Biochemistry **1990**, *29*, 9269–9273

Validity of the Nearest-Neighbor Approximation in the Evaluation of the Electrophoretic Manifestations of DNA Curvature[†]

Pasquale De Santis,[*,‡] Antonio Palleschi,[§] Maria Savino,[∥] and Anita Scipioni[‡]

Dipartimento di Chimica and Dipartimento di Genetica e Biologia Molecolare, Università di Roma "La Sapienza", P. le A. Moro 5, 00185 Roma, Italy, and Istituto di Chimica, Università della Basilicata, Via N. Sauro 85, 85100 Potenza, Italy

Received September 13, 1989; Revised Manuscript Received May 14, 1990

ABSTRACT: The validity of a theoretical nearest-neighbor model based on conformational energy calculations, for translating the deterministic fluctuations of DNA base sequences in superstructural elements, is tested on the electrophoretic data of 450 multimeric, duplex oligonucleotides with different sequence, periodicity, and complexity, corresponding to all the experimental data so far published. An accurate semiempirical linear relation between the gel electrophoretic retardation and the dispersion of the theoretical curvature was found which allows a prediction of electrophoretic anomalies of DNA tracts, e.g., the cyclic permutation gel assays, with great confidence. Such a pattern of agreement allows the interpretation of the gel electrophoresis retardation as due to an increase of activation energy required in straightening the DNA axis.

The existence of sequence-dependent DNA curvature is now becoming a general concept and the object of many investigations. Essentially, two classes of models have been proposed to explain the superstructural feature of DNA at molecular level. The first one derives from the original idea of Trifonov (1980), based on the formation of wedges between adjacent base pair planes in the dinucleotide steps of the double helix, particularly significant for the AA sequence; repetition of wedges in phase with the DNA helical periodicity produces macroscopic curvatures of the double helix as a result of additive local effects. Models of such a kind are known as nearest-neighbor models.

The other class of models proposes that distal interactions beyond the nearest-neighbor level contribute significantly to the curvature of DNA. The first relevant example is represented by the junction model proposed by Crothers and coworkers (Koo et al., 1986; Koo & Crothers, 1988), which assumes that AA tracts containing more than three adenines adopt a helical conformation different from the canonical B structure; as a consequence, a bend of the helical axis is produced at the conformational transition between the two structures.

Three years ago we developed a model based on the theoretical evaluation of the local deviations from the standard B-DNA structure of the 16 different dinucleotide steps using energy conformational calculations (De Santis et al., 1986). The relevant local deviations on a given sequence of nucleotides were easily integrated by adopting an original mathematical method resulting from the reduction of the transformation matrices at the limit of small angles. A strikingly good correlation between electrophoretic retardation and the modulus of the theoretical curvature was found for all the periodical biosynthetic DNAs until then reported in literature.

The model was more recently deeply investigated, and a very good linear dependence was found for the electrophoretic retardation of tracts of DNAs with different lengths, sequences, and sequence periodicities and the pertinent dispersion of the curvature σ^2 (De Santis et al., 1988b, 1989; Cacchione et al., 1989).

The present paper illustrates the application of this nearest-neighbor model to the theoretical evaluation of the electrophoretic behavior of all the periodical biosynthetic double helix polynucleotides so far reported in the literature as well as of some representative tracts of DNAs. Sixty-two different monomeric sequences of various polymerization degrees were investigated corresponding to about 450 experimental data, and a correlation coefficient of 0.98 was obtained between experimental and theoretical electrophoretic retardations.

RESULTS

The problem of obtaining the macroscopic curvature by integrating over a tract of sequence the local distortions from the standard B-DNA structure is correctly set up in terms of transformation matrices; however, taking advantage of the smallness of the deviation angles involved, we reduced the problem, at the first order of the pertinent Taylor series, and used the complex representation of the rotation matrices, obtaining the following simple form as shown in the appendix:

$$\bar{C}(n) = \nu(n_2 - n_1)^{-1} \sum_{s=n_1}^{n_2} \mathbf{d}_s \exp(2\pi i s/\nu)$$

where $\bar{C}(n)$ is the average curvature vector, namely, the orientational deviation of the curved helical axis between n_1 and n_2 sequence numbers, per turn of DNA of helical periodicity ν, assigned to $n = (n_2 + n_1)/2$, the position of the average sequence number of that DNA tract, and $\mathbf{d}_s$ is the local deviation of the sth base pair plane from the canonical B-DNA represented as a complex vector $\mathbf{d}_s = \rho_s - i\tau_s$ in terms of the roll (ρ) and tilt (τ) angles. These parameters were obtained theoretically by using conformational energy calculations for the 16 possible dinucleotide steps. In the energy calculations we have used our set of van der Waals functions and point charges (Remigopalakrishnan et al., 1971) and a variational method to find the local distortions from the uniform B-DNA. For each of 16 different dinucleotide double helix fragments from the standard B-DNA structure, we assumed the propeller twist for the AT and GC base pairs to correspond to the average values of 17° and 11°, respectively, as found in the crystal structures of double-helical oligonucleotides, ignoring the other small perturbations of the base paring (buckling and

---

[†] This work was supported by M.P.I. 40% and Progetto Finalizzato Biotecnologie C.N.R.

[*] Author to whom correspondence should be addressed.

[‡] Dipartimento di Chimica, Università di Roma "La Sapienza".

[§] Università della Basilicata.

[∥] Dipartimento di Genetica e Biologia Molecolare, Università di Roma "La Sapienza".

0006-2960/90/0429-9269$02.50/0

Chart I

ρ (deg)	A	T	G	C	τ (deg)	A	T	G	C	Ω (deg)	A	T	G	C
T	8.0	−5.4	6.7	2.0	T	0.0	0.5	−0.4	1.7	T	34.6	35.9	34.5	35.8
A	−5.4	−7.3	1.0	−2.4	A	−0.5	0.0	−1.6	−2.7	A	35.9	35.0	35.6	34.6
C	6.7	1.0	4.6	1.3	C	0.4	1.6	0.0	0.6	C	34.5	35.6	33.7	33.0
G	2.0	−2.4	1.3	−3.7	G	−1.7	2.7	−0.6	0.0	G	35.8	34.6	33.0	33.3

opening). The resulting structures were refined according to a multiple gradient method in order to select the van der Waals and electrostatic minimum energy structures with the stereochemical constraint of the correct base pair–phosphodiester–base pair–phosphodiester ring closure. A 0.25-Å increment was added to the van der Waals radius of the thymine methyl group to mimic hydrophobic interactions. The structural deviations of the base pairs were given in terms of differential tilt and roll of the base pairs for each of the 16 different dinucleotides with respect to their average values. The roll angles are in satisfactory agreement with the X-ray parameters obtained in double-helical dodecamers in spite of terminal and crystal packing effects (De Santis et al., 1988b; Cacchione et al., 1989). In agreement with the results obtained theoretically (Zhurkin, 1985; Tung & Harvey, 1986; Srinivasan et al., 1987) and empirically (Dickerson, 1983; Ulanovsky & Trifonov, 1987) the roll angles represent the main deviation from the B-DNA canonical structure. Such a representation of the curvature using the average ν^0 periodicity of DNA structure equal to 10.4 turned out in a good correlation with the experimental electrophoretic retardation factors of a number of multimeric periodical polynucleotides until then published.

Furthermore, the dispersion of the curvature, $\sigma^2 = \langle C(n)\cdot C(n)^*\rangle - \langle C(n)\rangle\cdot\langle C(n)\rangle^*$ [$C(n)$ = the orientational deviation between the nth and the first nucleotide local helical axes], was found in a good linear correlation with the electrophoretic retardation for different molecular weights, sequences, and periodicity of the multimers (Hagerman, 1986; Koo et al., 1986; Diekman, 1987; our unpublished data). The model worked successfully for natural DNA tracts in agreement with the electron microscopy visualization (Griffith et al., 1986) as well as with the permutation gel electrophoresis experiments (Wu & Crothers, 1984; Milton & Gesteland, 1988).

More recently, doubts on the validity of the nearest-neighbor models were raised (Haran & Crothers, 1989), in this journal, and cooperative effects were advanced as determinant factors of DNA curvature.

In this connection we have calculated the theoretical electrophoretic retardation of all the biosynthetic periodical polynucleotides so far published, but releasing the constancy of the twist previously adopted and allowing the roll and tilt angles to change until a maximum variation of 1°, in order to optimize the linear correlation with the experimental data. We have used 450 electrophoretic retardation data (8% acrylamide and room temperature) corresponding to 62 oligonucleotides different for sequence, periodicity, and complexity ranging between 50 and 200 bp (Koo & Crothers, 1986; Haran & Crothers, 1989; Cacchione et al., our unpublished data). In spite of different experimental conditions (12% acrylamide at 25 °C) the Hagerman data are also satisfactorily fitted probably because of a sort of compensation effect between a higher temperature than room temperature and the higher percentage of acrylamide in the gel.

The correlation diagram is shown in Figure 1 and corresponds to a correlation factor of 0.98. The roll, tilt, and twist angles are reported in degrees in the matrix form shown in Chart I. Note the symmetry of ρ and Ω and the antisymmetry of τ, which assure that the same curvature is obtained for the sequence of the one or the other strand (see the appendix) as logically expected. These values differ slightly from those previously published (De Santis et al., 1988b), but they represent an optimized semiempirical base to predict curvatures and electrophoretic behaviors for any DNA tracts. In any case, the range of variability of twist is very restricted with respect to the observed X-ray values around the average value of Ω = 34.7°. Generally, AT-rich tracts are characterized by values of twist ~2° higher than those of GC-rich tracts, which accounts, qualitatively, for the tendency of CC and GG dinucleotide steps to stabilize the A form of DNA as well as for the resistance of AT-rich DNA to adopting the A form. In any case, they represent variational parameters which mediate different approximation effects of the model.

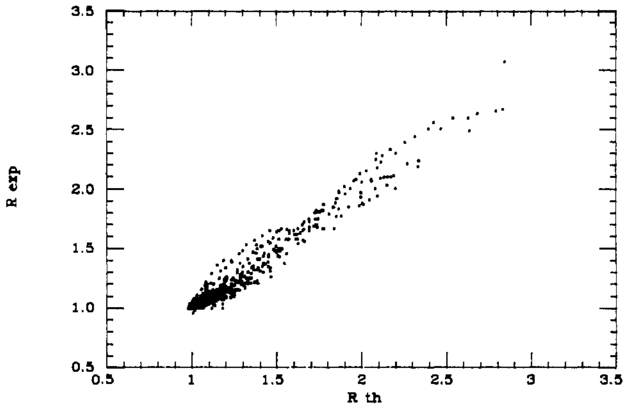

FIGURE 1: Experimental vs theoretical electrophoretic retardation factors for 450 synthetic multimeric oligonucleotides with 62 different sequences ranging between 50 and 200 bp. Rth = 0.97 + (1.56 × 10^{-4})σ^2 (correlation coefficient = 0.98)

In this connection, it is worthwhile to note that the periodical oligonucleotides represent a more severe test of the model with respect to the natural DNAs because of the sequence periodicity, which results in an amplification of the defects of approximation. Figure 2 shows, in fact, the practical coincidence of the modulus of curvature of the kinetoplast DNA averaged over tracts of 31 nucleotide steps recurring along the sequence, as calculated by using the previously published matrices (De Santis et al., 1986, 1988a) and the refined ones. The same result is obtained for the corresponding dispersion σ^2.

Figure 3 shows the comparison of the theoretical permutation gel electrophoresis diagrams (σ^2 is linearly related to the retardation factor) obtained with the old and the refined set of parameters. They are very similar and practically coincident with the experimental diagrams (Wu & Crothers, 1984). Other examples are reported in previous papers (De Santis et al., 1988b, 1989).

CONCLUSIONS

The results reported are a very critical convincing test of the validity of our theoretical model of DNA curvature; in spite of the complexity of the system and the variety of free energy contributions which contribute to the thermodynamic stability of curved DNA structures, it represents a surprisingly good tool for investigating the superstructures of DNA tracts and

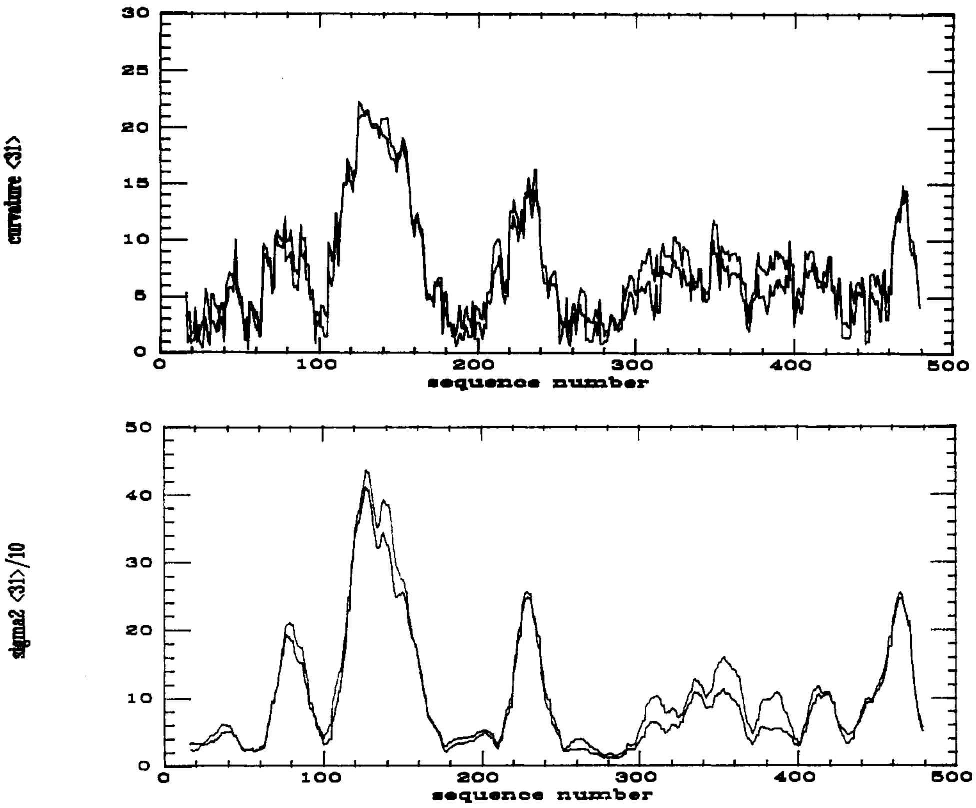

FIGURE 2: Comparison of curvature and σ^2 diagrams of a K-DNA fragment of 490 bp averaged on three turns of DNA (31 bp) as calculated by use of our old and the refined matrices for roll, tilt, and twist.

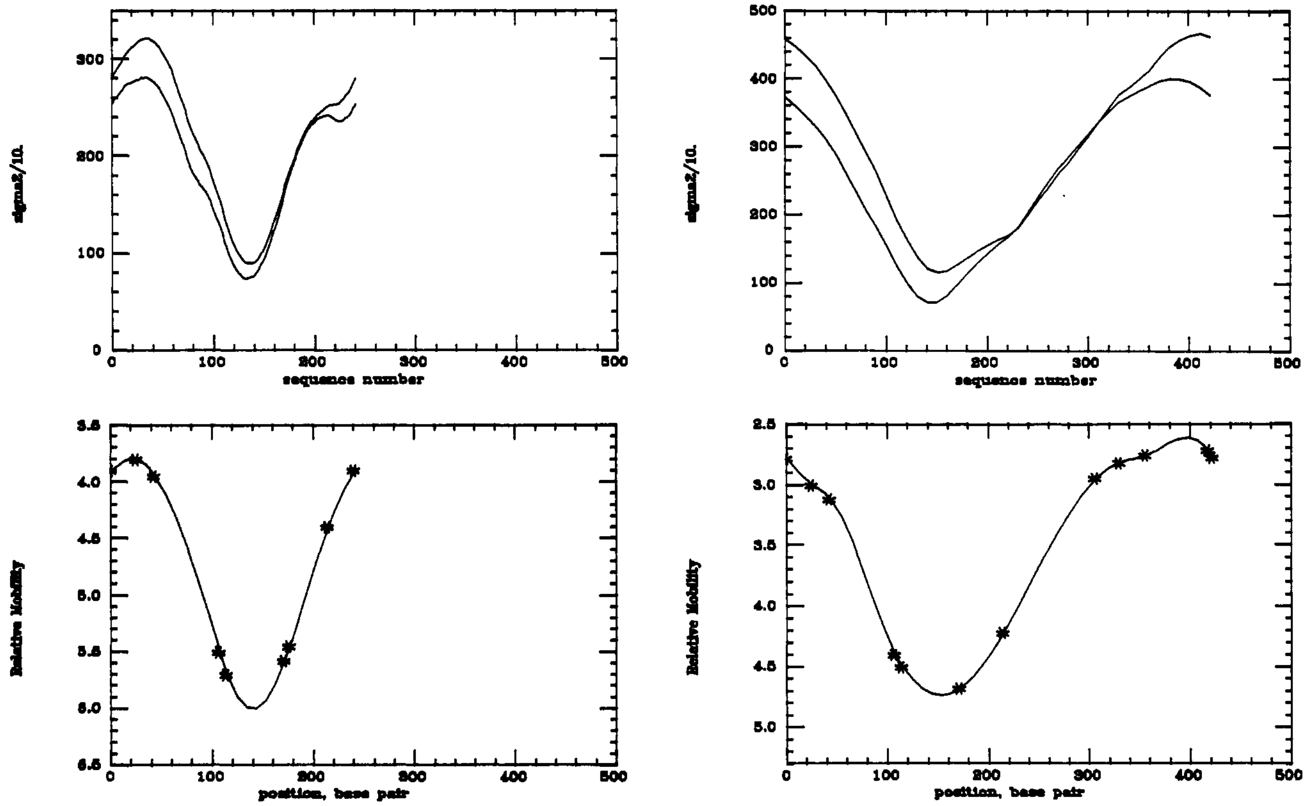

FIGURE 3: Comparisons between the theoretical gel electrophoresis permutation diagrams of two fragments of K-DNA (1–241 and 1–423) as calculated by use of our old and the refined matrices for roll, tilt, and twist. The diagrams below are the experimental gel electrophoresis as reprinted by Wu and Crothers (1984).

their biological relevance starting only from the sequence.

APPENDIX: DERIVATION AND PROPERTIES OF THE COMPLEX PLANE REPRESENTATION OF DNA CURVATURE

The Curvature Vector. The DNA curvature $\bar{C}(n)$ is a local property of B-DNA axis and represents its directional change along the sequence. It is defined as having a modulus equal to the reciprocal of the curvature radius of the writhing B-DNA axis assuming as a unitary step the average length of one turn of the double helix. This quantity can be calculated

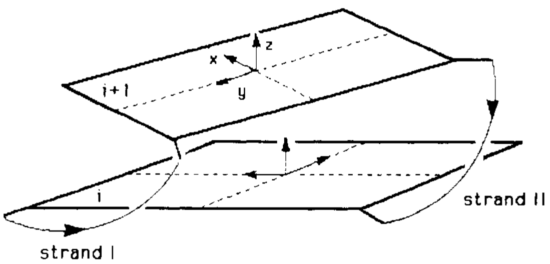

FIGURE 4: Scheme of the dinucleotide step following the recent international convention.

following the Cambridge convention (see Figure 4) by transforming the coordinate system fixed at the first base pair average plane of the chain as in B-DNA (with the z axis in the direction of the helical axis and the x axis along the pseudo dyad axis of the base pair) into that equivalently fixed at the first base pair of the next turn of double helix.

If $\mathbf{A}_s = \boldsymbol{\Omega}_s\mathbf{R}_s\mathbf{T}_s$ represents the transformation matrix of a coordinate system at $s - 1$ into one of sequence number s, corresponding to the product of the rotation matrices of argument Ω_s (the twist angle around the z axis), ρ_s (the roll angle around the y axis), and τ_s (the tilt angle around the x axis), the change in direction of the B-DNA axis after one nucleotide step is represented by the vector:

$$(\mathbf{A}_s{}^+ - \mathbf{U})\begin{vmatrix}0\\0\\1\end{vmatrix} \qquad (1)$$

Thus, the change in the helical axis direction between the sequence number n_1 and n_2 is given by

$$[(\prod_{s=n_1}^{n_2} \mathbf{A}_s{}^+) - \mathbf{U}]\begin{vmatrix}0\\0\\1\end{vmatrix} \qquad (2)$$

The low values of ρ and τ allow the reduction of the matrices involved at the first order of Taylor series. Thus, expressions 1 and 2 become respectively

$$\begin{vmatrix}\boldsymbol{\Omega}_s{}^+ - \mathbf{U} & \mathbf{d}_s\\ -(\boldsymbol{\Omega}_s\mathbf{d}_s)^+ & 0\end{vmatrix}\begin{vmatrix}0\\1\end{vmatrix} = \begin{vmatrix}\mathbf{d}_s\\0\end{vmatrix} \qquad (3a)$$

$$\begin{vmatrix}\prod_{s=n_1}^{n_2} \boldsymbol{\Omega}_s{}^+ - \mathbf{U} & \mathbf{C}(n)\\ -[\prod_{s=n_1}^{n_2} \boldsymbol{\Omega}_s\mathbf{C}(n)]^+ & 0\end{vmatrix}\begin{vmatrix}0\\1\end{vmatrix} = \begin{vmatrix}\mathbf{C}(n)\\0\end{vmatrix} \qquad (3b)$$

in supermatrix form; where $\mathbf{U}$ is the unitary matrix

$$\boldsymbol{\Omega}_s{}^+ = \begin{vmatrix}\cos\Omega_s & -\sin\Omega_s\\ \sin\Omega_s & \cos\Omega_s\end{vmatrix} \qquad \mathbf{d}_s = \begin{vmatrix}\rho_s\\ -\tau_s\end{vmatrix}$$

The curvature vector is defined in the plane perpendicular to the helical axis as

$$\mathbf{C}(n) = \mathbf{d}_{n1} + \boldsymbol{\Omega}_{n1}{}^+\mathbf{d}_{n1+1} + ... + \prod_{s=n_1}^{n_2-1} \boldsymbol{\Omega}_s{}^+\mathbf{d}_{n2}$$

and assigned to the central nucleotide of the tract considered $n = (n_1 + n_2)/2$. The deviation vector, $\mathbf{d}_s$, and the curvature vector, $\mathbf{C}(n)$, lie in the xy plane. This reduction is a very good approximation because the intrinsic curvature is generally less than 20–25° per turn.

The curvature vector of DNA is conveniently calculated as a vectorial summation in the complex plane by changing the x and y axes into the real and imaginary axes, respectively, and adopting an average value for $\boldsymbol{\Omega}_s$ corresponding to the periodicity of B-DNA, ν, in the tract considered. In such a representation:

$$\mathbf{d}_s = \rho_s - i\tau_s$$

and

$$\mathbf{C}(n) = \sum_{s=n_1}^{n_2} \mathbf{d}_s \exp(2\pi is/\nu)$$

A useful representation of the distortions of the B-DNA axis along the chain is given by the "curvature diagram", which gives the value of curvature per turn $\bar{\mathbf{C}}(n)$ in modulus and phase calculated for recurrent turns along the sequence. In order to reduce the noise and to better approximate the period of B-DNA ($\nu = 10.4$) in the tract of sequence considered, it is useful to average the curvature over a tract of 21 or 31 nucleotide steps. Thus, in general:

$$\bar{\mathbf{C}}(n) = \nu(n_1 - n_2)^{-1} \sum_{s=n_1}^{n_2} \mathbf{d}_s \exp(2\pi is/\nu)$$

In the case of tracts corresponding to complete turns of B-DNA, $\bar{\mathbf{C}}(n)$ is different from zero only when the distribution of $\mathbf{d}_s$ contains a harmonic component of frequency ν. It is worthwhile to note that this symmetry feature is true independently by the approximation adopted, namely, also when $\bar{\mathbf{C}}(n)$ is calculated by the standard matrix transformations.

It is interesting to note the logically expected invariance of the curvature to the interchange of the sequence considered with that of the other strand in the correct 5′–3′ direction, because the symmetry properties of the deviation vectors; it corresponds to a change in the $\bar{\mathbf{C}}(n)$ formula $\mathbf{d}_s$ with the complex conjugate $\mathbf{d}_s{}^*$ and the sequence number s with $N - s$ (N = total number of bp); so that the curvature of the a strand is the complex conjugate of the curvature of the other strand times a phase factor $\exp(2\pi iN/\nu)$. It is a consequence of the commutability of the roll and tilt matrices in their linearized form valid for low angles.

A more direct derivation of the complex plane representation springs out of the consideration that the vectorial chain of the local deviation vectors, rotated by the pertinent phase angle of B-DNA in the complex plane, corresponds to the vectorial chain on the unit sphere of the directional changes of the local helical axes.

Starting from the curvature diagrams, it is possible to reproduce the three-dimensional superstructure of DNA tracts with a very good approximation, using the product of the pairs of rotation matrices $|\bar{\mathbf{C}}(n)||\Phi(n)|$, where $|\bar{\mathbf{C}}(n)|$ represents the pure rotation around the x axis of an angle equal to the average curvature at the sequence number n, and $|\Phi(n)|$, the pure rotation matrix around the z axis of the difference in phase of the curvature vectors at n and $n + 1$ sequence numbers. The structure obtained is practically identical with that calculated via standard matrix transformations. This is identical with that calculated via standard matrix transformations This is illustrated in Figure 5 in the extreme case of the looping DNA tract of *Crithidia fasciculata.*

The same result is obtained for multimeric oligonucleotides; in these cases the sequence periodicity results in superhelical structures which are completely defined by the structure of the pertinent oligomeric unit, generally, a short fragment of double helix (9–21 bp) where the mathematical approximation adopted is quite appropriate. Therefore, the results obtained are quite adequate to the comparison with the available experimental data.

The Curvature Dispersion. It is interesting to evaluate the central angular dispersion, $\sigma^2 = \langle \mathbf{C}(n)\cdot\mathbf{C}(n)^*\rangle - \langle\mathbf{C}(n)\rangle\cdot\langle\mathbf{C}$-

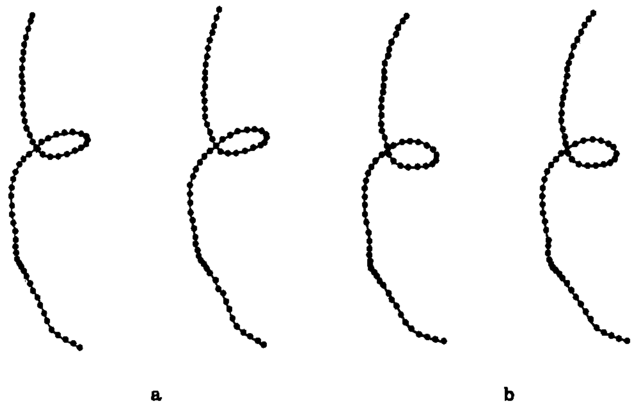

FIGURE 5: Stereoprojections of the superstructures of the looping mitochondrial DNA of *C. fasciculata*, 660 bp, as calculated by using (a) the curvature diagrams in the complex plane representation and (b) the standard transformation matrices.

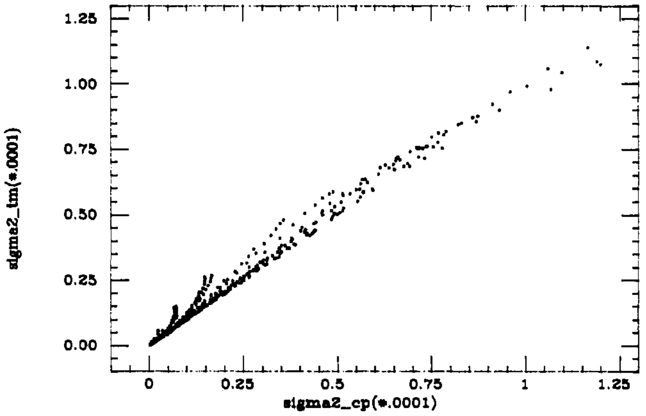

FIGURE 6: Correlation diagram of the curvature dispersion σ^2 as calculated by using the complex plane representation (sigma2_cp) and the transformation matrices (sigma2_tm) for 450 multimeric oligonucleotides of Figure 1; correlation factor $r = 0.996$.

$(n))^*$ (the mean values calculated over the N bp DNA tract examined), of the directions of the local helical axes in tracts of multimeric oligonucleotides as well as of DNAs. More explicitly

$$\sigma^2 = (1/N)\sum_{n=1}^{N}[[\sum_{s=1}^{n}\mathbf{d}_s \exp(2\pi is/\nu)] \times [\sum_{s=1}^{n}\mathbf{d}_s \exp(2\pi is/\nu)]^*] - [[(1/N) \times \sum_{n=1}^{N}\sum_{s=1}^{n}\mathbf{d}_s \exp(2\pi is/\nu)][(1/N)\sum_{s=1}^{N}\sum_{s=1}^{n}\mathbf{d}_s \exp(2\pi is/\nu)]^*]$$

In the former case it can be conveniently obtained with a very good approximation as the radius of gyration of the chain of the deviation vectors in the complex plane, as demonstrated by the very good correlation ($r = 0.996$) with the values of σ^2 calculated from the angular distributions on the unit sphere of the local helical axes using the pertinent matrix transformations (see Figure 6).

In the case of long tracts of DNAs as those analyzed to reproduce the permutation gel electrophoresis, the evaluation of σ^2 in the complex plane introduces minor errors for σ^2 values less than 4000 as in the majority of the cases examined. The position of the minimum is in any case practically invariant, as shown in Figure 7, where the tract of K-DNA (1–241) is analyzed according to both the complex plane and the transformation matrices methods.

This suggests that the relative mobility in polyacrylamide gel electrophoresis of curved polynucleotides with respect to the corresponding straight ones should be strictly related to the central dispersion of curvature, namely, to the central dispersion of the angular distribution of the local helical axes. It is interesting that in Hooke's harmonic approximation, assuming a elastic rod as a rough model of DNA, σ^2 is proportional to the minimum energy for straightening the double-helical axis. This activation energy increases the relaxation time of DNA fluctuations about the straight DNA conformations with a consequent retardation effect.

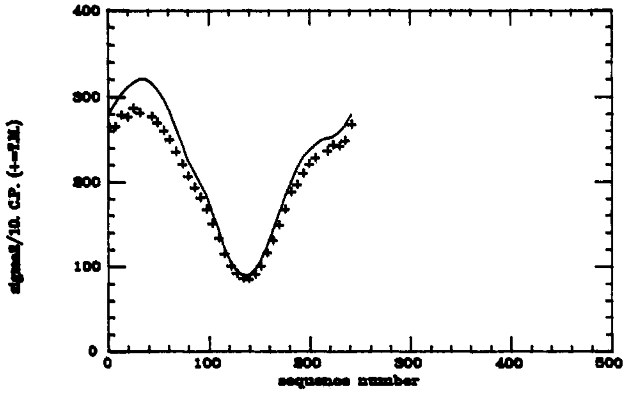

FIGURE 7: Comparison between the theoretically predicted permutation gel electrophoresis of tract 1–241 of K-DNA as calculated by using the complex plane representation (—) and the transformation matrices (+++).

REFERENCES

Cacchione, S., De Santis, P., Foti, D. P., Palleschi, A., & Savino, M. (1989) *Biochemistry 28*, 8706–8713.

De Santis, P., Morosetti,, S., Palleschi, A., & Savino, M. (1986) in *Structure and Dynamics of Nucleic Acids, Proteins and Membranes* (Clementi, E., & Chin, S., Eds.) pp 31–49, Plenum Publishing, New York.

De Santis, P., Morosetti, S., Palleschi, A., Savino, M., & Scipioni, A. (1988a) in *Biological Artificial Intelligence Systems* (Clementi, E., & Chin, S., Eds.) pp 143–161, ESCOM, Leiden.

De Santis, P., Palleschi, A., Savino, M., & Scipioni, A. (1988b) *Biophys. Chem. 32*, 305–317.

De Santis, P., Gallo, G., Palleschi, A., Savino, M., & Scipioni, A. (1989) *J. Mol. Liq. 41*, 291–304.

Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419–441.

Diekmann, S. (1987) *EMBO J. 6*, 4213–4217.

Griffith, J., Bleyman, M., Rauch, C. A., Kitchin, P. A., & Englund, P. T. (1986) *Cell 46*, 717–724.

Hagerman, P. S. (1986) *Nature 321*, 449–450.

Haran, T. E., & Crothers, D. M. (1989) *Biochemistry 28*, 2763–2767.

Koo, H.-S., & Crothers, D. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1763–1767.

Koo, H.-S., Wu, H.-M., & Crothers, D. M. (1986) *Nature 320*, 501–506.

Milton, D. L., & Gesteland, R. F. (1988) *Nucleic Acids Res. 16*, 3931–3949.

Remigopalakrishnan, V., Lakshminarayanan, A. V., & Sasisekharan, V. (1971) *Biopolymers 10*, 1159–1167.

Srinivasan, A. R., Torres, R., Clark, W., & Olson, W. K. (1987) *J. Biomol. Struct. Dyn. 5*, 459–496.

Trifonov, E. N. (1980) *Nucleic Acids Res. 8*, 4041–4053.

Tung, C. S., & Harvey, S. C. (1986) *Nucleic Acids Res. 14*, 381–387.

Ulanovsky, L. E., & Trifonov, E. N. (1987) *Nature 326*, 720–722.

Wu, H.-M., & Crothers, D. M. (1984) *Nature 308*, 509–513.