Periodical Polydeoxynucleotides and DNA Curvature†

Stefano Cacchione,‡ Pasquale De Santis,§ Dolly Foti,‡ Antonio Palleschi,‖ and Maria Savino*,‡

Dipartimento di Genetica e Biologia Molecolare, Centro per lo Studio degli Acidi Nucleici del CNR, and Dipartimento di Chimica, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, 00185 Roma, Italy, and Istituto di Chimica, Università della Basilicata, Via Nazario Sauro 85, 85100 Potenza, Italy

Received February 14, 1989; Revised Manuscript Received June 1, 1989

ABSTRACT: A theoretical method to predict DNA curvature was developed, and a strikingly good correlation between the experimental retardations and theoretical curvature of all the periodical biosynthetic DNAs so far reported in the literature was found. The analysis has been extended to G- and C-rich synthetic polynucleotides, which show a behavior in agreement with the theoretical prediction. A possible application of the method to biologically significant DNA tracts is shown in the case of the regulative region of one of the genes which code for the small subunit of ribulose-1,5-bisphosphate carboxylase in *Pisum sativum*. While curvature measurements have not so far been reported for this system, biochemical analysis has indicated short nucleotide sequences (boxes I–III) as recognition sites for regulative proteins. On the basis of the theoretical curvature profile of the region and of the electrophoretic retardation measurements of synthetic polynucleotides, obtained by ligating monomers mimicking the boxes, we suggest that the proteins could use DNA local curvature as structural motif in the recognition process.

The importance of DNA curvature is convincingly emerging in the last few years, on the basis of the hypothesis that the harmonic component of the base distribution observed in eukaryotic nucleotide sequences with the period of B DNA could induce DNA curvature (Trifonov, 1980).

This hypothesis received experimental support from the observation that some "Curved" DNAs migrate more slowly than normal DNA molecules having the same molecular weight on polyacrylamide gel electrophoresis (Marini et al., 1983). More recently, Koo et al. (1986) and Hagerman (1986) approached the problem of determining the relationship between nucleotide sequence and DNA curvature by chemical synthesis of appropriate oligonucleotides, which were then ligated in a unique polarity to form longer periodical DNA molecules. The electrophoretic gel mobility of these multimers, compared with a scale of fragments of known molecular weights, showed different retardation factors, namely, different ratios between the apparent and the real molecular weight.

Essentially, two kinds of hypothesis were advanced to explain the origin of DNA curvature.

The first associates the deviation from linearity of the helical axis to a structural transition that should occur between A stretches longer than three base pairs, which are assumed to adopt the H (heteronomous) DNA structure derived from X-ray diffraction analysis of poly(dA)·poly(dT), and the normal B DNA structure (Koo et al., 1986); this transition could also explain cooperative effects found by Diekmann and von Kitzing (1988) for oligonucleotides containing A_n tracts, if n is higher than three.

The second hypothesis explains the DNA curvature as a constructive summation of the small distortions from B DNA standard structure caused by differential interactions between the different dinucleotide steps along the double helix. We adopted this hypothesis, generally indicated as the nearest-neighbor model, and used an original method based on matrix transformations and the local distortion parameters for each of the 16 base pairs' stacking obtained theoretically from energy calculations (De Santis et al., 1986). We found that dinucleotide roll angles have a predominant role, and using these parameters, we were able to predict the different electrophoretic behavior of sequences such as CA_4, GA_4, CT_4, and GT_4 (Hagerman, 1986), which were not satisfactorily explained on the basis of the first nearest-neighbor, or "wedge", model (Trifonov & Sussman, 1980). However, more recently, Ulanovsky and Trifonov (1987) recognized as a main cause of DNA curvature the roll angle distortion that in their analysis was attributed only to the AA step.

On the basis of our method we were able to find a strikingly good correlation between the experimental retardations and the theoretical curvature of all the periodical biosynthetic DNAs than have been so far reported in the literature (De Santis et al., 1986, 1988a). Afterward the analysis was extended to many biologically relevant DNA tracts, which showed anomalous electrophoretic behavior. We were able to predict correctly (De Santis et al., 1988b) permutation gel electrophoresis experiments in the case of kinetoplast DNA (K-DNA) of the parasite *Leishmania tarentolae*, of the DNA of SV40 virus, of a fragment of *Xenopus* 5S gene, of the DNA binding domain of *Escherichia coli* CAP protein, and of two fragments of the plasmid pT181. This last example is particularly significant because, although the average base composition of pT181 is about 30% in G+C, the origin region contains about 50% G+C base pairs, showing that in some cases it is possible to find curvature also in the sequences richer in G+C. The relevance of roll angles of GC steps is in agreement with the results of X-ray structural analysis of double-helical dodecanucleotides, which contain a sufficient number of base pairs to be considered as models of B DNA. In fact the distortions from the canonical B DNA structure can be considered the effect of local deviations of many kinds of dinucleotide steps (Fratini et al., 1982; Nelson et al., 1987; Yoon et al., 1988). The same conclusion was reached on the

† This work was supported by MPI (60%, Grandi Progetti d'Ateneo, University of Rome I) and by Fondazione "Istituto Pasteur–Fondazione Cenci-Bolognetti". S.C. and D.F. were supported by ENI fellowships.

* Author to whom correspondence should be addressed.

‡ Dipartimento di Genetica e Biologia Molecolare, Centro per lo Studio degli Acidi Nucleici del CNR, Università di Roma "La Sapienza".

§ Dipartimento di Chimica, Università di Roma "La Sapienza".

‖ Istituto di Chimica, Università della Basilicata.

0006-2960/89/0428-8706$01.50/0

basis of conformational energy calculations by different authors (Zurkin et al., 1979; Srinivasan et al., 1987).

Therefore, in the present communication, as a complement of previously synthesized DNAs, sequences rich in G and C bases were investigated to obtain information about their role in DNA curvature.

Furthermore, as an example of the possible applications of our predictive method, we report the theoretical analysis of the curvature of a biologically relevant sequence, namely, the sequence conferring the property of photoactivation to one of the genes which code for the small subunit of the ribulose-1,5-bisphosphate carboxylase protein (rbcS) in pea plants (Green et al., 1987). The theoretical curvature profile was then compared with the electrophoretic properties of synthetic multimers of the decamers corresponding to some of the regions with higher curvature in the profile.

Materials and Methods

Oligodeoxyribonucleotides were synthesized on an automated DNA synthesizer and purified on a 20% polyacrylamide gel in the presence of 7 M urea, followed by high salt (0.1 M ammonium bicarbonate) elution from a Sephadex G-50 column. Oligomer sequences were verified by using the chemical sequencing technique of Maxam and Gilbert with minor modifications for short oligomers (Wu et al., 1984).

Polydeoxyribonucleotides were obtained according to the following procedure. Two micrograms of each oligomer was 5′-labeled with 5 μCi of [γ-^{32}P]ATP (5 μCi pmol^{-1}) in 10 μL of solution, at 37 °C for 15 min. The reaction solution contained 70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM DTT,[1] and 3 units of polynucleotide kinase. After labeling with [γ-^{32}P]ATP, 50 nmol of cold ATP and an additional 2 units of kinase were added, and the reaction volume was increased to 15 μL. Phosphorylating with cold ATP was continued in the same buffer and temperature as above, for a period of 1 h. Phosphorylated oligonucleotides were heated to 60 °C and slowly cooled to 0 °C to form hybrids. To 10 μL of the hybridized mixture was added 1 unit of T4 DNA ligase, and the reaction volume was adjusted to 20 μL; the ligation mixture had the same concentration of Tris-HCl, $MgCl_2$, and DTT as the phosphorylating step, plus 2.3 mM cold ATP. The ligation reaction was allowed to proceed on ice overnight, after which it was quenched by addition of EDTA (pH 8) to 25 mM final concentration.

Ligated products were run on nondenaturing 10% polyacrylamide gels (mono:bis acrylamide ratio = 29:1; 90 mM Tris–borate, 2.5 mM EDTA, pH 8.3) until the bromophenol blue dye had migrated 25 or 50 cm. The applied voltage was 4 V cm^{-1}. Electrophoretic experiments were carried out at room temperature.

Results

A Theoretical Method To Predict the Curvature and the Electrophoretic Retardation Factor of Natural and Synthetic Nucleotide Sequences. We have previously shown that DNA curvature can be predicted by integrating the slight local deviations from the canonical B DNA given in terms of roll (De Santis et al., 1986, 1988a) and tilt (De Santis et al., 1988b) angles, obtained on the basis of energy conformational calculations for the sixteen dinucleotide steps and shown in matrix form in Chart I.[2]

Chart I

roll (rad)	A	T	G	C	tilt (rad)	A	T	G	C
T	0.16	−0.09	0.12	0.04	T	0.00	0.00	−0.01	0.03
A	−0.09	−0.12	−0.02	−0.04	A	0.00	0.00	−0.03	−0.03
C	0.12	−0.02	0.10	0.01	C	0.01	0.03	0.00	0.02
G	0.04	−0.04	0.01	−0.07	G	−0.03	0.03	−0.02	0.00

Table I: Repeating Sequences of Multimers Reported in Figure 1

	name	sequence (5′ → 3′)
1	A_5N_4	CAAAAACGG[a]
2	CA_4	CAAAATTTTG[b]
3	CT_4	CTTTTAAAAG[b]
4	GA_4	GAAAATTTTC[b]
5	GT_4	GTTTTAAAAC[b]
6	G_3TC	GGGTCGACCC[b]
7	IAT	GGCAATAACG[a]
8	A_4N_6	GGCCAAAACG[a]
9	A_9N_1	CAAAAAAAAA[a]
10	A_8N_2	CCAAAAAAAA[a]
11	FGG	CCGAAAAAGG[a]
12	A_6N_4	GGCAAAAAAC[a]
13	A_5N_5	GGCAAAAACG[a]
14	FCT	GGCAAAAATG[a]
15	IAC	GGCAACAACG[a]
16	IAG	GGCAAGAACG[a]
17	A_3N_7	GGCCAAACCG[a]
18	*Eco*RI linker	CGGAATTCCG[c]
19	G_2A_3	GGAAATTTCC[d]
20	AC_4	ACCCCGGGGT
21	AG_4	AGGGGCCCCT
22	TG_4	TGGGGCCCCA
23	TC_4	TCCCCGGGGA
24	A_3CT	AAACTAGTTT
25	TTCGC	TTCGCGCGAA
26	rbcS-2	ATCATTTTCA
27	rbcS-3	TGCAAACTTT
28	rbcS-4	AATTTCAAAT
29	rbcS-6	GTGTGGTTAA
30	A_5N_6	GGCCAAAAACG[a]
31	A_5N_7	GGCCAAAAACCG[a]
32	A_5N_{10}	CCGGCAAAAACGGGC[a]
33	A_6T_6	AAAACGGGTTTTTTGGGCAA[a]
34	A_{8-5}	CCAAAAAAAACGGGCAAAAA[a]
35	A_{5-8}	CCAAAAACGGGCAAAAAAAA[a]
36	rbcS-1	TGCAAACTTTATCATTTTCA
37		CGGAATTCCGCGGAATTCCGG[c]

[a] Koo et al., 1986. [b] Hagerman, 1986. [c] Diekmann, 1987. [d] Hagerman, 1985.

The curvature vector $\bar{C}$ per turn of DNA can be conveniently given as a complex function of the sequence with the modulus representing the deviation and the phase indicating the relative direction:

$$\mathbf{C} = \nu^0(n_2 - n_1)^{-1} \sum_{j=n_1}^{n_2} (\rho_j - i\tau_j) \exp(2\pi ij/\nu^0)$$

where ν^0 is the double-helix average periodicity that we assume equal to 10.4 according to the experimental evidence; $n_2 - n_1$ (n_2 and n_1 sequence numbers) represents the integration step that should be chosen near to a multiple of 10 in order to minimize the signal to noise ratio. In the particular case of periodical polynucleotides it is convenient that the integration step coincides with the corresponding oligomeric unit so that $\bar{C}$ becomes a constant. Electrophoretic retardation factors of the synthetic periodical polynucleotides analyzed by different authors correlate very satisfactorily with the modulus of curvature so evaluated, but by fitting different curves for different molecular weights. We have, however, shown that all the representative points of the electrophoretic retardations

[1] Abbreviations: DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; RUBISCO, ribulose-1,5-bisphosphate carboxylase; rbcS, small subunit of ribulose-1,5-bisphosphate carboxylase; bp, base pairs.

[2] The signs of the roll (ρ) and tilt (τ) angles are in agreement with the recently established nomenclature (Dickerson et al., 1989).

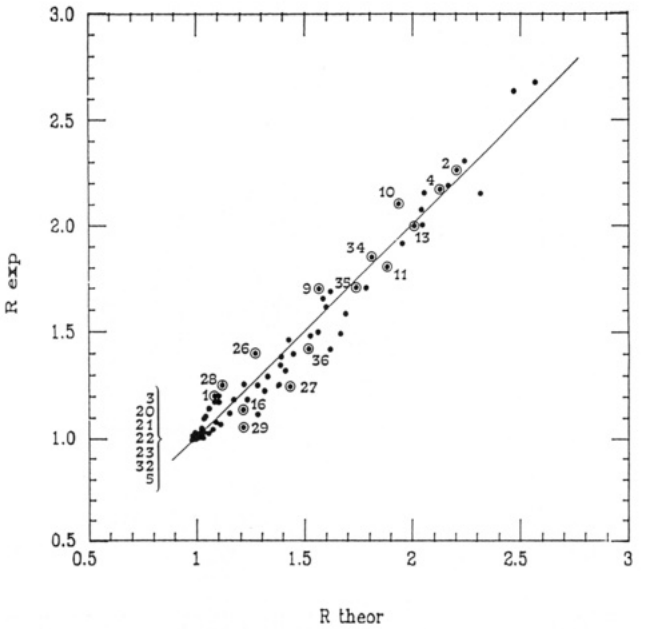

FIGURE 1: Linear correlation between the gel electrophoresis retardation (R_{exp}) and the theoretical retardation (R_{theor}). The diagram reports polynucleotides listed in Table I, different either for the sequence or for the molecular weight (N = 100, 150, 200). The labels correspond to N = 150.

are in good linear correlation (=0.99) with the dispersion (or variance) of the curvature vectors (De Santis et al., 1988b):

$$\sigma^2 = \langle \mathbf{C}|\mathbf{C} \rangle - \langle \mathbf{C} \rangle^2$$

The correlation diagram between the experimental and theoretical retardation relative to 37 periodical polynucleotides, reported in Table I, with different periodicity (9, 10, 11, 12, 15, 20, 21) and molecular weight (N = 100, 150, 200) is shown in Figure 1. While the correlation is generally very satisfactory, it is higher in the case of more curved sequences. There could be two causes for this behavior; first, it is more difficult to measure precisely electrophoretic gel retardation in the case of sequences having low curvature; second, in this range the influence of small differences between sequence periodicity of the fragments and their actual helical twist could be larger.

Theoretical and Experimental Analysis of the Curvature of Guanine- and Cytosine-Rich Synthetic Oligonucleotides. Decanucleotides of six different sequences were ligated to multimer distributions. For comparison with the electrophoretic behavior of periodical polynucleotides previously reported (Koo et al., 1986; Hagerman, 1986) the sequences A_5N_5 and G_3TC were also synthesized and ligated. In fact, it was previously shown that sequence G_3TC presented a normal electrophoretic behavior, whereas the sequence A_5N_5 showed a large retardation that was interpreted as due to large curvature. The sequences AC_4, AG_4, TC_4, TG_4, and A_3CT correspond to the palindromic sequences GT_4, GA_4, CT_4, CA_4, and G_3TC previously studied by Hagerman; in order to obtain evidence on the role of G and C bases on DNA curvature the sequences were planned with the interchange of A with G and T with C. The sequence TTCGC corresponds to the dodecamer CGCGAATTCGCG (the restriction site of *Eco*RI) whose structure was determined by Dickerson and found to be curved (Fratini et al., 1982).

The electrophoretic migrations of multimer distributions were determined in 10% polyacrylamide gels, and in all cases, since the repeat length of the oligomer sequence is close to the B DNA periodicity (ν^0 = 10.4), this technique allows sensitive

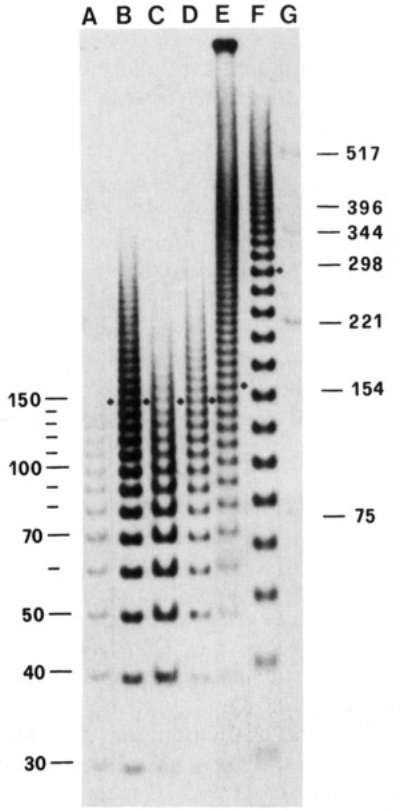

FIGURE 2: Autoradiogram of the multimer series of G- and C-rich polynucleotides electrophoresed on a nondenaturing 10% polyacrylamide gel. Asterisks indicate multimers of 150-bp sequence length. The ligated products were run until the bromophenol blue dye had migrated 25 cm. (A) G_3TC; (B) TC_4; (C) TG_4; (D) AC_4; (E) A_3CT; (F) A_5N_5; (G) pBR322–*Hin*fI digest.

measurement of the local curvature of the multimer helix axis. A typical polyacrylamide gel is shown in Figure 2. By comparison of the mobility of the different multimers with that of multimers of sequence G_3TC which shows a normal electrophoretic behavior, and also with the mobility of a pBR322–*Hin*fI digest, the apparent length of every multimer can be determined.

The migrations obtained from this and similar gels allow us to evaluate the ratio (R) of the apparent number of base pairs and the true number versus the number of base pairs N.

The R values of the multimers corresponding to N = 15 are as follows: R is equal to 1 in the case of AC_4, AG_4, TC_4, and TG_4, and to 1.05 for A_3CT. In other words, these polynucleotides show a normal electrophoretic behavior in full agreement with theoretical prediction. The value, R, of the sequence A_5N_5 at N = 15 is 1.97, in fairly good agreement with the retardation values previously obtained by Koo et al. (1986). The obtained results are also reported in Figure 1. It is worth noting that the lack of curvature in the examined sequences derives mainly from the fact that GG and CC dinucleotides have a very low roll angle, namely, ρ = 0.01, in comparison with AA and TT dinucleotides, which have a roll angle of −0.09. Only the sequence TTCGC shows a substantial retardation factor; i.e., R_{15} is equal to 1.38, while theoretically evaluated R is equal to 1.36; this behavior is due to the dinucleotides CG and GC, which have roll angles of 0.10 and −0.07, respectively.

The reported results and the numerous examples of good agreement between theoretical and experimental curvature quoted in reference (De Santis et al., 1988b) allow us to use the theoretical method in the case of an interesting biological system for which experimental analysis of DNA curvature has

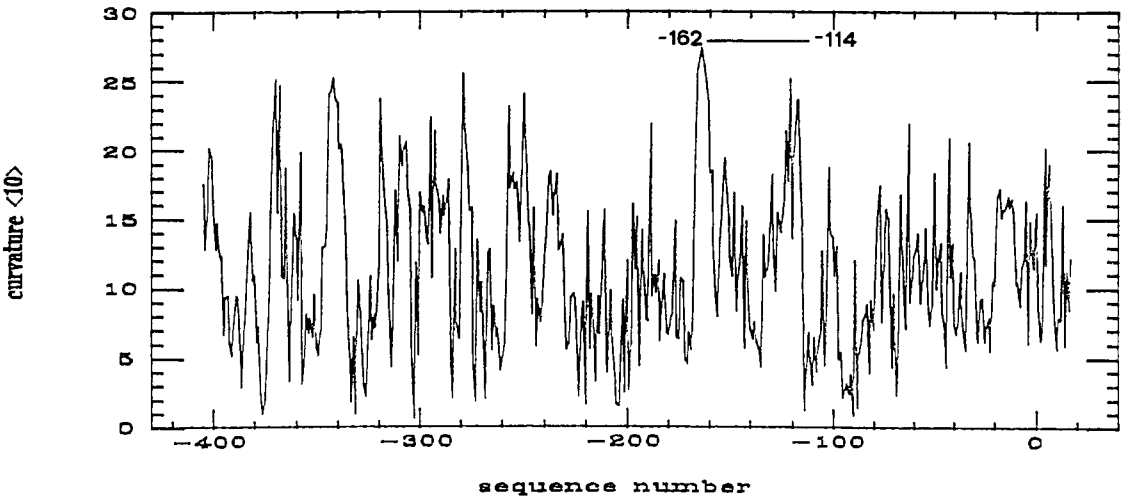

FIGURE 3: Theoretical curvature profile of the regulatory region from −410 to +15 of the rbcS-3A gene. The curvature is reported as $|c|$, the curvature modulus, averaged over 10 nucleotides versus the sequence, n. The region containing the three boxes derived from biochemical analysis, between −162 and −114, is indicated by the bar.

not so far been reported.

Theoretical Analysis of the Curvature of the Regulatory Region of the Ribulose-1,5-bisphosphate Carboxylase 3A Gene. The regulatory region of one of the genes which code for the small subunit of the ribulose-1,5-bisphosphate carboxylase (RUBISCO), the rbcS-3A gene, was chosen.

The small subunit of RUBISCO is encoded in higher plants by the rbcS nuclear gene family (Fluhr et al., 1986a; Pichersky et al., 1986); the expression of these genes is controlled by both tissue-specific factors and light (Fluhr et al., 1986b; Simpson et al., 1986). RbcS upstream sequences as far as about 400 bp from the transcription start point have been shown to be relevant for light-regulated expression in transgenic plants. Within this DNA region are contained sequence elements (designated box I, box II, and box III), between −162 and −114 which are conserved among all known pea rbcS genes (Kuhlemeier et al., 1987). At present, the opinion that well-defined sequences which demonstrate in vivo regulatory function are likely to act as targets for the binding of protein factors seems shared by numerous researchers (Sassone-Corsi & Borrelli, 1986). In fact, recently, Green et al. (1987), identified, using gel retardation assays and DNase I footprinting experiments, protein factors that specifically interact with the regulatory sequence upstream of the pea rbcS-3A gene, or, more precisely, with the sequence between −162 and −114. On the basis of the hypothesis that DNA curvature could be the structural motif recognized by regulatory proteins, we have carried out, using the theoretical method reported in the previous sections, the analysis of the curvature of the DNA sequence from −410 to +15 upstream of the rbcS-3A gene.

The profile is reported in Figure 3. The modulus of DNA curvature averaged over 10 nucleotides appears very complex; however, in the region from −162 and −114, corresponding to the sequences identified by biochemical analysis as regulatory proteins targets, two peaks clearly emerge.

To obtain experimental support to the theoretical curvature profile, we have adopted a method based on the synthesis of oligonucleotides, with sequences mimicking those of the relevant DNA regions; we used as an additional constraint the periodicity of B DNA; that is, the length of the monomers was chosen equal to 10 bp or a multiple. After synthesis, the monomers were ligated by using T4 ligase and the mobilities of polynucleotides were measured on polyacrylamide gel electrophoresis. The retardation of the multimers allows the evaluation of DNA local curvature, as described in a previous paper (De Santis et al., 1988b).

Table II: Sequences Relative to the Upstream Region −170/−111 of the rbcS-3A Gene

−170

ACACAAAATTTCAAATCTTGTGTGGTTAATATG-

box I box II

−111

GCTGCAAACTTTATCATTTTCACTATC

box III

	name	sequence (5′ → 3′)
box III	rbcS-1	TGCAAACTTTATCATTTTCA
	rbcS-2	ATCATTTTCA
	rbcS-3	TGCAAACTTT
box I	rbcS-4	AATTTCAAAT
box II	rbcS-5	AATCTTGTGT
	rbcS-6	GTGTGGTTAA

Electrophoretic Analysis of the Mobility of Multimers, Models of Elements of the Theoretical Curvature Profile. The synthesized decanucleotides and one 20-mer are reported in Table II and correspond to the two peaks and to the sequence in between. The gel electrophoretic mobilities of the reported monomers were measured after ligation up to at least 15 units.

Figure 4a shows a typical polyacrylamide gel of the multimers corresponding to the sequences reported in Table II.

Figure 4b shows a similar gel electrophoresis of the same multimers, except that the samples were run for about twice as long as in Figure 4a. This allows us to measure the mobilities of multimers ligated up to 40 monomers.

In both the autoradiographies the retardation factors of the sequences corresponding to box I and box III are clearly evident, while the decamer corresponding to box II shows an almost normal behavior. The values of the ratio between the apparent and the real number of base pairs, R, as a function of number of base pairs, N, are reported in Figure 5.

In all cases the trend corresponds to a quick increase of R as a function of N to reach a maximum, in the range 20–25, and then to a slow decrease. As an example, Figure 6 shows, in the case of A_5N_5, the theoretical retardation, R, versus N; it is worth noting the presence of a maximum followed by a slow decrease, in good agreement with experimental results.

The sequences rbcS-1, rbcS-2, and rbcS-3 that correspond to the region of box III show a high curvature; in addition, the larger retardations of the multimers of rbcS-2 with respect to those of rbcS-1 fit well the profile of the peak. The re-

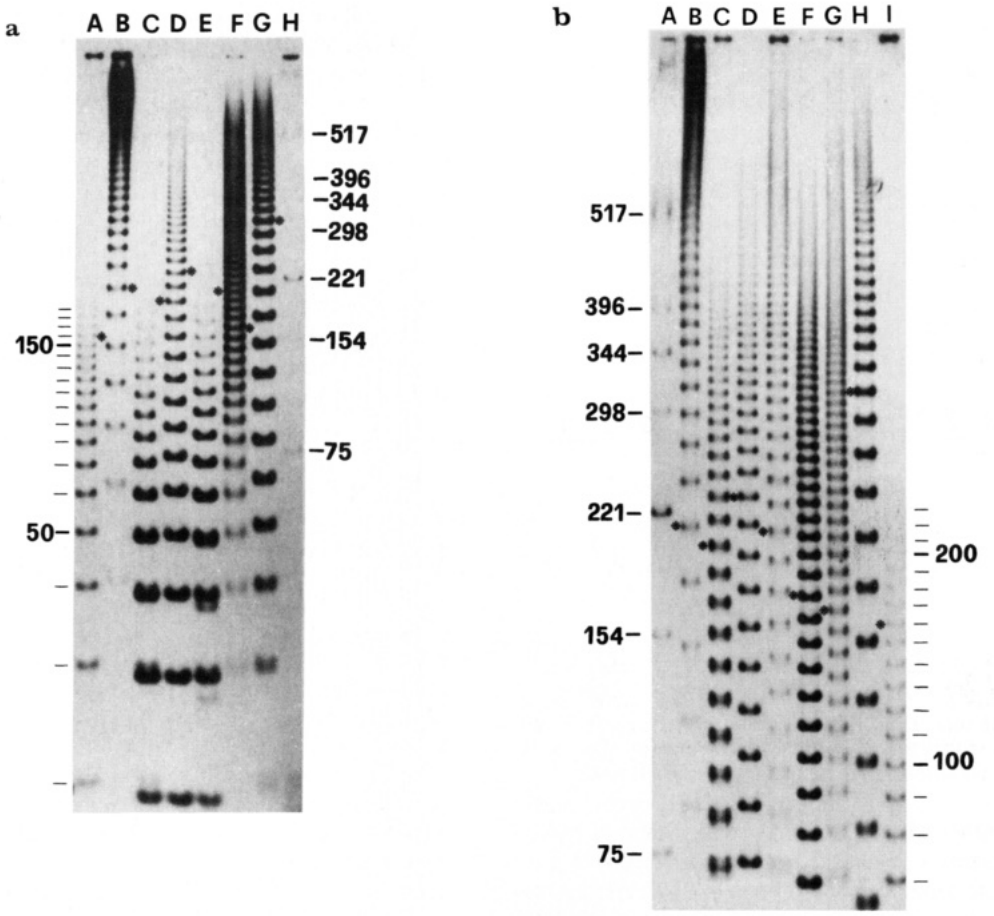

FIGURE 4: Autoradiograms of the multimer series (derived from ligation of the monomers reported in Table II) electrophoresed on a nondenaturing 10% polyacrylamide gel. Asterisks indicate multimers of 160-bp sequence length. (a) The ligated products were run until the bromophenol blue dye had migrated 25 cm. (A) G_3TC; (B) rbcS-1; (C) rbcS-3; (D) rbcS-2; (E) rbcS-4; (F) rbcS-6; (G) A_5N_5; (H) pBR322–*Hin*fI digest. (b) The ligated products were run until the bromophenol blue dye had migrated 50 cm. (A) pBR322–*Hin*fI digest; (B) rbcS-1; (C) rbcS-3; (D) rbcS-2; (E) rbcS-4; (F) rbcS-5; (G) rbcS-6; (H) A_5N_5; (E) G_3TC.

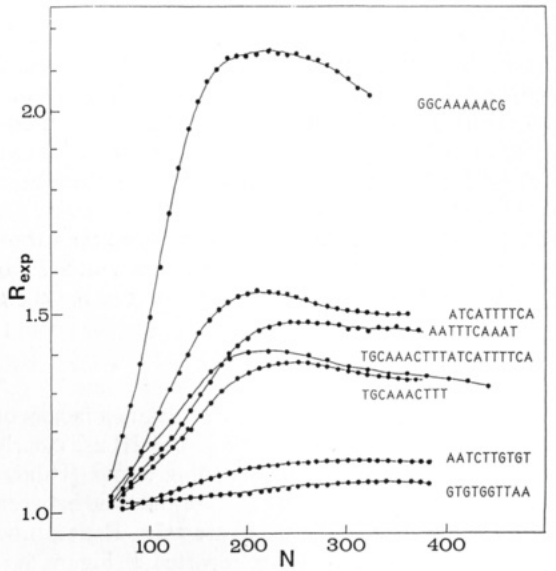

FIGURE 5: Mobilities of duplex polymers derived from oligomers reported in Table II and the 10-mer A_5N_5, represented as R, are plotted as function of N. R is determined from comparison with the electrophoresis markers.

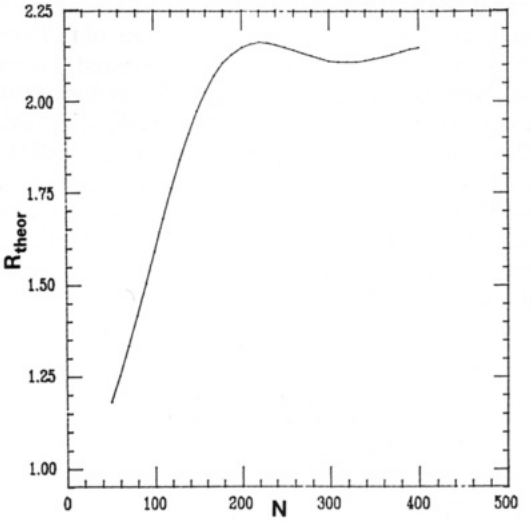

FIGURE 6: Trend of R_{theor} versus N in the case of A_5N_5. The twist number is equal to 10.4.

tardation of the multimers of the sequence rbcS-4, corresponding to the region of box I, similarly shows a noticeable retardation effect; in addition, their behavior shows a sort of biphasic trend, which is now under investigation. In full agreement with the theoretical prediction is also the small R corresponding to the sequences rbcS-5 and rbcS-6. In fact, in the theoretical profile, reported in Figure 3, these sequences correspond to a region of lower curvature.

Discussion

The nearest-neighbor theoretical model described in this paper, while starting from an approximate picture of the plausibly more complex factors which give rise to the intrinsic DNA curvature, nevertheless appears to be surprisingly quite suitable for translating the base sequence in superstructural

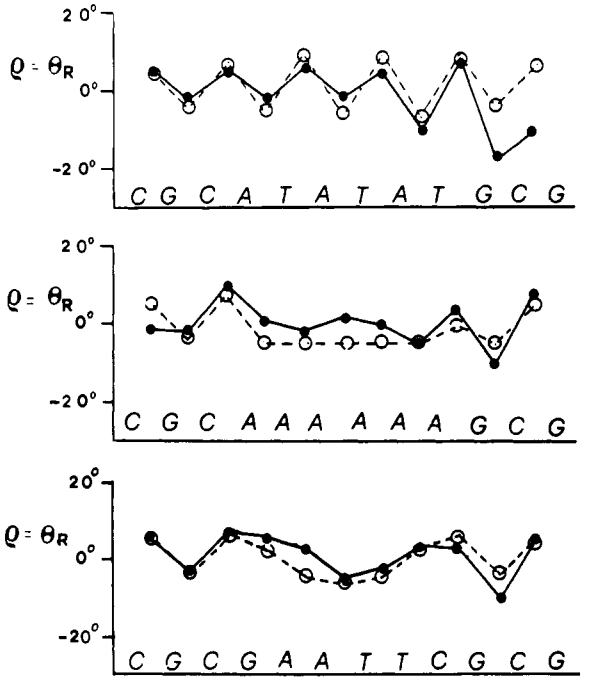

FIGURE 7: Comparison of the roll angles, ρ, obtained by conformational analysis (O), with those obtained by X-ray diffraction analysis (●), for the reported sequences. From top to bottom: (a) Yoon et al. (1988); (b) Nelson et al. (1987); (c) Fratini et al. (1982).

elements in synthetic as well as natural DNAs.

In fact, it explains with good accuracy the gel electrophoresis experiments that depend on the global superstructure of the considered sequence, as well as it satisfactorily fits the local deviations from the standard idealized B DNA structure, observed in double-helical oligonucleotides investigated by X-ray single-crystal analysis (Fratini et al., 1982; Nelson et al., 1987; Yoon et al., 1988); this agreement is shown in Figure 7, where the theoretical and experimental roll angles (De Santis et al., 1988b) of the dinucleotide steps of three different double-helical oligonucleotides are compared. It must be remembered that distortions from the B DNA structure were recently evidenced in synthetic oligonucleotides by NMR studies (Sarma et al., 1987, 1988). While the NMR analysis and our conformational analysis agree on the greater importance of the AT step, however, the dynamic effects near the terminals could influence the local distortions in NMR, which seem different with respect to those obtained from X-ray analysis.

An additional evidence of the influence of all the 16 dinucleotide steps on DNA curvature is the presence of a large curvature in the region corresponding to the replication origin of the plasmid pT181 (Koepsel & Khan, 1985). Using our matrices, it is possible to predict correctly the gel permutation analysis on fragments of this plasmid (De Santis et al., 1988b) and to find large curvature in a DNA sequence lacking phased A tracts. In Figure 8 the curvature profile averaged on 21 bp from the sequence 800–1000 is reported.

To suggest possible applications of our method, we have also tried to obtain experimental evidence of DNA local curvature of biologically important DNA tracts, with respect to the theoretical curvature profile. To this purpose, we have synthesized oligonucleotides 10 or 20 base pairs long, with sequences equal to those of peaks of the theoretical curvature profile or to regions of low curvature, and have measured their electrophoretic mobility after ligation up to at least 15 monomers. This approach multiplies DNA local curvature and allows us to obtain experimental evidence of local curvature by the measurement of a global property such as electrophoretic mobility.

The chosen system was the regulatory region of the rbcS-3A gene, since this region was extensively studied by biochemical analyses and shown to be the target for specific protein binding. The agreement between theoretical and experimental analysis seems very satisfactory; also, if it must be taken into account that to obtain a full agreement between theoretical and experimental retardation factors, the synthesized sequences must be chosen taking care that the two dinucleotide steps, resulting after ligation, do not sensibly change the curvature when different from those present in the natural DNA. The curvature characteristics of the three boxes connected with the protein binding seem satisfactorily proved by the reported results and prompted us to ask whether the regulative regions of different genes of the multigenic rbcS family could present different curvature.

We have considered the regulative region of the three genes: rbcS-3A (Green et al., 1987), rbcS-E9 (Coruzzi et al., 1984), and rbcS-SS 3.6 (Herrera-Estrella et al., 1984), which are characterized by the fact that the transcription of the first can be estimated to be 40%, with respect to the 7% of the last two together (Fluhr et al., 1986a). The theoretical curvature profiles of the three sequences are reported in panels a, b, and c, respectively, of Figure 9.

The analysis has been carried out by using the parameter σ^2, the curvature dispersion. The overall characteristics of the curvature profile are the same as in the case of the curvature

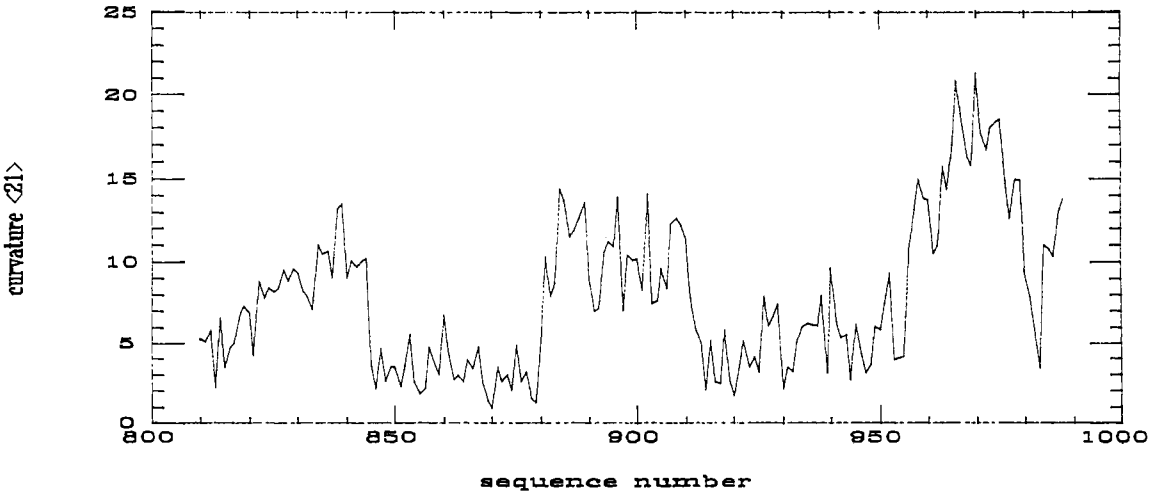

FIGURE 8: Theoretical curvature profile of a DNA fragment of the plasmid pT181, containing the origin region.

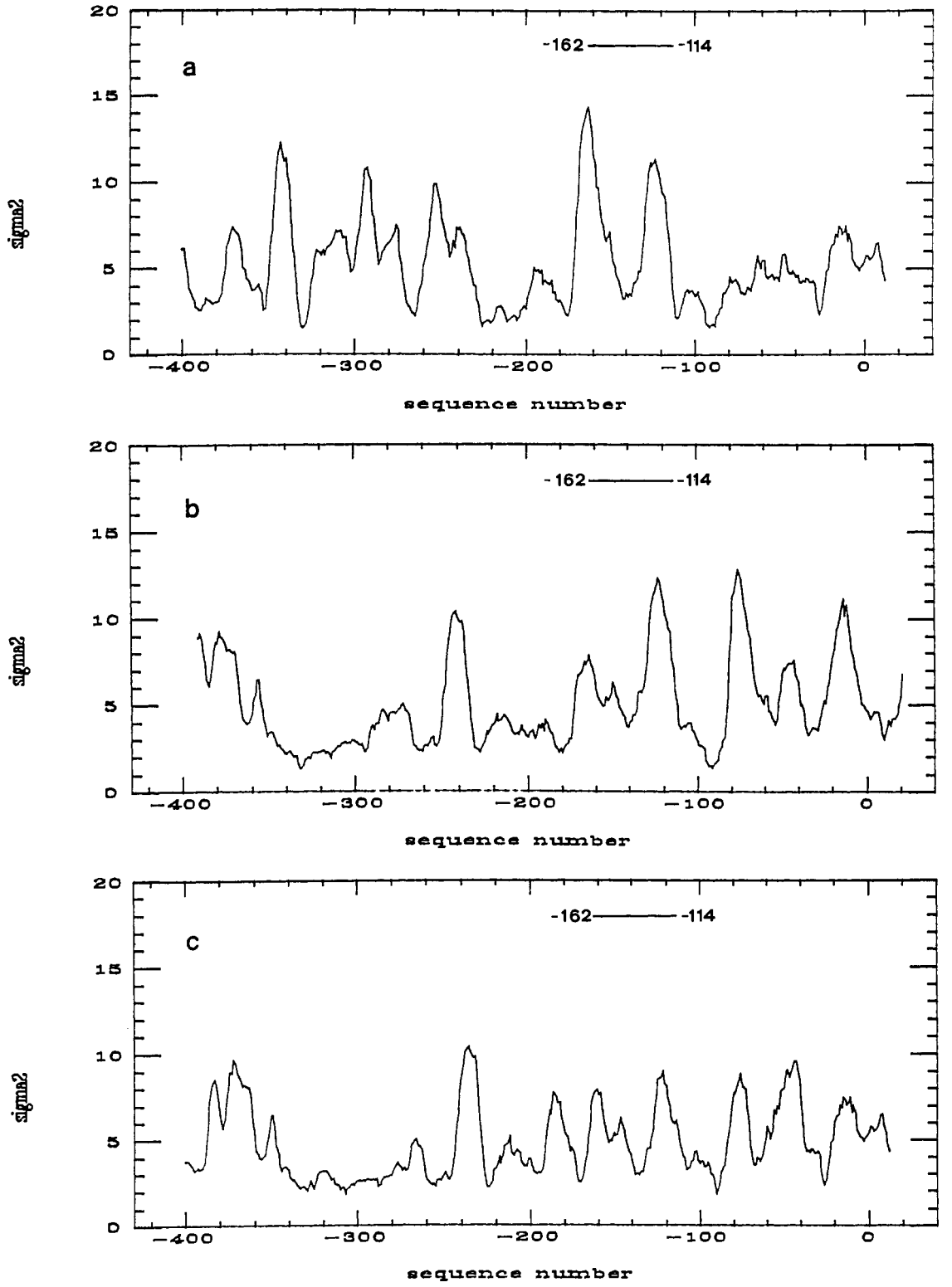

FIGURE 9: Theoretical curvature profiles of the regulatory region from −410 to +15 (from the start of transcription) in the case of (a) rbcS-3A, (b) rbcS-E9, and (c) rbcS-SS 3.6. The curvature is reported as σ^2, the curvature dispersion. The regions between −162 and −114, containing the three boxes, are indicated by bars.

modulus, but σ^2 more easily allows the individuation of the sequences of higher curvature, since with this method the noise, represented by the fluctuations of the curvature along the sequence, is cancelled.

The comparison between the curvature profiles of the sequences between −162 and −114, which in the three genes share a large chemical homology, shows interesting features. Specifically, rbcS-3A presents, as previously discussed, two peaks at about −160 and −120, while rbcS-E9 shows only a peak at −120 and rbcS-SS 3.6 lacks defined peaks in the same DNA region. These results suggest that if the sequences between −162 and −114 are the sites for specific binding in all the examined genes, the curvature of B DNA could be the structural motif recognized by the regulatory proteins.

However, at present, the relationship between biological function of defined nucleotide sequences and their curvature is only a stimulating research hypothesis. The behavior of the biological system is surely more complex. In fact, aside from the changes in the local curvature, the regulatory regions of the three genes are characterized by a different global cur-

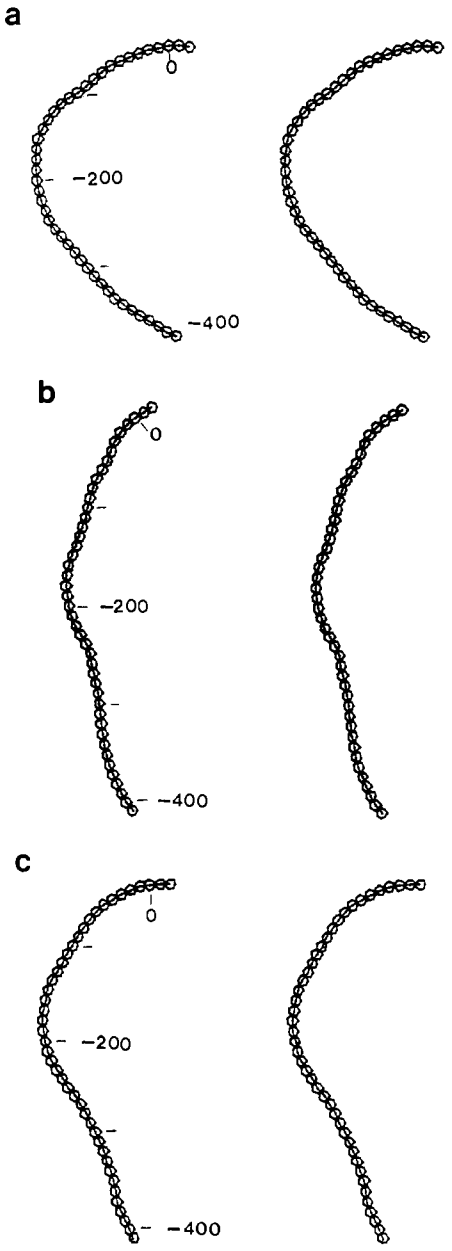

FIGURE 10: Stereoviews of the regulatory region from −410 to +15 in the case of (a) rbcS-3A, (b) rbcS-E9, and (c) rbcS-SS 3.6.

vature as shown in panels a, b, and c of Figure 10, where the stereoviews of the three regions between −410 and +15 are reported.

ACKNOWLEDGMENTS

Thanks are due to G. Gallo and A. Tufillaro for clever technical aid and G. Tecce for his interest and encouragement in the research project.

Registry No. **1**, 102432-35-9; **2**, 122799-65-9; **3**, 122799-66-0; **4**, 114094-49-4; **5**, 116054-40-1; **6**, 122799-67-1; **7**, 102432-41-7; **8**, 102432-45-1; **9**, 102432-48-4; **10**, 102432-47-3; **11**, 102432-43-9; **12**, 102432-46-2; **13**, 102432-36-0; **14**, 102432-42-8; **15**, 102432-39-3; **16**, 102432-40-6; **17**, 102432-44-0; **18**, 122762-11-2; **19**, 117940-93-9; **20**, 100186-85-4; **21**, 115259-34-2; **22**, 122762-12-3; **23**, 122762-13-4; **24**, 122762-14-5; **25**, 122762-15-6; **26**, 122762-16-7; **27**, 122762-17-8; **28**, 122762-18-9; **29**, 122762-19-0; **30**, 102432-37-1; **31**, 102432-38-2; **32**, 102426-47-1; **33**, 102426-55-1; **34**, 102426-54-0; **35**, 102426-53-9; **36**, 122873-79-4; **37**, 113956-59-5; rbcS-5, 122762-20-3.

REFERENCES

Coruzzi, G., Broglie, R., Edwards, C., & Chua, N.-H. (1984) *EMBO J. 3*, 1671–1679.

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 Corp., New York.

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

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

Dickerson, R. E., et al. (1989) *Nucleic Acids Res. 17*, 1797–1803.

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

Diekmann, S., & von Kitzing, E. (1988) in *DNA Bending & Curvature* (Olson, W. K., Sarma, M. H., Sarma, R. H., & Sundaralingam, M., Eds.) pp 57–67, Adenine Press, New York.

Fluhr, R., Moses, P., Morelli, G., Coruzzi, G., & Chua, N.-H. (1986a) *EMBO J. 5*, 2063–2071.

Fluhr, R., Kuhlemeier, C., Nagy, F., & Chua, N.-H. (1986b) *Science 232*, 1106–1112.

Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686–14707.

Green, P. J., Kay, S. A., & Chua, N.-H. (1987) *EMBO J. 6*, 2543–2549.

Hagerman, P. J. (1985) *Biochemistry 24*, 7033–7037.

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

Herrera-Estrella, L., Van der Broeck, G., Maenhaut, R., Van Montagu, M., Schell, J., Timko, M., & Cashmore, A. R. (1984) *Nature 310*, 115–120.

Koepsel, R. R., & Khan, S. A. (1986) *Science 233*, 1316–1318.

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

Kuhlemeier, C., Green, P. J., & Chua, N.-H. (1987) *Annu. Rev. Plant Physiol. 38*, 221–257.

Marini, S. C., Levene, S. D., Crothers, D. M., & Englund, P. T. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 279–501.

Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.

Pickersky, E., Bernatzky, R., Tanskley, S. D., & Cashmore, A. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3880–3884.

Sarma, M. H., Gupta, G., Sarma, R. H., Bald, R., Engelke, U., Oei, S. L., Gessner, R., & Erdmann, V. A. (1987) *Biochemistry 26*, 7707–7714.

Sarma, M. H., Gupta, G., & Sarma, R. H. (1988) *Biochemistry 27*, 3423–3432.

Sassone-Corsi, P., & Borrelli, E. (1986) *Trends Genet. 8*, 215–219.

Simpson, J., Van Montagu, M., & Herrera-Estrella, L. (1986) *Science 233*, 34–38.

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.

Trifonov, E. N., & Sussman, J. L. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3816–3820.

Ulanovsky, L., & Trifonov, E. N. (1987) *Nature 362*, 720–722.

Wu, R., Wu, N.-H., Hanna, Z., Georges, F., & Narang S. (1984) in *Oligonucleotide Synthesis* (Gait, M. J., Ed.) pp 135–151, IRL Press, Oxford.

Yoon, C., Privè, G. G., Goodsell, D. F., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332–6336.

Zurkin, V. B., Lysov, Y. P., & Ivanov, V. I. (1979) *Nucleic Acids Res. 6*, 1081–1096.