Sequence specificity of curved DNA

Stephan Diekmann

Max-Planck-Institut für biophysikalische Chemie Am Fassberg, D-3400 Göttingen-Nikolausberg, FRG

Received 30 September 1985

Anomalously slow migration of DNA fragments on polyacrylamide gels is interpreted as resulting from curvature of the DNA fragment. Different models have been suggested to explain DNA curvature. In this work a number of DNA fragments were synthesized, cloned, and electrophoretically characterized to distinguish between these models. Strong anomaly of migration is found for sequence stretches $(dA)_n$ repeated in phase with the helix turn with n at least 4. For n smaller than 4 only negligible anomaly is observed. The results contradict the purine-clash hypothesis. The data can be explained by assuming longer stretches of As to be in a B′-form, and that tilt of this structure might be the reason for its curvature.

DNA structure *DNA bending* *DNA curvature* *kinetoplast DNA* *Poly(dA)·poly(dT)*

1. INTRODUCTION

A restriction fragment of a minicircle of mitochondrial DNA of *Leishmania tarantolae* (kinetoplast DNA) exhibits unusually low electrophoretic mobility in acrylamide gels, whereas in agarose gels migration is observed in accordance with fragment size [1–4]. Migration anomalies of this kind have been observed before [5,6] but the magnitudes were not as pronounced as for the kinetoplast DNA. Physicochemical measurements suggest that this abnormal electrophoretic mobility is due to a curve DNA structure [3,4,7–9] (review [10]). In the kinetoplast DNA fragment 2 particular subfragments are important for its abnormal gel electrophoretic behaviour [11]. One common feature of their sequences is the periodic presence of short runs of As (4–6 As in each) with a period of about 11 bp [11]. Here it is shown that such recurring stretches of As repeated with the period of the helix turn, are indeed the origin of anomalous gel migration.

Several models have been suggested to explain this abnormal behaviour [3,7–9]. In general, dinucleotide wedge angles would impose a curvature on the helix axis. The AA dinucleotide and, for symmetry reasons, the TT dinucleotide were found to have large wedge angles [12,13]. If these dinucleotides are positioned periodically every helix turn, their wedge angles may sum up to produce a detectable curvature, explaining the observed effect [3,11]. A different origin of curvature might be the steric hindrance of neighbouring purines on opposing strands ('purine-clash') [9,14–16]. Furthermore, the stretches of As might have a non-B-form structure with curvature resulting from a kink at the junction of the different helix types [8]. Here evidence is presented that the purine-clash hypothesis does not explain the observed altered gel electrophoretic mobility. A 'tilt' model is suggested.

2. MATERIALS AND METHODS

A series of 10 bp long oligonucleotides have been kindly provided by Ott and Eckstein [17] and Preuss and Scheit. These precursors 1–5 (sequences see sections 3 and 4) were kinased, ligated and cloned into the *Cla*I site of pJW200 [11]. The plasmid pJW200 is derived from pBR322 [18] by the deletion of the sequence in between 2 *Hae*II sites at bp 236 and 2352. The obtained plasmids were transformed into *E. coli* HB101 cells. The cells were grown to saturation, the plasmids

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

alkaline extracted and purified by HPLC on Nucleogen 4000 columns [19] (Machery and Nagel). The sequence of each insert was determined following the protocol of Maxam and Gilbert [20].

The altered mobility of the analysed fragments is measured in 10% polyacrylamide gels (30:1 acrylamide to bis; 45 mM Tris-borate, 1.25 mM EDTA at 20°C, 5 V/cm, pH 8.6) relative to marker fragments (pBR322 cut by the restriction endonucleases *Hae*III and *Hin*fI [21], both from New England Biolabs) and a 123 bp ladder (BRL). The abnormal behaviour is presented by the '*k* factor' defined as the ratio of the gel electrophoretic mobility of a fragment to that expected of a 'normal' DNA fragment of the same sequence length. A reduced mobility results in a *k* factor larger than unity, while normal migrating fragments have a *k* factor of 1.00. The experimental error of the *k* factors is estimated to about ± 0.02.

3. RESULTS

To determine the sequence specificity of DNA curvature and the validity of the different models explaining the effect, a series of 10 bp long oligonucleotides were synthesized [17]. After annealing they form the following double-helical precursors:

(1) 5′ d C G A C A G G A C T 3′ family 1A10
3′ d T G T C C T G A G C 5′ containing no A_2 dinucleotide ('A_1')

(2) 5′ d C G A C A A G A C T 3′ family 2A10
3′ d T G T T C T G A G C 5′ containing one A_2 in a repeat of 10 bp

(3) 5′ d C G A C A A A G C T 3′ family 3A10
3′ d T G T T T C G A G C 5′ containing A_3

(4) 5′ d C G A C A A A A C T 3′ family 4A10
3′ d T G T T T T G A G C 5′ containing A_4

(5) 5′ d C G A A A A A G C T 3′ family 5A10
3′ d T T T T T C G A G C 5′ containing A_5

The synthetic oligomers above are denoted according to the number of As in a row. The precursors were ligated to each other and cloned into the *Cla*I site of pJW200. However, the 5 types were

Table 1

Plasmid	Family	Insertion arrangement (see legend)	Insert length (bp)	*k* factor *Hae*III
pK1A107	1A10	1A1T1A3T1A	70	1.023
pK1A109	1A10	1T8A	90	1.025
pK2A104	2A10	1A3T	37	1.017
pK2A107	2A10	−6A	70	1.011
pK3A101	3A10	1A	10	1.020
pK3A107	3A10	3T1A1T2A	70	1.025
pK3A108/1	3A10	4T4A	79	1.042
pK3A108/2	3A10	1A−3T2A1T	80	1.051
pK4A101	4A10	1A	10	1.005
pK4A102	4A10	2A	20	1.023
pK4A104	4A10	4T	41	1.094
pK4A106/1	4A10	2T4A	60	1.254
pK4A106/2	4A10	2A2T1A1T	60	1.290
pK4A107/1	4A10	2T5A	70	1.324
pK4A107/2	4A10	4A1T2A	70	1.360
pK4A108	4A10	4A1T3A	80	1.386
pK4A1010	4A10	1A−8A	100	1.524
pK5A101	5A10	1A	10	1.020
pK5A102	5A10	2A	20	1.038
pK5A105	5A10	3T2A	50	1.080
pK5A106/1	5A10	1T5A	60	1.250
pK5A107/1	5A10	5T2A	70	1.260
pK5A107/2	5A10	2T5A	70	1.220
pK5A108	5A10	5T3A	79	1.188

The table presents the *k* factors of fragments containing the indicated insert cut out of the given plasmids by the restriction endonuclease *Hae*III. The clones have to be classified not only due to their sequence family (see sequences 1–5) and to the insert length, but also due to the repeat arrangement of the insert. This is done by reading the sequence 5′ to 3′ (starting at the *Eco*RI site) and representing the 10 bp of the repeat by an 'A' if the As are read, or 'T', if the Ts are in this strand. For example, the insert

*Eco*RI 5′ dGAGTTTTGTC′GAGTTTTGTC′
T T

GACAAAACTC′GACAAAACTC′
A A

GACAAAACTC 3′ *Hin*dIII
A

(only one strand shown) belongs to the family 4A10 (4 As in a repeat of 10 bp) and has the arrangement 2T3A (with the total insert length 50 bp). A dash in the sequence arrangement (for example, pK2A107 or pK4A1010) indicates the presence of 10 bp but a mutation in the A stretch

not mixed together so that each precursor is ligated to its own type only. For all 5 sequences a number of plasmids is obtained (belonging to one sequence 'family') each of which contains an insertion at the *Cla*I site built up of a different number of 10 bp precursors. Each 10 bp long precursor can be ligated to the next either head-to-head, tail-to-tail, or in tandem. According to its insert, each plasmid is classified not only by the total length of the insert and the number of As within its precursor sequence but also by the arrangement of the repeats within the insert (see table 1).

The migration anomaly of the fragments is measured on acrylamide gels by determining the apparent length of these fragments relative to marker fragments. It is presented as the k factor [11] (i.e. apparent divided by sequence length, see section 2). For a series of clones, the family, insertion arrangement, and the corresponding k factors are presented in table 1. The results clearly show that clones of the families 1A10, 2A10 and 3A10 (i.e. up to 3 As in a repeat of 10 bp) do not show migration anomaly (k factors close to 1.00) for any insertion length and arrangement. Strong anomaly is observed for the families 4A10 and 5A10. The anomaly does not increase going from 4 to 5 As per precursor (instead, it slightly decreases, for example compare pK4A107/1 with pK5A107/2 in table 1). Short insertions do not show migration anomaly for any clone family indicating either a low sensitivity of the gel matrix to small degrees of curvature or thermally driven bending of the normal marker fragments. High symmetry of the insert (runs of tandem repeats) as well as nearly randomly arranged inserts show similar degrees of migration anomaly (cf. pK4A106/1 with pK4A106/2). This might be due to 10 bp not being exactly the helix repeat [11].

4. DISCUSSION

The sequences 1–4 are designed in such a way that they all have the same purine-pyrimidine sequence. Following the purine-clash hypothesis all insertions should lead to comparable curvature independent of the family, clearly in contrast to the experimental finding. Thus, obviously purine-clash is not the origin of the observed migration anomaly. But repeated A_n stretches with $n \geq 4$ in phase with the helix turn are a sufficient sequence element. However, the effect does not seem to be continuously built up by AA dinucleotides. Following this explanation, the retardation should continually increase in the order 2A10–5A10. Instead, migration anomaly is observed for the families containing A_n with $n \geq 4$ only. In addition, no increase of the effect is observed going from 4 to 5 As. It should be noted, however, that the 'jump' in migration behaviour between longer inserts of the families 3A10 and 4A10 might be an artifact of the detection system: the gel matrix might detect curvature only beyond a threshold value too high for the 3A10 family. DNA fragments containing homopolymer stretches of poly(dA)·poly(dT) (42 and 71 bp length) do not show anomalous gel migration (S. Diekmann, unpublished). Thus, not the presence of this sequence but its alteration with other DNA sequences seems to be the origin of gel migration anomaly.

Those sequences in the kinetoplast fragment responsible for the curvature have a helical repeat of 10.4 bp [11], between 10.1 bp for poly(dA)·poly(dT) and 10.6 bp for normal B-form DNA [22,23]. This suggests that the stretches of A_{4-6}, which constitute half a helix turn within these sequences, have the typical structure of poly(dA)·poly(dT) [22–24]. This conclusion is supported by other experimental evidence: the kinetoplast DNA fragment injected into rabbits gives rise to immunoglobulins which specifically recognize poly(dA)·poly(dT) [25]. Thus, I propose that the A_n with $n \geq 4$ is in a B′-form structure which cannot form for $n \leq 3$. The negligibly small anomaly of the families 1A10, 2A10, and 3A10 might be due to curvature caused by repeated dinucleotide wedge angles or purine-clash. However, this effect is too small to explain the strong anomaly observed for the 4A10 and 5A10 families, as well as for the kinetoplast DNA fragment. A recent 500 MHz NMR study [26] using one-dimensional NOE concluded that poly(dA)·poly(dT) adopts the classical B-form, however, other experimental evidence accumulates from Raman [27,28], linear dichroism [29], enzyme digestion [30] and X-ray diffraction [31] studies suggesting a B′-form structure for poly(dA)·poly(dT).

The results show that 4 or 5 consecutive As with a repeat distance in phase with the helix turn lead to strong abnormal mobility in acrylamide gels. Perhaps, neighbouring influences of the B-form

sequences do not allow A stretches shorter than A_4 to be in the B′-form structure. A sufficiently long stretch of As is assumed to adopt a B′ structure, possibly an A-B hybrid structure [24,28,31]. Linear dichroism [29] and theoretical energy-minimization studies [32] indicate that this hybrid structure has a tilt, which, by normal base stacking at the junction of this structure (with tilt) to B-form (roughly without tilt), might give rise to the curvature of the helix axis [8,33]. This tilt model which assumes a smooth transition [33] between the structures, is distinguished from a junction bending model [8] which suggests a base stacking disruption (kink) at the junction. In the kinetoplast sequence no kink could be detected by DNase I digestion [7].

ACKNOWLEDGEMENTS

I wish to thank Drs F. Eckstein, J. Ott, K.H. Scheit and D. Preuss for oligonucleotides, Drs R. Clegg, E. Trifonov and L. Ulanovsky for helpful discussions, B. Preitz and M. Meyer for technical assistance, and Dr M. Kahlweit and the Deutsche Forschungsgemeinschaft (grant Di 258/3-1) for support.

REFERENCES

[1] Simpson, L. (1979) Proc. Natl. Acad. Sci. USA 76, 1585–1588.
[2] Challberg, S.S. and Englund, P.T. (1980) J. Mol. Biol. 116, 447–472.
[3] Marini, J.C., Levene, S.D., Crothers, D.M. and Englund, P.T. (1982) Proc. Natl. Acad. Sci. USA 79, 7664–7668.
[4] Marini, J.C. and Englund, P.T. (1983) Proc. Natl. Acad. Sci. USA 80, 7678.
[5] Brown, N.L. and Smith, M. (1977) J. Mol. Biol. 116, 1–30.
[6] Ross, W., Shulman, M. and Landy, A. (1982) J. Mol. Biol. 156, 505–529.
[7] Marini, J.C., Effron, P.N., Goodman, T.C., Singleton, C.K., Wells, R.D., Wartell, R.M. and Englund, P.T. (1984) J. Biol. Chem. 259, 8974–8979.
[8] Wu, H.M. and Crothers, D.M. (1984) Nature 308, 509–513.
[9] Hagerman, P.J. (1984) Proc. Natl. Acad. Sci. USA 81, 4632–4636.
[10] Trifonov, E.N. (1985) CRC Crit. Rev. Biochem., in press.
[11] Diekmann, S. and Wang, J.C. (1985) J. Mol. Biol., in press.
[12] Trifonov, E.N. and Sussman, J.L. (1980) Proc. Natl. Acad. Sci. USA 77, 3816–3820.
[13] Trifonov, E.N. (1980) Nucleic Acids Res. 8, 4041–4053.
[14] Ulyanov, N.B. and Zhurkin, V.B. (1984) J. Biomol. Struct. Dyn. 2, 361–385.
[15] Calladine, C.R. (1982) J. Mol. Biol. 161, 343–352.
[16] Dickerson, R.E. (1983) J. Mol. Biol. 166, 419–441.
[17] Ott, J. and Eckstein, F. (1984) Nucleic Acids Res. 23, 9137–9142.
[18] Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 77–90.
[19] Colpan, M. and Riesner, D. (1985) submitted.
[20] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[21] Stellwagen, N. (1983) Biochemistry 22, 6186–6193.
[22] Peck, L.J. and Wang, J.C. (1981) Nature 292, 375–378.
[23] Rhodes, D. and Klug, A. (1981) Nature 292, 378–380.
[24] Prunell, A., Goulet, I., Jacob, Y. and Goutorbe, F. (1984) Eur. J. Biochem. 138, 253–257.
[25] Diekmann, S. and Zarling, D. (1985) submitted.
[26] Sarma, M.H., Gupta, G. and Sarma, R.H. (1985) J. Biomol. Struct. Dyn. 2, 1057–1084.
[27] Thomas, G.A. and Peticolas, W.L. (1983) J. Am. Chem. Soc. 105, 993–996.
[28] Jolles, B., Laigle, A., Chinsky, L. and Turpin, P.Y. (1985) Nucleic Acids Res. 13, 2075–2085.
[29] Edmondson, S.P. and Johnson, W.C. (1985) Biopolymers 24, 825–841.
[30] Diekmann, S., Robert-Nicoud, M. and Wang, J.C., in preparation.
[31] Arnott, S., Chandrasekaran, R., Hall, I.H. and Puigjaner, L.C. (1983) Nucleic Acids Res. 11, 4141–4151.
[32] Kitzing, E.V. and Diekmann, S., in preparation.
[33] Selsing, E., Wells, R.D., Alden, C.J. and Arnott, S. (1979) J. Biol. Chem. 254, 5417–5422.