Correlation of intrinsic DNA curvature with DNA property periodicity

Andrei Gabrielian**, Sándor Pongor*

International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park, Padriciano 99, 34012 Trieste, Italy

Received 27 June 1996; revised version received 19 July 1996

Abstract Twelve di- and trinucleotide parameter sets representing various structural, thermodynamic or bendability-related properties of DNA were tested in the prediction of DNA curvature applying Fourier analysis on curved and straight, A/T-type or G/C-type DNA sequence motifs. The best predictions were obtained with a new consensus bendability scale created by combining a nucleosome-based and a deoxyribonuclease I-based parameter set. Geometry calculations on the same sequences showed that the helical parameters derived from NMR structures can correctly predict curvature, as distinct from the parameters derived from X-ray crystallographic analysis.

Key words: DNA structure; DNA curvature; Sequence periodicity; Fourier analysis

1. Introduction

Sequence-dependent DNA bending plays a crucial role in many biological events such as packaging, transcription, site-specific recombination and DNA replication [1]. It is important therefore to predict both intrinsically curved regions and flexible, easy-to-bend sites in DNA sequences. There are a number of models and algorithms available (see e.g. [2,3] for references) but there is still no complete consensus on either the nature of intrinsic DNA curvature or the sequence-dependent parameters suitable for its prediction.

The central concept of intrinsic DNA curvature is helical phasing [1]: Repeating a uniform structural deformation at about every 10 or 10.5 basepairs in a sequence (i.e. 'in phase' with the helical repeat of B-DNA) can result in a macroscopic curvature of the DNA trajectory. By far the best characterized elements involved in curvature are phased A/T tracts. A/T-tract-based curvature can be correctly predicted by most algorithms [2,4], moreover it can be simply recognized by visual inspection or by pattern recognition methods. Also, it was shown that DNA curvature can be caused by the periodic repeats of other motifs, such as GGGCCC [5], as well as by such elements as nicks, gaps [6] and, to some extent, unpaired loops [7]. These examples lead us to suppose that a helically phased repetition of a certain property (such as those related to stability or flexibility of DNA) and not only a given sequence pattern might be a sufficient basis for intrinsic curvature.

The aim of the present work was to determine the correlation of experimentally known curvature with the periodicity of various sequence-dependent properties of DNA. For this purpose we chose 11 representative sets covering various structural, thermodynamic and bendability-related properties. We compare these parameters on a selected set of 'difficult DNA motifs' [5,8,9] that are known to cause problems in predicting curvature from sequence. We show that helical roll angles deduced from NMR measurements, but not the ones based on X-ray crystallographic data, can correctly predict curvature in these examples, and present a consensus bendability scale which is especially suitable for this purpose.

2. Materials and methods

2.1. Sequence-dependent parameters and calculation of property profiles

Twelve di- and trinucleotide property scales were used (Tables 1 and 2). Six of them are helical roll angles [10–15] that are known to reflect the major deviations of the helix axis in DNA structures from the ideal straight line [3]. Three scales are free energies (ΔG) that characterize the relative stability of B-DNA with respect to the transitions to A, Z or coil forms, respectively [16–18]. The two bendability scales are based on fractional occurrences of trinucleotides in bent or bendable regions, derived from deoxyribonuclease I (DNase I) cutting [19] and nucleosome positioning experiments [2,20], respectively. All scales were taken from the literature and applied in their original form. A consensus bendability scale was calculated as the average of the last two scales.

The property profiles were constructed by dividing the sequence into overlapping di- or trinucleotides and assigning a corresponding scale value from Table 1 or Table 2, respectively. For the uniformity of the further calculations, we constructed test sequences of 210 nucleotides from the short repeat sequences given in Table 3.

2.2. Fourier analysis

For the analysis of periodicity in a DNA parametric profile, we used a sensitive and robust variant of the discrete Fourier transform

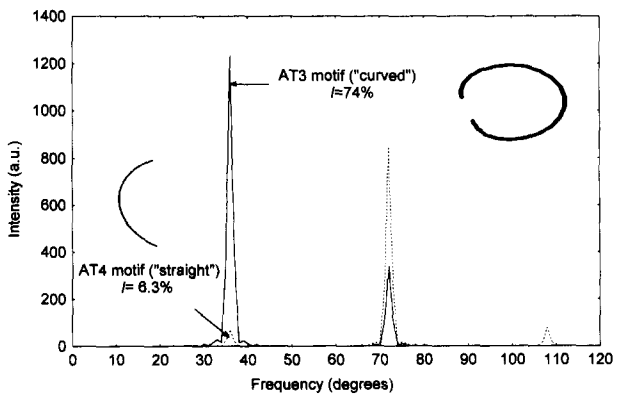

Fig. 1. Examples of Fourier spectra. Fourier spectrum of the sequence AT3 $(AAAATTTTGC)_n$ (continuous line) and of the sequence AT4 $(TTTTAAAAGC)_n$ (dotted line) calculated according to Eq. 1 using the consensus bendability scale (Table 2). The discrimination index D is calculated according to Eq. 3. The sequences are taken from [9]. Representations of the DNA path were calculated from the helical roll, tilt and twist angles of Ulyanov and James [10] using standard matrix transformation technique [14].

*Corresponding author. Fax: (39) (40) 226 555.
E-mail: pongor@icgeb.trieste.it

**Permanent address:* Institute of Molecular Biology, Russian Academy of Sciences, Vavilov St. 32, 117984 Moscow, Russian Federation.

PII S0014-5793(96)00855-1

Table 1
Dinucleotide parameter sets

		AATT	CGCG	ACGT	CCGG	ATAT	AGCT	GATC	GCGC	CATG	TATA
1	G(B→A), Aida [16], (kJ/mol)	−0.96	−10.79	4.52	2.26	6.82	0.79	3.18	8.27	5.10	0.42
2	G(B→Z), Hartmann et al. [17], (kJ/mol)	16.30	16.70	5.40	10.00	10.00	14.20	10.50	2.90	19.20	24.70
3	G(B→C), Breslauer et al. [18], (kJ/mol)	7.90	15.10	5.40	13.00	6.30	6.70	6.70	13.00	7.90	3.80
4	Roll, Bolshoy et al. [13], (°)	−6.50	6.70	−0.90	1.20	2.60	8.40	−2.70	−5.00	1.60	0.90
5	Roll, De Santis et al. [14], (°)	−5.40	4.60	−2.40	1.30	−7.30	1.00	2.00	−3.70	6.70	8.00
6	Roll, Calladine et al. [15], (°)	0.00	3.30	3.30	3.30	3.30	3.30	3.30	3.30	3.30	6.60
7	Roll, Bansal et al. [12], (°)	2.66	3.21	−0.70	5.82	−0.29	7.10	0.31	−6.40	−7.50	0.54
8	Roll, Gorin et al. [11], (°)	0.50	6.60	0.40	6.50	−0.60	2.90	−0.10	−7.00	1.10	2.60
9	Roll, Uljanov and James [10], (°)	1.40	11.10	0.90	6.20	−2.10	3.10	3.40	5.70	12.10	4.00

[21]. By this method, the periodicity of any property along a sequence can be detected by calculating a Fourier transform power spectrum:

$$P(\omega) = \left[\sum_{k=1}^{L} v_k \cos(k\omega)\right]^2 + \left[\sum_{k=1}^{L} v_k \sin(k\omega)\right]^2 \quad (1)$$

where v_k is a property values at the k-th sequence position, L is the length of the sequence and ω is the helical twist angle. This expression is expected to give a maximum around the repeat length of B-DNA (10.5 nucleotides per turn) if a sequence is curved and the property scale is suitable for curvature detection. A quantitative index of periodicity in the range of DNA helical repeat is:

$$I = \frac{\sum_{\omega=30}^{40} P(\omega)}{\sum_{\omega=1}^{180} P(\omega)} \cdot 100\% \quad (2)$$

This formula expresses that portion of the whole Fourier spectrum which is in the range corresponding to the periodicity of B-DNA. In this study a range of 30–40° in twist angle (9–12 nucleotides per turn) was used. It includes the dominant Fourier peak of all the curved sequences analyzed in this work, since they all contain 10 or 10.5 nucleotide repeats. Small changes of this range do not alter the qualitative picture obtained with our examples of curved and straight DNA. Cornette et al. [21] also describe a least-squares method for testing periodicities which is an alternative to the Fourier power spectrum. While the least-squares method may be preferable for very short sequence segments [21], we did not find any appreciable difference with our examples, so we based our study on Fourier analysis.

Given a set of i straight and j curved sequences one can define a discrimination index D which is the quantitative difference between the average of the I values calculated for the two groups:

$$D = \frac{1}{i}\sum_{i} I_{\text{curved,i}} - \frac{1}{j}\sum_{j} I_{\text{straight,j}} \quad (3)$$

The discrimination index was calculated for all sequences as well as for the groups of A/T-type and G/C-type motifs (Table 4). In fact, the present form of discriminant index is quite arbitrary. Therefore, it should be considered only as a coarse ranking indicator to highlight the qualitative differences between parameter sets.

2.3. Geometry calculations

Curvature geometry of DNA was calculated with the BEND program of Goodsell and Dickerson [2] kindly provided by the authors. In this program, the normal vectors of successive basepairs are added up vectorially and the angle between two averaged normal vectors 15 basepairs apart is used as the indicator of curvature as described [2].

3. Results and discussion

Fig. 1A shows typical Fourier power spectra calculated using the consensus bendability parameters. The dominating peak around 36° (10 nucleotides per helical turn) in the spectrum of the curved motif is much less pronounced in the spectrum of the straight motif. These spectra illustrate that a given feature (i.e. DNA bendability) is 'in phase' with the helical repeat of B-DNA within the curved sequence, and is 'out of phase' (with a periodicity of 5 nucleotides, i.e. half of a helical turn) in the opposite case. The rather obvious difference is quantitatively shown by the I indexes given in the figure.

Table 4 shows a comparison of the various scales on 7 test

Table 2
Trinucleotide parameter sets

Trinucleotide	Bendability (DNase I) [19]	Bendability (nucleosome) [20]	Consensus bendability scale	Trinucleotide	Bendability (DNase I) [19]	Bendability (nucleosome) [20]	Consensus bendability scale
AAA/TTT	0.1	0.0	0.05	CAG/CTG	9.6	4.2	6.90
AAC/GTT	1.6	3.7	2.65	CCA/TGG	0.7	5.4	3.05
AAG/CTT	4.2	5.2	4.70	CCC/GGG	5.7	6.0	5.85
AAT/ATT	0.0	0.7	0.35	CCG/CGG	3.0	4.7	3.85
ACA/TGT	5.8	5.2	5.50	CGA/TCG	5.8	8.3	7.05
ACC/GGT	5.2	5.4	5.30	CGC/GCG	4.3	7.5	5.90
ACG/CGT	5.2	5.4	5.30	CTA/TAG	7.8	2.2	5.00
ACT/AGT	2.0	5.8	7.80	CTC/GAG	6.6	5.4	6.00
AGA/TCT	6.5	3.3	4.90	GAA/TTC	5.1	3.0	4.05
AGC/GCT	6.3	7.5	6.90	GAC/GTC	5.6	5.4	5.50
AGG/CCT	4.7	5.4	5.05	GCA/TGC	7.5	6.0	6.75
ATA/TAT	9.7	2.8	6.25	GCC/GGC	8.2	10.0	9.10
ATC/GAT	3.6	5.3	4.45	GGA/TCC	6.2	3.8	5.00
ATG/CAT	8.7	6.7	7.70	GTA/TAC	6.4	3.7	5.05
CAA/TTG	6.2	3.3	4.75	TAA/TTA	7.3	2.0	4.65
CAC/GTG	6.8	6.5	6.65	TCA/TGA	10.0	5.4	7.70

Table 3
DNA sequence motifs used for the comparison of parameter sets

Name	Sequence	Curvature	Reference
AT1	$(caaaattttg)_n$	curved	Hagerman [9]
AT2	$(cttttaaaag)_n$	straight	Hagerman [9]
AT3	$(gaaaattttc)_n$	curved	Hagerman [9]
AT4	$(gttttaaaac)_n$	straight	Hagerman [9]
GC1	$(agggccctagaggggccc\ tag)_n$	curved	Brukner et al. [5]
GC2	$(aaaaactctctaaaaactctcgggccctagaggggccctaga)_n$	straight	Brukner et al. [8]
GC3	$(aaaaactctctaaaaactctcgaggggccctagagggcccta)_n$	curved	Brukner et al. [8]

sequences. Four of these are the A/T-type motifs of Hagerman [9], in which a seemingly similar A/T periodicity gives rise to a drastic difference in curvature, while the other three are G/C-type motifs [5,8]. The performance of each scale can be conveniently judged from the *D* value which will be high if the scale can distinguish curved and straight sequences. The higher the *D* value, the better the discriminative ability of a scale. A negative *D* value, on the other hand, shows that, with the given scale, the straight motifs appear to be more curved than the curved ones. According to the results in Table 4, none of the scales performed equally well on both the A/T-type and G/C-type motifs. The best overall results were obtained with the consensus bendability scale and with the NMR-derived roll angles of Ulyanov and James [10]. Several scales (Brukner's bendability [19] and DeSantis' roll angles [14]) performed well on A/T sequences but not on G/C motifs. On the other hand, the nucleosome bendability scale, calculated from the data of Satchwell et al. [20] by Goodsell and Dickerson [2], which showed the highest *D* value on the G/C-type motifs, could not distinguish curved and straight motifs in the A/T-rich group.

The consensus bendability scale can be best pictured as an attempt to get rid of the particular features of the experimental systems upon which the two scales are based (i.e. DNase I cutting and nucleosome binding), and to emphasize their common properties. The consensus scale is a simple arithmetic average of these two scales; no weighting was used to optimize its predictive power on the present examples. Nevertheless, as the results in Table 4 show, the consensus scale performs quite well on both the A/T-type and G/C-type motifs and can therefore be considered a suitable qualitative indicator of DNA curvature. It is worth mentioning here that bendability scales should not be used for quantitative curvature calculations, since they cannot be rationally decomposed into helical geometry parameters (such as roll, tilt, twist) necessary for the reconstruction of the helical path of DNA. Also, frequency statistics on bend/bendable sites will reflect both the flexibility and the inherent geometry of these sites, but will not in themselves allow one to distinguish between these [19].

The performance of any given property is expected to be better for trinucleotide or tetranucleotide scales than for a dinucleotide scale, simply because the former incorporate a larger sequence context. One should also take into account that much larger experimental data sets are available for building statistical bendability scales than the current set of X-ray or NMR-based DNA structures. In this respect it is interesting to mention that the dinucleotide scale of NMR-based roll angles was one of the best predictors of curvature, comparable to the trinucleotide-based consensus bendability scale. It was unexpected, on the other hand, that roll angles derived from the X-ray structures [11,12] did not perform very well in distinguishing curved and straight motifs (Table 4). Since this difference can be, in principle, related to the nature of the technique we used to distinguish between curved and straight motifs, we carried out an additional independent

Table 4
Quantitative comparison of DNA parameter scales using Fourier analysis on selected sequences

	Scale	Periodicity index (%)[a]							Discrimination index[b]		
		Curved motifs				Straight motifs			All motifs	AT motifs	GC motifs
		AT1	AT3	GC1	GC3	AT2	AT4	GC2			
1	G(B→A), Aida et al. [16], (kJ/mol)	17.30	15.15	32.82	30.92	29.51	0.61	7.18	11.61	1.16	24.69
2	G(B→Z), Hartmann et al. [17], (kJ/mol)	0.49	2.51	49.14	40.18	49.69	43.56	8.71	−10.91	−45.13	35.95
3	G(B→C), Breslauer et al. [18], (kJ/mol)	31.96	18.30	76.69	21.47	23.50	13.13	32.52	14.06	6.81	16.56
4	Roll, Bolshoy et al. [13], (°)	4.48	10.10	21.17	5.28	20.12	21.47	29.63	−13.48	−13.51	−16.41
5	Roll, De Santis et al. [14], (°)	43.44	68.40	29.75	13.25	0.18	0.18	33.87	27.30	55.74	−12.37
6	Roll, Calladine et al. [15], (°)	20.80	22.90	32.64	36.87	35.03	54.54	8.28	−4.31	−22.94	26.48
7	Roll, Bansal et al. [12], (°)	49.88	0.18	3.44	0.80	0.25	8.53	3.01	9.65	20.64	−0.89
8	Roll, Gorin et al. [11], (°)	12.02	16.80	0.31	0.80	12.15	7.55	0.43	0.77	4.56	0.13
9	Roll, Uljanov and James [10], (°)	56.50	52.70	68.71	23.50	17.85	8.16	3.31	40.58	41.60	42.80
10	Bendability, Brukner et al. [19], (au)[c]	73.62	78.30	4.85	24.05	1.47	1.96	19.88	37.44	74.25	−5.43
11	Bendability, Satchwell et al. [20], Goodsell and Dickerson [2], (°)	65.77	62.80	73.01	52.88	57.91	48.65	4.23	26.69	11.01	58.72
12	Consensus bendability, this work (au)[c]	73.01	74.11	51.66	39.45	22.33	6.32	3.56	48.82	59.24	42.00

[a]Calculated according to Eq. 2.
[b]Calculated according to Eq. 3.
[c]au, arbitrary units.

Table 5
Comparison of various geometric parameters in curvature calculation using the BEND program [2]

	Scales (roll, tilt, twist angles)	Angle of curvature (°)[a]							Discrimination index	
		Curved motifs				Straight motifs			AT motifs	GC motifs
		AT1	AT3	GC1	GC3	AT2	AT4	GC2		
1	Bolshoy et al. [13]	18.91	48.13	37.82	67.61	50.42	43.55	78.50	−13.47	−25.79
2	De Santis et al. [14]	62.46	57.87	12.03	44.69	8.60	20.06	48.13	45.84	−19.77
3	Calladine et al. [15]	15.47	15.47	9.74	19.48	5.73	5.73	22.92	9.74	−8.31
4	Bansal et al. [12]	88.81	29.80	10.89	31.52	29.80	18.34	20.06	35.24	1.15
5	Gorin et al. [11]	17.19	18.34	12.03	13.18	18.91	12.03	18.34	2.29	−5.73
6	Ulyanov and James [10]	74.49	49.85	32.66	32.66	14.90	20.63	21.77	44.41	10.89

[a]The angle between the normal vectors of two basepairs situated 15 nucleotides apart [2]. The maximum value is given in those cases when the value varied along the sequence motif.

comparison based a more rigorous geometry calculation using the BEND program of Goodsell and Dickerson [2]. To calculate helical curvature, this program employs, in addition to roll angles, sequence-dependent tilt and twist helical angles. This comparison (Table 5) confirms the results of the Fourier periodicity analysis. We obtained the best results with the helical parameters derived from NMR structures by Ulyanov and James [10]. It is especially noteworthy that the NMR-based scales gave by far the best discrimination for G/C motifs, while most other models gave negative D values, i.e. showed curvature in straight motifs and vice versa. It thus appears that, at least with the present examples, the helical rotational parameters derived from X-ray crystallography are less suitable for detecting curvature then are the NMR-based rotational parameters. As the set of dinucleotide parameters can be considered as abstracted generalized form of a structural database, the difference is noteworthy. The systematic differences that may exist between DNA structure in solution and in crystals warrant further interest in this field.

References

[1] Travers, A.A. and Klug, A. (1990) in: Cozzarelli, N.R. and Wang, J.C. eds.), DNA Topology and its Biological Effects, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 57–106.
[2] Goodsell, D.S. and Dickerson, R.E. (1994) Nucl. Acids Res. 22, 5497–5503.
[3] El Hassan, M.A. and Calladine, C.R. (1995) J. Mol. Biol. 251, 648–664.
[4] Calladine, C.R. and Drew, H.R. (1996) J. Mol. Biol. 257, 479–485.
[5] Brukner, I., Dlakic, M., Savic, A., Susic, S., Pongor, S. and Suck, D. (1993) Nucl. Acids Res. 21, 1025–1029.
[6] Mills, J.B., Cooper, J.P. and Hagerman, P.J. (1994) Biochemistry 33, 1797–803.
[7] Kahn, J., Yun, E. and Crothers, D.M. (1994) Nature 368, 163–166.
[8] Brukner, I., Susic, S., Dlakic, M., Savic, A. and Pongor, S. (1994) J. Mol. Biol. 236, 26–32.
[9] Hagerman, P. (1986) Nature 321, 449–450.
[10] Ulyanov, N. and James, T. (1995) Methods Enzymol. 261, 90–115.
[11] Gorin, A., Zhurkin, V. and Olson, W. (1995) J. Mol. Biol. 247, 34–48.
[12] Bansal, M., Bhattacharyya, D. and Ravi, B. (1995) CABIOS 11, 281–287.
[13] Bolshoy, A., McNamara, P., Harrington, R.E. and Trifonov, E.N. (1991) Proc. Natl. Acad. Sci. USA 88, 2312–2316.
[14] De Santis, P., Palleschi, A. Savino, M. and Scipioni, A. (1990) Biochemistry 29, 9269–9273.
[15] Calladine, C.R., Drew, H.R. and McCall, M.J. (1988) J. Mol. Biol. 201, 127–137.
[16] Aida, M. (1988) J. Theor. Biol. 130, 327–335.
[17] Hartmann, B., Malfoy, B. and Lavery, R. (1989) J. Mol. Biol. 207, 433–444.
[18] Breslauer, K., Frank, R., Blocker, H. and Marky, L. (1986) Proc. Natl. Acad. Sci. USA 83, 3746–3750.
[19] Brukner, I., Sanchez, R., Suck, D. and Pongor S. (1995) EMBO J. 14, 1812–1818.
[20] Satchwell, S.C., Drew, H.R. and Travers, A.A. (1986) J. Mol. Biol. 191, 639–659.
[21] Cornette, J., Cease, K., Margalit, H., Spouge, J., Berzofsky, J. and DeLisi, C. (1987) J. Mol. Biol. 195, 659–685.