presented in this paper, d(GACCGCGGTC), is shown here with the conserved d(ACC) in bold. Note that the d(G) after the conserved d(ACC) is a preferred sequence and the initial 5′ d(G) corresponds to HPV 11, 31 and 35 (Fig. 1). The sequence after the underlining is included to complete dyad symmetry for crystallographic purposes. Thus, half the specific recognition sequence d(ACC) of the E2-BS may be investigated in the context of a real E2-BS sequence, with d(G) residues around the d(ACC).

To gain a better understanding of the structural recognition between the E2 protein and the E2-BS, we report the X-ray crystal structure of the self-complementary DNA decamer d(GACCGCGGTC) containing half the E2-BS. This structure has been solved at two different ionic conditions resulting in different helical curvatures. By comparing these structures to similar sequences, a discussion of conserved structural features is possible. The two salt conditions represent crystals grown from a relatively high or low ionic strength: 50 mM MgCl₂ and 50 mM spermine for the high salt or 1.56 mM MgCl₂ and 1.56 mM spermine for the low salt structure. The terms high and low will be used throughout this paper in reference to the structures solved in the above two salt conditions. In both ionic conditions, the DNA sequence adopts a similar overall structure in the A-DNA conformation. But more interesting, are the similar DNA helical curvatures and conserved water molecules. Also, the structure is compared with previously solved, similar DNA structures, including the E2-BS bound by the bovine E2 protein (18). Finally, a discussion of biological relevance is presented.

MATERIALS AND METHODS

Synthesis, crystallization and data collection

The DNA decamer d(GACCGCGGTC) was purchased from The Midland Certified Reagent Co. (Midland, TX).

Crystals of varying salt conditions (high and low salt crystals) were grown from 5 µl hanging drops at room temperature. Aliquots of 1 µl each of the following stock solutions were mixed to form the drop: 10 mM DNA; 200 mM cacodylic acid, pH 6.5 (free acid adjusted with NaOH); 30% (v/v) 2-methyl-2,3-pentanediol (MPD); 250 mM MgCl₂; 250 mM spermine for the high salt structure;
crystals grew to pH 6.5 (free acid adjusted with NaOH), 30% (v/v) MPD. Hexagonal drops were equilibrated over 1 ml of 200 mM cacodylic acid, enantiomorph P6\(^{5}\)\(^{2}\). showed the general reflection conditions: \(7.8\text{ mM MgCl}_2\), \(7.8\text{ mM spermine}\) for the low salt structure. The were recorded with a 0.25 crystal to detector distance was 10 cm and the diffraction images pinhole collimator were used to select the CuK\(_\alpha\) radiation (Ni filtered), 50 kV, 108 mA, 40% bias and \(\alpha\) radiation. The \(\alpha\) radiation. The \(\alpha\) radiation. The were collected for the low salt crystal, 180\(^\circ\) oscillation for 2 min; 180\(^\circ\) at each point, \(R_{\text{free}}\) and 2F\(_{\text{o}}\) – F\(_{\text{c}}\) maps were used to monitor each round of all atom minimization. Finally, a B factor refinement was performed on all atoms and an occupancy analysis completed on all water molecules. The final refinement statistics for both structures are shown in Table 1.

### Table 1. Crystallographic parameters

<table>
<thead>
<tr>
<th></th>
<th>low salt</th>
<th>high salt</th>
</tr>
</thead>
<tbody>
<tr>
<td>space group</td>
<td>P6(^{1})(^{2})</td>
<td>P6(^{2})(^{2})</td>
</tr>
<tr>
<td>unit cell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a = b)</td>
<td>38.61</td>
<td>38.35</td>
</tr>
<tr>
<td>(c)</td>
<td>80.60</td>
<td>80.21</td>
</tr>
<tr>
<td>asymmetric unit</td>
<td>single stranded DNA</td>
<td>single stranded DNA</td>
</tr>
<tr>
<td>(V_m)</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>resolution(^{1})</td>
<td>1.7</td>
<td>1.9</td>
</tr>
<tr>
<td># unique reflections</td>
<td>4447</td>
<td>3599</td>
</tr>
<tr>
<td>% complete</td>
<td>97.6</td>
<td>96.0</td>
</tr>
<tr>
<td>(R_{sym})</td>
<td>5.23</td>
<td>5.05</td>
</tr>
<tr>
<td>(R_{free})</td>
<td>21.0</td>
<td>25.5</td>
</tr>
<tr>
<td># waters</td>
<td>52</td>
<td>47</td>
</tr>
<tr>
<td>RMSD</td>
<td>0.009</td>
<td>0.010</td>
</tr>
<tr>
<td>bond length</td>
<td>1.640</td>
<td>1.430</td>
</tr>
<tr>
<td>bond angle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vs. high salt structure (DNA only)</td>
<td>0.08</td>
<td>0.00</td>
</tr>
</tbody>
</table>

1 Mathew’s number.  
2 Resolution cut-off where \(< I > > 3\sigma\).  
3 \(R_{sym} = \Sigma I - \langle I \rangle/\Sigma I\).  
4 \(R_{free} (10\% \text{ of data}) = \Sigma \{F_{\text{o}}(hkl)\} - k\{F_{\text{c}}(hkl)\}/\Sigma \{F_{\text{o}}(hkl)\} - E(hkl)\}/F_{\text{o}}(hkl).\)  
5 \(R = \Sigma \{F_{\text{o}}(hkl)\} - k\{F_{\text{c}}(hkl)\}/\Sigma \{F_{\text{o}}(hkl)\} - E(hkl)\}/F_{\text{o}}(hkl).\)

### Structure determination and refinement

The low salt structure was solved first. A double-stranded DNA model in the A or B conformation was created with the program Quanta (Molecular Simulations Inc., Burlington, MA). The program AMoRe (23) was used to find the molecular replacement solution. The space groups P6\(^{1}\)\(^{2}\) and P6\(^{2}\)\(^{2}\) were tried with each model and the only correct solution was the A-form model in space group P6\(^{1}\)\(^{2}\). After an AMoRe rigid body refinement the correlation coefficient and the R factor were 49.3 and 50.2 respectively. X-PLOR v3.85 (24–26) was used to refine the structure with a single DNA strand in the asymmetric unit because the \(V_m\) for a single strand is 2.8 and the molecular replacement solution sets the DNA two-fold axis on a crystallographic two-fold axis. X-PLOR Powell minimization was carried out first with the phosphate backbone, then with three bases at a time (1–3, 2–4…8–10) and finally with all atoms. The X-PLOR parameter file param_ndbx_highdna for high resolution DNA structures was used for all refinements, including hydrogen atoms (27). At each point, \(R_{\text{free}}\) and 2F\(_{\text{o}}\) – F\(_{\text{c}}\) maps were used to monitor refinement progress. At this point, anisotropic scaling was performed with the XtalView program (28) decreasing the \(R_{\text{free}}/R_{\text{factor}}\) from 0.341/0.313 to 0.315/0.299. Finally, four rounds of ordered water molecule addition were performed. The program Quanta was used to locate potential water molecules using a F\(_{\text{o}}\) – F\(_{\text{c}}\) map contoured at 1.5\(\sigma\). Water molecules were kept only when 2–3.5 Å from an appropriate H bonding partner and when electron density appeared in a 2F\(_{\text{o}}\) – F\(_{\text{c}}\) map contoured at 2.5\(\sigma\), or 1\(\sigma\) during the final water addition round. After each round of water addition, the water molecules only were energy minimized using X-PLOR Powell minimization, followed by a round of all atom minimization. Finally, a B factor refinement was performed on all atoms and an occupancy analysis completed on all water molecules. The final refinement statistics for both structures are shown in Table 1.

### Figure 1.

(A) Sequence homology alignment of one of the E2-BS located in the URR of various HPV types. Different viral strains contain varying copy numbers of the E2-BS. BPV 1 stands for bovine papillomavirus type 1 and is included as a reference. The sequences are in the conserved form, ACCnungGT (underlined portion), and flanked by 1 nt on the 5′- and 3′-ends. Note that the internal GAAA(A)(C) are also highly preferred. (B) Sequence and numbering scheme of the presented structure.
Table 2. Helical parameters

<table>
<thead>
<tr>
<th></th>
<th>Global Base pair-Axis Parameters</th>
<th>Global Inter-Base Pair Parameters</th>
<th>Minor Groove</th>
<th>Major Groove</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Salt</td>
<td>High Salt</td>
<td>Low Salt</td>
<td>High Salt</td>
</tr>
<tr>
<td>Adenine</td>
<td>4.62</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Guanine</td>
<td>4.07</td>
<td>3.06</td>
<td>2.13</td>
<td>2.36</td>
</tr>
<tr>
<td>Thymine</td>
<td>3.65</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Cytosine</td>
<td>4.56</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Guanine</td>
<td>4.07</td>
<td>3.06</td>
<td>2.13</td>
<td>2.36</td>
</tr>
<tr>
<td>Thymine</td>
<td>3.65</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Cytosine</td>
<td>4.56</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Adenine</td>
<td>4.62</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Guanine</td>
<td>4.07</td>
<td>3.06</td>
<td>2.13</td>
<td>2.36</td>
</tr>
<tr>
<td>Thymine</td>
<td>3.65</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Cytosine</td>
<td>4.56</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Adenine</td>
<td>4.62</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Guanine</td>
<td>4.07</td>
<td>3.06</td>
<td>2.13</td>
<td>2.36</td>
</tr>
<tr>
<td>Thymine</td>
<td>3.65</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
<tr>
<td>Cytosine</td>
<td>4.56</td>
<td>3.01</td>
<td>2.68</td>
<td>2.44</td>
</tr>
</tbody>
</table>

1Helical parameters calculated with the Curves program (29).
2Groove widths and depths are averages of two to three slices through the groove.
3The low and high salt structures have only one unique strand. Only unique figures are given.
4Standard A- and B-form DNA [d(GACCGCGGTC)] generated with Quanta (Molecular Simulations Inc) and parameters calculated with Curves (29).
5Roll relative to the axis followed by the local axis.

The high salt structure is isomorphous with the low salt structure. Therefore, the final single-stranded, refined, low salt structure was directly used in an X-PLOR rigid body refinement giving an initial $R_{free}$/ $R_{factor}$ of 0.31/0.35 (without the addition of water molecules). After this point, refinement continued as described above.

RESULTS

Overall DNA structure

The decamers of the low and high salt structures are isomorphous and crystallize in the space group P6$_1$22 where they form a right handed screw axis around a large hydrated opening. Two DNA molecules pack with their terminal base pairs into the minor groove of a third molecule, between residues C4 and G7. Additionally, the DNA two-fold axis is coincident with a crystallographic two-fold axis. Thus the asymmetric unit contains only one DNA strand.

In both cases, the DNA structures are well defined by their respective electron densities (Fig. 2). Figure 3 shows the overall DNA conformation viewed along and perpendicular to the helical axis including a vector showing the overall helical curvature as generated with the Curves program (29). It can be seen from Figure 3 that the DNA exhibits typical A-form conformation, evidenced by inclination of the base pairs relative to the overall DNA axis (view perpendicular to the helical axis in Fig. 3) and that the base pairs stack around the helical axis (view along helical axis in Fig. 3).

DNA conformation

The various helical parameters of both structures are compared with canonical A-form and B-form values in Table 2. Among them are several A-form indicators, such as X-displacement (Xdisp), inclination and rise, confirming that the DNAs are in the A conformation. All helical values were calculated with the Curves program (29). Canonical A- and B-form DNA helical values were calculated from DNA oligomers generated with the Quanta program (Molecular Simulations Inc).

The DNA helix of both the structures is curved toward the major groove, with the overall angle of curvature for the low and high salt structures being 18° and 31°, respectively, as calculated with the Freehelix program (30). Excluding the terminal base pairs, the greatest curvature occurs in both cases between the bases C3 and C4 (numbering scheme presented in Fig. 1). The result of this curvature seems to be a narrowing and deepening of the minor groove (see Table 2).

In general, the backbone torsion angle and furanose ring conformations for the low and high salt structures are in good agreement with canonical A-form DNA. There are, however, two deviations from canonical values. The first is an extended trans $\alpha$ and $\gamma$ torsion angle around the P8 atom. This conformation has been seen in other DNA crystal structures such as the A-form DNAs d(ACCGGCCGTC) (31) and d(CCCGCGCGGG) (32), each having one trans conformation. This extended trans conformation may be the result of an interesting packing and intramolecular hydrogen bonding network shown in Figure 4. Water molecule numbers 47 and 44 mediate hydrogen bonding.

Figure 2. Final 2Fo – Fc electron density maps of the base pair C4-G17. This and other figures were generated with the Ribbons program (62). (A) Low salt structure contoured at 2$\sigma$. (B) High salt structure contoured at 1$\sigma$. 
between the atoms O3' of G7 and O5' of G8, respectively, to neighboring oxygens of symmetry-related DNA molecules. These water molecules may also form hydrogen bonds with O2P of G8. Further stabilizing this structure are two other water bridges. The first from water molecule numbers 44, 54 and 45 to OP1 of G8 and the second from water molecule numbers 47, 28 and 17 to O1P of G8. The high salt structure has similar features in this region. However, it lacks two water molecules, numbers 54 and 17. Recently, Tippin and Sundaralingam have argued for the existence of a similar structural water molecule stabilizing an extended trans conformation in the A-form octamer d(CCTAGG) (33). Also, in the DNA structure d(CCCGCGG), waters seemingly stabilize an extended trans conformation around a backbone P atom (32). This structure has waters corresponding to water molecules 47, 45 and 44 in the present low salt structure.

The second deviation is that the terminal sugars are in the C2'-exo conformation. Although C2'-exo is an allowed A-form sugar pucker (34), it is not the canonical C3'-endo conformation. A hydrogen bond may exist between the terminal O4' atom of G1 and the N2 atom of a symmetry-related G7 in both structures. This interaction, unique to the terminal sugars, is probably responsible for stabilizing the C2'-exo sugar conformation. The terminal sugar pucker phase angles are 356° and 354° for the high and the low structures, respectively, calculated with the Curves program (29). Another effect of this hydrogen bond may be the low twist values exhibited by the G7:C14 base pair (Table 2).

Base stacking
The base stacking of both structures is similar regardless of the salt concentration. Comparisons of the base stacking in these structures with the canonical A-form DNA generated with Quanta (Molecular Simulations Inc.) makes it evident that the base stacking of the main body is nearly ideal, while some deviations are present primarily in the terminal base pairs. A typical stacking interaction is that the small ring of a guanine nucleotide is directly over the large ring of the next guanine nucleotide, as in C3:G18/C4:G17 (35). The stacking of the first G-C and the second A-T base pair is furthest away from the canonical A-form DNA in which the large ring of G1 is not perfectly positioned over the small ring of A2. Another deviation from the canonical A-form DNA is the fourth base pair stack C4-G5/C16-G17. As seen in other A-form DNA structures (31,32,36), this stacking interaction exhibits a common configuration of the G5 base crossing over and stacking with the G17 base of the other strand.

Hydration
The low and high salt structures both contain very well-ordered water molecules (on average, 50 ordered water molecules per DNA strand) with a high degree of similarity among them. The high and low salt structures contain 52 and 47 ordered water molecules, respectively, associated with the single DNA strand. Among them, 28 water molecules are conserved between both structures.

The phosphate oxygens are typically hydrated extensively in A-form DNA structures (37) with water molecules often bridging the O1 atoms between successive phosphates (38). Waters are not observed on five out of 18 phosphate oxygens in the low salt structure and three out of 18 in the high salt structure. Both the low and high salt structures have waters bridging the O1P atoms between P2-P3, P5-P6 and P6-P7.

The major grooves of both structures are extensively hydrated with nearly all possible atoms in good hydrogen bonding distances to water molecules. The minor groove, however, is hydrated to a lesser extent. The most notable exceptions are G5,
Figure 4. Hydrogen bonding of the extended trans configuration around the P8 atom of the low salt structure. The labeled dotted lines indicate distances in Å. Note that the high salt structure is missing waters corresponding to the low salt water molecule numbers 17 and 54 (62).

DISCUSSION

Effects of salt, hydration and packing on the DNA structure

The helical parameters seem to vary little with salt concentration. Thus, it seems that the differences in the helical parameters from canonical values may be the result of the sequence itself. Sodium ions are also present in the crystallization conditions of each crystal (used to adjust the buffer pH) and may account for the structural similarities, as well as the exclusion of magnesium and spermine from the structures. The fact that magnesium and spermine are not seen is not completely unexpected. The NDB (Nucleic Acid Database) contains >300 A- or B-form DNA crystal structures of which only 44 report magnesium and only 15 report spermine. Evidence that the salt concentration does not play a significant role in determining the DNA conformation is seen by examining helical rise (Table 2). The average rise for both salt conditions is 2.8 Å, indicating that the DNA conformation is relatively independent of salt conditions (within the limits of 1.56–50 mM MgCl2 and spermine; Materials and Methods). The Xdisp is slightly less than that of canonical A-form DNA but the inclination is ∼8° less than canonical A-form DNA. These deviations may be complemented by the large tilt, roll and rise (see Table 2). The twist is near that of canonical A-form DNA with one exception. The low twists of bases C3-C4 and G7-G8 could be a result of packing interactions. However, low twist values have been seen in C-C stacks as well as in G-C stacks (31,32,36). As mentioned above, the furanose ring of G1 makes contact with the N2 atom of the symmetry-related G7 (base paired to C3), which may be responsible for the low twist.

The variations in salt concentration do not seem to influence the DNA hydration or the overall DNA conformation significantly. The high salt structure in this study has only five ordered water molecules less than the low salt structure. This may indicate that the increase in ionic strength (magnesium or spermine) may
reduce water activity so less water molecules bind DNA tightly. The salt concentration does have a slight effect, however, on the unit cell parameters and the overall curvature of the DNA molecules. Inspection of Table 1 reveals that when the salt concentration is increased, the unit cell dimensions decrease. Axes along a and b decrease almost 0.3 Å, while the c axis decreases nearly 0.4 Å. The DNA conformation, however, remains relatively unchanged (Table 2). The helical rise, larger than that of canonical A-form DNA, remains unchanged with an average rise of 2.8 Å for both structures. The DNA molecules apparently can accommodate the unit cell changes by adjusting the overall angle of curvature. As the unit cell dimensions shrink with the increased salt concentration, the angle of curvature increases from 18° (low salt) to 31° (high salt), which may indicate structural flexibility.

The hydration and packing has two specific effects on the DNA conformation. As already pointed out, the extended trans conformation around the P8 atom may be stabilized by several conserved water molecules with a hydrogen bonding network through symmetry-related DNA strands (Fig. 4). The second effect is the C2'endo conformation of G1 and the low twist of G7, probably caused by a hydrogen bond between these two residues in symmetrically related DNA strands.

**Biological implications: the A-form, curved helical axis and protein binding**

Crystal structures of DNA oligomers in the A-form have traditionally been questioned, especially when crystallized with the commonly used MPD reagent (which may cause dehydration as well as induce the A conformation). This is because the central dogma has been the belief that the B-form is the relevant biological DNA molecule; while the A conformation has been reserved only for RNA molecules, the natural conformation of RNA and RNA–DNA chimeric duplexes (40). Although it is known that MPD may have a straightening effect on poly(A)-derived curvature (41), CD spectra and topoisomerase assays of curved DNAs rule out conformational changes to structures induced by MPD outside the B-DNA family (42). Additionally, the A-form is believed to exist in spores of the gram-positive bacteria *Bacillus* (43). Finally, the idea has been advocated that an intermediate B/A solution conformer may be adopted by non-alternating d(G+C)-rich oligos (44-48). Certain DNA sequences may contain A-form characteristics (49). In particular, the GpG base step, which has been shown to form A-type DNA by fiber diffraction, Raman spectroscopy and recent NMR studies of a 34mer (40,50-54). Many DNA regulatory regions are known to contain repeats of GpG, including the E2-BS d(ACC):d(GGT). Several of these GpG-containing DNA elements have now been crystallized in the A conformation, for instance, the promoter region of simian virus 40 with which the transcription factor Sp1 interacts. The structure of half the Sp1 binding site has been solved and is in the A-form, d(GATGGGAG) (55). Furthermore, the crystal structure of the TATA-containing octamer d(GGATACCC) is in the A-form (56). TFIIA can recognize both the 5S RNA gene, the 5S RNA molecule and its cognate DNA element crystallized in the A-form (57). Although the biological role of A-form DNA remains unclear, these recent results may suggest a role for A-form DNA in these regulatory regions.

The overall DNA curvature is towards the major groove. This is reflective of the large global roll values of the C3-C4 and C4-G5 base steps (Table 2). It appears that this DNA sequence may have a natural tendency for curvature. Biburger et al. have shown that properly phased d(C-G) runs have intrinsic curvature (45). It is striking that the human virus strains have a preference for three to four adenines between the conserved d(ACC) regions of the E2-BS (Fig. 1; 19-21). Adenine tracts have a general tendency to be bent in solution (42,58-60). d(G)d(C) runs were also shown to enhance A-tract curvature (61). More importantly, the bending between C3 and C4 in the present structures may be a unique characteristic signature of this DNA sequence necessary for high affinity protein binding. The present A-form structures, the A-form DNA sequence d(ACCGGCGCGT) (31), the A-form DNA sequence d(CCCGCGCGCG) (32) and the B-form E2-BS of the bovine papilloma virus (BPV) E2 protein–DNA complex (18) are all bent in a similar manner. The overall angle of curvature for the low and high salt structures are 18° and 31°, respectively, with the greatest curvature occurring in each case between bases C3 and C4. The C-terminal domains of the BPV 1 and HPV 16 E2 proteins can induce DNA curvature of the E2-BS (16,17). The change seen in helical curvature may be reminiscent of the protein-induced curvature. Frederick's structure is curved ~18° with the greatest base step curvature accruing between C2 and C3. Overall, Ramakrishnan's structure is curved 25°; the greatest base step curvature is observed at the G4-G5 step. The DNA d(CCGACCGACGTCGGGT) complexed with the BPV E2 protein has the greatest overall curvature of 33°. Although the complexed DNA is in the B-form, the bending similarities are striking. The greatest base step curvature is between the base steps C5:C6 and C6:G7. Figure 5 summarizes the helical axis curvature of the discussed DNA structures. Note that all curving data have been calculated with the FreeHelix program (30) and terminal residues are not included in determining the greatest base step curvature.

We hypothesize that the curving of d(CC) in the d(ACC) specific recognition sequence is intrinsic to this sequence and may be required for high affinity protein recognition. HPVs also have a preference for a short poly(A) tract within the E2-BS (Fig. 1; 19-21). We suggest that the inherent curvature of poly(A) tracks in solution (42,59,60) is the reason for this preferential selection by the HPVs. The curvature of the d(CC) combining and enhancing the poly(A) tract curvature (as discussed above) may create a pre-curving situation of the E2-BS allowing high affinity protein recognition. This hypothesis suggests that the overall three-dimensional structure of the DNA helix, not just the specific nucleotide sequence in the binding site, plays a critical role in specific DNA recognition by regulatory proteins.

**CONCLUSIONS**

In conclusion, we found the DNA to exhibit sequence-specific curvature at the d(CC) step. We believe that this curvature is intrinsic to the present sequence and provides the appropriate three-dimensional structure for protein binding. This belief is based on the comparison of the presented structure at two different ionic strengths to similar DNA structures, all exhibiting similar curvature, the most relevant being the BPV-1 E2 protein complexed with its E2-BS DNA target sequence (18). Overall, the various salt conditions used in crystallization did not have a large effect on the DNA conformation or hydration. However, a greater DNA curvature is observed at higher salt concentrations,
Figure 5. Graphs displaying the helical curvature of the four DNA structures discussed in the text. The DNA sequence and base steps are indicated in each graph: (A) presented structure (low salt); (B) Frederick et al. (31); (C) Ramakrishnan and Sundaralingam (32); (D) Hegde et al. (18). (A)–(C) A-form DNAs; (D) B-form DNA in a protein–DNA complex. Note that in (D), the terminal cytosine overhang is not included in the curvature calculations.

which may suggest inherent structural flexibility and account for the small unit cell change (Table 1). Coordinates have been deposited in the Nucleic Acid Database, Rutgers University and assigned the NDB accession nos AdJ0102 and ADJ0103 for the low and high salt structures, respectively.

ACKNOWLEDGEMENTS

Many thanks to Dr Loren Williams (School of Chemistry and Biochemistry, Georgia Institute of Technology) whose advice on DNA crystallography was intricate in our study. We thank Dr David Borhani (Southern Research Institute) for his assistance in the use of the precession camera. We also thank Dr Yu Luo (Center for Macromolecular Crystallography, University of Alabama-Birmingham) for generating a computer program for the water analysis. Finally, we thank Dr Champion C.S.Deivanayagam (Center for Macromolecular Crystallography, University of Alabama-Birmingham), Dr Steve Harvey (Department of Biochemistry and Molecular Genetics, University of Alabama-Birmingham) and Barbara Lommer (Center for Macromolecular Crystallography, University of Alabama-Birmingham) for discussions and critical reading of the manuscript. J.B.F. was supported by training grant T32AI07150 from the NIAID.

REFERENCES

49 Frederick, C.A. and et. al. (1987) Biopolymers, 26, 551–552.