Curved DNA without A-A: Experimental estimation of all 16 DNA wedge angles

(DNA curvature/anomalous migration/Eulerian angles)

A. Bolshoy*, P. McNamara†, R. E. Harrington‡, and E. N. Trifonov*

*Department of Polymer Research, The Weizmann Institute of Science, Rehovot 76100, Israel; and †Department of Biochemistry, University of Nevada Reno, Reno, NV 89557-0014

Communicated by Peter H. von Hippel, December 10, 1990

ABSTRACT The principal sequence feature responsible for intrinsic DNA curvature is generally assumed to be runs of adenines. However, according to the wedge model of DNA curvature, each dinucleotide step is associated with a characteristic deflection of the local helix axis. Thus, an important test of a more general view of sequence-dependent DNA curvature is whether sequence elements other than A-A cause the DNA axis to deflect. To address this question, we have applied the wedge model to a large body of experimental data. The axial path of DNA can be described at each step by three Eulerian angles: the helical twist, the deflection angle (wedge angle), and the direction of the deflection. Circularization and gel electrophoretic mobility data on 54 synthetic DNA fragments, both from other laboratories and from our own, were used to compare the theoretical predictions of the wedge model with experiment. By minimizing misfit between calculated and observed DNA curvature, we have found that the stacks AG/CT, CG/CG, GA/TC, and GC/GC, in addition to AA/TT, have large wedge values. We have also synthesized seven sequences without AA/TT elements but with these other wedges correctly phased to cause appreciable predicted curvature. All appear curved as demonstrated by anomalous gel mobilities. The full set of 16 roll and tilt wedge angles is estimated and, together with the known 10 helical twists, these allow prediction of the general sequence-dependent trajectory of the DNA axis.

Intrinsic DNA curvature (reviewed in refs. 1–4), as reflected in anomalous electrophoretic mobility (5) and the capacity to form very small circles (6), is a phenomenon that is generally believed to be a function of periodically repeating A-A dinucleotides (5, 7). DNA fragments known to exhibit large electrophoretic anomalies in polyacrylamide gels have A-A dinucleotides in the form of repeated runs of four to six adenines alternating with mixed sequences (5, 6, 8–11). The simplest model of DNA curvature, the junction model, describes global curvature as the result of a deflection at each junction between the axes of the normal B-DNA and the B'-DNA of the AA/TT stretches (11, 12). The deflection is proposed to be caused by the difference in the inclination of base pairs in these two helical forms and is localized at the junctions of the two averaged helical axes.

In addition to runs of adenines, however, there are other sequence elements that may also cause deflection of the helical axis. These may be either isolated A-A dinucleotides (13–15) or other dinucleotides (16–18). Thus, a complete description of DNA curvature must be in terms of a more general model in which every base pair step is associated with some deflection of the DNA axis. The wedge model accounts for the three-dimensional trajectory of DNA as the vectorial sum of all dinucleotide wedge deflections along the length of the molecule (7, 19).

The simplest form of the wedge model is one in which neighboring base-pair stacks are geometrically independent. In this model, the DNA axis deflections are described by 16 wedge angles, each comprised of roll and tilt components, for the stacks AA/TT, AC/GT, AG/CT, CA/TG, CC/GG, and GA/TC (six rolls and six tilts) and the self-complementary stacks AT/AT, CG/CG, GC/GC, and TA/TA (four rolls). The experimental estimation of these 16 angles is a formidable task requiring a very large amount of experimental data. If it is assumed that the A-A wedge (roll and tilt combined) makes the only significant contribution to DNA curvature, its magnitude may be estimated to be 7° to 11°, based on the circularization of synthetic DNA fragments (6). From the gel migration anomalies of two DNA sequences with A-A and T-T elements in different orientation (20), values for the roll and tilt components of the A-A wedge have been estimated (21). Any number of experiments (equations) may always be approximately satisfied by one or another sufficiently large set of angles (22–25). However, a realistic solution must fit experimental data closely, such that the dispersion of each predicted wedge angle is no larger than 1 SD.

Absolute values for DNA curvature can be obtained from several ring-closure experiments with different repeating sequences (6, 26–28). In addition, a large number of independent measurements of the gel electrophoretic mobility of synthetic DNA fragments with various periodic sequences is available today, all obtained under virtually identical experimental conditions (10, 11, 18, 20, 29–31). We have supplemented this data base by gel migration data for 17 specifically designed additional fragments. Altogether, 54 equations are available, one for each independent experiment. The resulting system of equations allows the estimation of all 16 roll and tilt wedge angles. The solution that we have derived is unique and the distribution of misfits between experimental and calculated curvature and relative mobility values is very close to normal with the dispersion equal to 1.0 of the standard error. Using the preliminary solution of a smaller system of equations, we have designed and synthesized seven DNA molecules without A-A (or T-T) dinucleotides in their sequences. These clearly exhibit curvature as manifested in polyacrylamide gel mobility anomaly (this work) and ring-closure experiments (P.M. and R.E.H., unpublished work).

MATERIALS AND METHODS

Computational Data Base. The data base for the computations consists of published and our own gel migration data on synthetic DNA fragments, both straight and curved, as manifested by anomalous gel mobilities. These data are self-consistent and all have been obtained under "standard" gel conditions as defined in ref. 10—i.e., nondenaturing 8% polyacrylamide gels, 90 mM Tris borate/2.5 mM EDTA, pH 8.3, applied voltage 7 V-cm⁻¹, room temperature. We have

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
Since the solution conditions for DNA circularization are not identical to the conditions for electrophoresis, DNA curvature and thus experimentally measured wedge angles may not be the same for the different ionic environments. In the calculations below, it is arbitrarily assumed (see, however, ref. 32) that the difference is proportional; i.e., the angles of one set are equal to those of another set multiplied by some unknown factor common to all of the angles. The only purpose of introducing the few ring-closure experiments into the system is to provide a degree of absolute calibration through these values for the wedge angles. The relative signs and values of the wedge angles are derived from the bulk of electrophoretic anomaly measurements, which therefore are calibrated by the absolute curvature measurements in the circularization experiments. Thus, the computational solution corresponds to the ionic and temperature conditions during ligation. The electrophoretic anomalies ($R_L$ values) of 90 base-pair-long DNA fragments have been obtained from our own or the respective published anomaly plots. This length was chosen as a compromise between larger anomaly values for longer fragments and smaller nonplanarity effects (10, 33) for shorter ones.

**Computation of Wedge Angles.** Taking a base pair as a rigid body, its position in space relative to either nearest neighbor can be described as a combination of a rotation and a translation. A rotation can be expressed in terms of three independent angles: (i) rotation by half of the helical twist $\Omega/2$ about the axis $z$; (ii) rotation by the wedge angle $\sigma$ about the line in the $x$-$y$ plane perpendicular to the direction of the DNA axis deflection, when the direction angle $\delta$ is measured from the new position of the axis $x'$; and (iii) rotation by half of the helical twist $\Omega/2$ about the new position of the axis $z'$ (A.B. and E.N.T., unpublished work). The angles $\Omega$, $\sigma$, and $\delta$ are analogous to Eulerian angles. An important property of these angles is that, for complementary dinucleotides, the direction angle $\delta$ changes its sign, while for self-complementary ones, it is equal either to zero or to 180°. Wedge roll and wedge tilt are expressed as a product of wedge angle value and cosine or minus sine of the direction angle, respectively, in accordance with the Cambridge convention (34).

The total number of unknowns to calculate is 19: 16 wedge angles and three parameters, $A$, $B$, and $C$, which define a “calibration” curve of $R_L$ value versus DNA curvature, assumed to be parabolic (3, 11, 13, 21). The actual calculation is done as follows. A product matrix $A$ is defined as

$$A(\Omega, \sigma, \delta) = \prod_{j=1}^{n} T(\Omega, \sigma, \delta) \times M(\sigma, \delta),$$

where $T$ is the twist rotation (see $i$ and $iii$ above) and $M$ is a “middle rotation” as defined in $ii$ above. The values of $\Omega$, $\sigma$, and $\delta$ are defined for the $j$th dinucleotide in the sequence of length $n$ and the product is carried out over the entire sequence. A total deflection angle is defined as the angle between the vectors normal to the mean planes of base pairs with repeat lengths of 10 or 11 base pairs are included, since the repeat length in the case of straight fragments is of no importance. Fragments 7 and 54 were excluded from the final system of equations (see Discussion).
of the first and last base pairs of a given fragment. It is obtained directly from the elements of $A$:

$$\text{total deflection angle} = \arccos(a_{33}),$$

where $(a_{33})$ is the right-bottom element of the product matrix $A$. Finally, a calculated $R_L$ is obtained from the computed curvature

$$R_L = A + B(\text{curvature})^C$$

$$= A + B(\text{[total deflection angle]})^C,$$

where $k$ is the proportionality between the total deflection angle and the curvature (expressed in curvature units; see legend to Table 1) and $A$, $B$, and $C$ are the coefficients of the quadratic form of $R_L$ and therefore the parameters that define the calibration curve. The deviation of $R_L$ in ref. 3 from the corresponding observed value for a given sequence is a measure of the misfit for that sequence. To estimate the wedge angles, the 54 equations corresponding to the sequences of Table 1 are solved simultaneously subject to the constraint that the average misfit, a global parameter, is minimized:

$$\text{Average misfit} =$$

$$\frac{1}{n} \left[ \frac{(\text{calculated } R_L - \text{observed } R_L)^2}{\text{experimental error}} \right]^{1/2}.$$  

Since $R_L$ is an exponential function of the parameter $C$ and the total deflection angle is a complex trigonometric function of the wedge angles, this becomes a problem in nonlinear parameter estimation (35).

To find the minimum of the total misfit, we used a modified Gauss-Newton gradient descent method (35, 36). This method is extremely sensitive to initial guess of parameters and generally can converge only to a global minimum from arbitrary initial values. In our calculations, the initial values of $A$ and $C$ were taken equal to 1 and 2, respectively, and the initial value of $B$ was estimated by assuming roll and tilt angles of A-A as in ref. 21. This also sets the directions (signs) of all the wedge angles relative to established signs for A-A components. Initially, the choice of wedge angles other than A-A was more or less arbitrary. The arbitrary initial guesses of the remaining wedge angles resulted in many solutions; however, all were indistinguishable within one standard deviation. After many cycles of computation and readjustment of these values, the system converged to a consistent global minimum in average misfit. The misfit surface became smooth after the calibration parameters $A$, $B$, and $C$ were averaged around this consistent minimum and fixed, leading to a unique global minimum. Final refinements were then made by consecutive small readjustments of the calibration curve parameters followed by descents along the gradient until the smallest average misfit value was attained.

Oligonucleotide Synthesis and Preparation. Oligonucleotides (Table 1, fragments 26-36 and 48-54) were made on a Du Pont Coder 300 DNA synthesizer. Each sequence is 21 nucleotides long so that sequence elements repeat at roughly helical intervals. The duplex 21-mers have unique 5-base overhangs, ensuring head-to-tail polymerization. DNA was phosphorylated with $[\gamma^32P]ATP$ (Du Pont/NEN) using T4 polynucleotide kinase (10), purified in 15% denaturing polyacrylamide gels, and electroeluted. Complementary strands were mixed, heated to 70°C, and cooled slowly to 4°C. Five milligrams of duplex DNA was ligated in 70 mM Tris Cl, pH 7.5/10 mM dithiothreitol/2.5 mM EDTA using 1 unit of T4 DNA ligase (BRL) in a total volume of 50 μl. Reactions were allowed to proceed for ice 1 hr. The products of self-ligation were resolved in 32-cm 8% polyacrylamide gels run in 90 mM Tris borate/2.5 mM EDTA, pH 8.3, room temperature, at 7 V cm⁻¹. Mobility measurements were made relative to the migration of a BamHI linker 10-base-pair ladder (known to run with normal mobility). $R_L$ values for 90-base-pair DNA were interpolated from plots of apparent to actual length in base pairs.

RESULTS

The first round of calculations included all experiments with fragments containing A-A dinucleotides and control straight fragments (Table 1, circles 1-3 and fragments 4-47). The best fit solution for this partial system of equations is shown in Table 2, which indicates that the dinucleotide elements AG/CT, CG/CG, GA/TC, and GC/GC possess appreciable wedge angles. To challenge the solution, additional curved DNA fragments were designed, which contain these elements but lack AA/TT stacks (Table 1, repeat units 48-54). These DNA fragments, indeed, are found to exhibit appreciable sequence-directed curvature as reflected in their gel migration anomalies. Circularization experiments also confirm that the fragments are, in fact, curved (P.M. and R.E.H., unpublished work). The solution for the complete system, including the anomaly data from the six sequences without A-A, is shown in Table 3. It is not significantly different from the original solution. This demonstrates both the stability of the system of equations and the uniqueness of the solution. Hence, the wedge angles for AG/CT, CG/CG, GA/TC, and GC/GC are verified.

The calibration curve, $R_L$ value versus curvature, is presented in Fig. 1. The average misfit value for the entire system (excluding fragments 7 and 54, which do not fit the pattern of the remaining 52 fragments well, probably for the reasons discussed below) is equal to 0.99. The expected range for a normal distribution is 1.0 ± 0.1, estimated (35) for the system of 52 equations.

DISCUSSION

The solution given in Table 3 has good predictive power and correctly describes gel electrophoretic mobility anomalies of DNA molecules of any given sequence. This includes select...

Table 2. Wedge angles derived from A-A-containing and control fragments

<table>
<thead>
<tr>
<th>Wedge angles, degrees</th>
<th>Orthogonal components, degrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eulerian</td>
<td>Direction</td>
</tr>
<tr>
<td>(σ)</td>
<td>(θ)</td>
</tr>
<tr>
<td>A-A</td>
<td>7.2</td>
</tr>
<tr>
<td>A-C</td>
<td>1.0</td>
</tr>
<tr>
<td>A-G</td>
<td>7.6</td>
</tr>
<tr>
<td>A-T</td>
<td>2.1</td>
</tr>
<tr>
<td>C-A</td>
<td>3.4</td>
</tr>
<tr>
<td>C-C</td>
<td>2.3</td>
</tr>
<tr>
<td>C-G</td>
<td>6.5</td>
</tr>
<tr>
<td>G-A</td>
<td>5.1</td>
</tr>
<tr>
<td>G-C</td>
<td>6.0</td>
</tr>
<tr>
<td>T-A</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Eulerian angles indicated are those actually used in and derived from the calculations. The twist angles $Ω$ (not shown) are taken from ref. 37. The orthogonal wedge components, roll (ρ) and tilt (τ), are calculated from the Eulerian angles by the formulas $\rho = \sigma(\cos \theta)$ and $\tau = -\sigma(\sin \theta)$. Signs corresponding to those defined by the Cambridge convention (34). Computational error bars are estimated from second partial derivatives (35) and indicated in parentheses, for wedge rolls and tilts only. The system included circles 1-3 and fragments 4-6 and 8-47 (Table 1). Average misfit value for the system is 0.86 SD.
The computational system included circles 1–3 and fragments 4–6 and 8–53 (Table 1). Further details are given in the legend to Table 2.

sequences that do not have AA/TT dinucleotides. It also evidently correctly defines DNA curvature, since the statistical fit of the system remains excellent even when the ring cyclization data are included. According to these estimates, deflections in the helical axis of DNA are caused primarily by AG/CT, AA/TT, CG/CG, GA/TC, and GC/GC base-pair stacks. Several other dinucleotides also contribute to DNA curvature to a lesser extent (Table 3). Given the limited accuracy of the experimental data, the present set of 16 wedge angle values can be considered only an estimate. More elaborate calculations, based on specifically designed experiments, are necessary to derive more accurate values for the wedge angles. However, these values are useful in describing the approximate three-dimensional path of specific DNA molecules. They have recently been used, for example, to rationalize nucleosome placement on certain DNA sequences (38).

Two sets of angles suggested earlier by other groups (22, 23) give either no estimates or incorrect ones for the AG/CT, CG/CG, GA/TC, and GC/GC wedges. The misfit values calculated for our experimental data set with the wedge angles in refs. 22 and 23 are 2.7 SD and 2.0 SD, respectively. The misfit value for solution with A-A roll and tilt only is 1.9 SD. Only by introducing several more DNA axis deflecting elements (Table 3) can agreement between the wedge model and the experiments be achieved. The existence of non-A-A wedge elements is demonstrated most clearly by the curved non-A-A fragments (Table 1, fragments 48 and 50–53).

Studies on curved DNA fragments with adenine runs suggest that there are some cooperativity effects—i.e., differences in reactivity, hydrogen exchange, and chemical shifts among A-A dinucleotides located near the middle of an adenine run and at or near its ends (39–41). These findings have been interpreted as evidence for structural variation along a poly(A) tract. Evidence for this kind of cooperativity has recently been reviewed by Hagerman (42). In general, the geometry of every dinucleotide stack should depend in some fashion upon neighboring stacks. However, existing manifestations of such cooperativity effects do not provide any specific geometrical or structural parameters that could be incorporated into the wedge model. We have therefore used the simplest noncooperative variant of the wedge model for the calculations reported here. This means that the AA/TT wedge angles presented in Tables 2 and 3 represent complicated (and, in general, nonlinear) averages over any cooperative effects in the poly(A) tracts of sequences given in Table 1. However, Table 1 includes sequences ranging from 0 to 20 consecutive adenines, including a number with fewer than 6. The fact that the solution of this system fits to experimental data with an average misfit of 1.0 SD is an indication that cooperativity effects are not statistically significant in the present context.

The two fragments excluded from the system of equations (Table 1, fragments 7 and 54) demonstrate anomalously high mobilities (low $R_1$ values) as compared to predicted values. Unusually high mobility is also observed in several DNA fragments at fragment lengths beyond 84–105 base pairs (Table 1, fragments 50–52). A maximum in the $R_1$ value versus length plot is also observed in these cases with $R_1$ actually decreasing at longer fragment lengths (43). Such a fall-off in friction with DNA length during migration in the gel is suggestive of a structural change in the DNA. One likely possibility is the formation of transient kinks at certain weak sequence elements. Of the sequences exhibiting this behavior, the stacks CA/TG and AC/CT appear to be likely sites for such kinking. Indeed, all fragments investigated here that show unusually high mobilities contain the elements CA and TG. NMR studies have suggested that CAC sites in DNA may possess unusual structure (44, 45), possibly involving partially unstacked bases (46). An apparently stable kink has recently been observed crystallographically at the CA step in the catabolite activator protein (CAP)–DNA complex (47). The ubiquitous nature of CAP triplets and runs of CA in a wide variety of protein–DNA recognition sites has been anecdotaly noted and has been qualitatively documented in at least one limited study (48). Finally, the unique and highly specific importance of CAC as a protein recognition feature has been clearly demonstrated in elegant studies on the catabolite activator protein system (49) and in variables (diversity) joining [V(D)J] recombination (50). We believe that certain other dinucleotide steps may have similar, although probably less dramatic, "kinkability" qualities as well (43). Thus, the concept of general sequence-directed DNA curvature and the potential formation of kinks, including the possibility of a structural basis for CA as a feature in protein recognition, may introduce a new dimension in DNA–protein interactions.

We acknowledge helpful discussions with Drs. E. Charney, D. M. Crothers, T. Jovin, D. Rao, and V. B. Zhurkin. Assistance with computer graphics by Dr. H. S. Ho is highly appreciated. This work was supported in part by National Institutes of Health Research Grant 33435 (R.E.H.), National Institutes of Health Cancer Training

---

**Table 3.** Wedge angles derived from the entire set of experimental data

<table>
<thead>
<tr>
<th>Eulerian angles, degrees</th>
<th>Orthogonal components, degrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wedge</td>
<td>Direction</td>
</tr>
<tr>
<td>(ρ)</td>
<td>(δ)</td>
</tr>
<tr>
<td>A-A</td>
<td>7.2</td>
</tr>
<tr>
<td>A-C</td>
<td>1.1</td>
</tr>
<tr>
<td>A-G</td>
<td>8.4</td>
</tr>
<tr>
<td>A-T</td>
<td>2.6</td>
</tr>
<tr>
<td>C-A</td>
<td>3.5</td>
</tr>
<tr>
<td>C-C</td>
<td>2.1</td>
</tr>
<tr>
<td>C-G</td>
<td>6.7</td>
</tr>
<tr>
<td>G-A</td>
<td>5.3</td>
</tr>
<tr>
<td>G-C</td>
<td>5.0</td>
</tr>
<tr>
<td>T-A</td>
<td>0.9</td>
</tr>
</tbody>
</table>

The regression curve (calibration curve) corresponds to the function $R_1 = 1.01 + 1.36 \times (\text{curvature})^{2.35}$. The average misfit is 0.99 SD. Open symbols correspond to those experiments for which the (individual) misfit exceeds 1 SD (Table 1).

**Fig. 1.** Calibration of DNA electrophoretic mobility measurements. The experimental $R_1$ values (Table 1) are plotted as a function of the DNA curvature of the corresponding fragments, calculated from the wedge and direction angles (see Table 3). The continuous regression curve (calibration curve) corresponds to the function $R_1 = 1.01 + 1.36 \times (\text{curvature})^{2.35}$. The average misfit is 0.99 SD. Open symbols correspond to those experiments for which the (individual) misfit exceeds 1 SD (Table 1).
Grant T32CA09563, and Hatch Project 142 from the University of Nevada (R.E.H.).