Nucleic Acids Research, 1993, Vol. 21, No. 20 4703–4710

Long-range organization and sequence-directed curvature of *Xenopus laevis* satellite 1 DNA

Philippe Pasero, Nikolay Sjakste, Christophe Blettry, Claude Got and Monique Marilley*
Laboratoire de Génétique, URA CNRS 1189, Faculté de Médecine, 27 Bd Jean Moulin, 13385 Marseille cedex 5, France

Received July 15, 1993; Revised and Accepted September 3, 1993

ABSTRACT

We have investigated the long-range organization and the intrinsic curvature of satellite 1 DNA, an unusual tandemly-repeated DNA family of *Xenopus laevis* presenting sequence homologies to SINEs. PFGE was used in combination with frequent-cutter restriction enzymes not likely to cut within satellite 1 DNA and revealed that almost all the repeating units are tandemly organized to form large arrays (200 kb to 2 Mb) that are marked by restriction length polymorphism and contain intra-array domains of sequence variation. Besides that, we have analysed the secondary structure of satellite 1 DNA by computer modelling. Theoretical maps of curvature obtained from three independent models of DNA bending (the dinucleotide wedge model of Trifonov, the junction model of Crothers and the model of de Santis) showed that satellite 1 DNA is intrinsically curved and these results were confirmed experimentally by polyacrylamide gel electrophoresis. Moreover, we observed that this bending element is highly conserved among all the members of the satellite 1 DNA family that are accessible to analysis. A potential genetic role for satellite 1 DNA based on this unusual structural feature is discussed.

INTRODUCTION

The genome of most complex eukaryotes contains large amounts of repetitive sequences, which may represent several percent of the total genetic material. These repeated DNA components are generally classified into two groups: tandemly repeated sequences, that are organized in a head-to-tail fashion and form arrays as large as several thousand kb [1] and interspersed sequences, such as SINEs or LINEs, that are flanked by unique sequences [2, 3].

One of the major reiterated components of the *Xenopus laevis* genome is an unusual type of repeated DNA that presents features common to both groups of eukaryotic repetitive sequences. This 750 bp repeating unit, which may represent up to 1.35% of the genomic DNA, has been described as satellite 1 DNA [4], OAX DNA (Oocyte-Activated Xenopus) [5] and RHM2 (Repetitive *Hin*dIII Monomer) [6]. Like other tandemly-repeated sequences, the repeats of the satellite 1 family are organized in a head-to-tail fashion. However, these sequences contain a RNA polymerase III transcription unit presenting striking homologies to tRNA coding genes [7]. Moreover, the presence of direct repeats suggests that this tRNA-related region has been generated by a RNA-mediated mechanism [7, 8]. Consequently, it has been recently proposed that the satellite 1 family was initially a SINE that was later amplified in tandem more than 30,000 times by a DNA-mediated mechanism [2].

The function of tandemly repeated sequences has been widely discussed through the past decades but still remains unclear [1]. Satellite DNAs are often confined to the heterochromatin present around centromeric or telomeric chromosome regions. This provocative location suggests that satellite DNA may have some structural function in the chromosomes. However, the molecular basis of genetic processes involved in higher-order chromatin condensation processes is still unknown.

New details concerning the structure of satellite DNA have been recently reported that might bring new insights into the study of their role in the folding of the chromatin fiber. A stable curvature of the DNA has been detected in mouse satellite DNA and this unusual secondary structure was shown to be directly involved in centromeric heterochromatin condensation [9]. Subsequently, curved domains were identified in several other tandemly-repeated sequences, and it has been proposed that bent DNA may be a conserved structural feature of satellite DNAs that may account for its genetic function [10, 11]. Indeed, it is now clear that bent DNA plays a significant role in nucleosome positioning [12, 13]. Moreover, it is assumed that the tandem repetition of hundreds of bent repeating unit would generate super-structures that may be able to attract specifically non-histone proteins and thus 'code' for a specific folding of the chromatin fiber [14, 15]. The specific binding of nuclear proteins to tandemly-repeated sequences has been evidenced [16–18] and strongly supports the 'chromatin folding code' hypothesis.

In this report, we have investigated the long-range organization of satellite 1 DNA by pulsed-field gel electrophoresis (PFGE) techniques. Almost all the repeating units were shown to be tandemly arranged to form arrays ranging from 200 kilobases to 2 megabases. Using a PFGE two-dimensional gel method [19], we also observed that satellite 1 sequence variants are not

* To whom correspondence should be addressed

randomly arranged within the arrays but may define locus-specific repetitive higher-order sequence structures. Besides that, we have analysed the structure of this unusual type of reiterated element by computer modelling. Three independent theoretical models that were proved to be viable for predicting the intrinsic curvature of DNA were used [20–23]. All these models predicted that satellite 1 DNA is intrinsically curved and this DNA bending was confirmed experimentally by polyacrylamide gel electrophoresis analyses. Moreover, we observed that this strong curve is highly conserved among all the members of the family that are accessible to analysis These results suggest that bent DNA is an important structural feature of satellite 1 DNA that may account for its genetic function in the nucleus.

MATERIALS AND METHODS

Sample preparation and digestion of chromosomal DNA

Genomic DNA samples were prepared from *Xenopus laevis* erythrocytes in low-melting agarose plugs (Boehringer-Mannheim) according to the procedure of Vollrath and Davis [45]. Blood cells were collected as described by Marilley and Gassend-Bonnet [46]. Each genomic DNA plug (8 mm×5 mm×2 mm) contains about 9.10^4 cells or 6μg of DNA. Before digestion, the DNA plugs are equilibrated in the appropriate digestion buffer two times for 1 hour with gentle shaking. Digestion occurs for 6 hours at 37°C in a 150 μl final volume of digestion buffer, with at least a tenfold excess of endonuclease (Boehringer-Mannheim). Enzymes are added a second time part way through the incubation.

Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis was performed on a contour-clamped homogeneous electric field system [47] from Biorad Laboratories in 0.5×TBE (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.0) at 14°C. Standard electrophoretic conditions are: 200 V, 0.8% chromosomal-grade agarose (Biorad), 120° reorientation angle, 90 s switch time for 15 h followed by 60 s switch time for 8 h. Intermitent-Field PFGE [26] was also performed in order to resolve restriction fragments larger than 1600 kb. In this case, electrophoresis conditions were a 70 hour run with 90 s switching times interrupted by 30 second pauses, 0.5% agarose, 160 V. *S.cerevisiae* chromosomes (strain 334/Beckman) were used as molecular weight markers. Gels are stained with ethidium bromide (0.5μg/ml) and photographed under UV light.

Two-dimensional gel electrophoresis

The intra-array organization of satellite 1 DNA sequence variants was investigated as described by Warburton and Willard [19]. Genomic DNA plugs are digested with *Bam*H1 and duplicated samples are submitted to PFGE in the standard running conditions described above. One half of the gel is stained with ethidium bromide and transferred to nylon filters. The other half of the gel is dissected into individual lanes with a scalpel. The gel slabs are thoroughly rinsed in distilled water and are equilibrated for several hours in freshly prepared restriction enzyme buffer. The buffer is aspirated and is replaced with 200 μl of a solution containing 2000 U of the appropriate restriction enzyme in the digestion buffer equally distributed over the gel slab. Digestion occurs overnight at 37°C in a hermetically sealed container. The electrophoresis is performed in the second dimension with a conventional system for 8 h at 2.4V/cm, although PFGE/PFGE 2D electrophoresis was also used.

Southern blotting and hybridization

The gel is exposed to 310 nm ultraviolet light to nick the DNA, then denatured with NaOH 0.5M, NaCl 0.5M for 5 mn, neutralized with NaCl 1.5M, Tris 1M pH 7.5 for 2×15 min with gentle shaking and transferred to nylon filters (Hybond N+, Amersham) using 20 SSC solution. Prehybridization and hybridization of the filters were carried out in a hybridization oven as described by the supplier. The satellite 1 DNA probe used (pE190) was kindly provided by Dana Carroll [4].

DNA trajectory modelling

Four independent computer programs based on the algorithm of Tung and Harvey [32] were developed in Turbo Pascal (Borland) according to the common nomenclature for nucleic acids structure parameters defined at the EMBO workshop on DNA curvature and bending [33]. The spatial path of the molecule was calculated from its primary sequence with tilt and roll angles given by the theoretical models [20–23] and twist values calculated by Kabsch et al. [48]. Points in the sugar-phosphate backbone or at the center of the base pair plates were used for the pictorial presentation [32]. Local variations of curvature along the DNA molecule (ENDS ratio) were calculated as the ratio of the curvilinear to the end-to-end distance for a 120 bp segment moving along the molecule by 10 bp steps [34].

DNA PAGE analysis

Polyacrylamide gel electrophoresis was performed on 7% gels at 4°C for 26 h at 5.3 V/cm in 1×TBE and agarose gel electrophoresis was performed on 1% gels at room temperature for 14 h at 1.8 V/cm in 1×TBE. K factor values, corresponding to the ratio of the observed length to the expected length, were calculated for each DNA fragments using a 123-bp DNA fragment ladder (Gibco-BRL) as molecular weigth standard. Gels were stained with ethidium bromide (0.5μg/ml) and photographed under UV light.

Positional autocorrelation analysis

The periodicity of the distibution of A-tracts, T-tracts, ApA and TpT dinucleotides was performed as described by Marini et al. (1982). The following correlation function was computed:

$$G_j = \left(\sum_{i=1}^{L-j-m+1} d_i\, d_{i+j}\right) \Big/ \sum_{i=1}^{L-j-m+1} d_i$$

where L refers to the length of the sequence (741 bp), m to the length of the subsequence ($m=2$ for dinucleotides, $m=3$ for $d(A.T)_3$ tracts) and j to the lag (1 to 80). d_i is set to 1 when the subsequence searched start at nucleotide i. The analysis was repeated for four 250 bp windows sliding along the sequence by 170 bp steps and the two DNA strands were analysed separately.

RESULTS

Long-range organization of satellite 1 DNA

There are about 20,000 to 40,000 copies of satellite 1 DNA tandemly organized in *X.laevis* genome [4]. Recently, we have shown that the use of PFGE in combination with a set of carefully

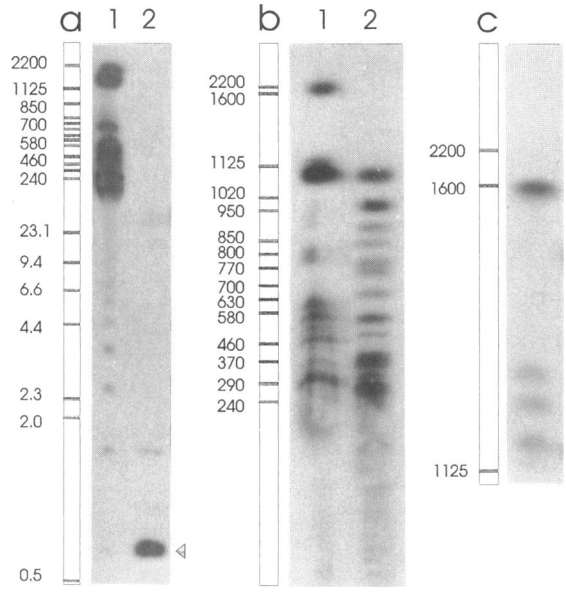

Figure 1. PFGE analysis of the long-range organization of satellite 1 DNA in *X.laevis* genome. Pulsed-field electrophoreses were performed with different running conditions and the gels were blotted and hybridized with a satellite 1 DNA probe. **a.** standard conditions with short running time (8h). lane 1, *X.laevis* genomic DNA digested with *SalI*, *XhoI* and *Eco*RV; lane 2, genomic DNA digested with *SalI*, *XhoI*, *Eco*RV and *Hind*III. Arrow head indicates the postion of satellite 1 monomers (approximately 750 bp). **b.** standard running conditions. lane 1, *X.laevis* genomic DNA digested with *SalI*, *XhoI* and *Eco*RV; lane 2, *X.laevis* genomic DNA digested with *SalI*, *XhoI* and *Eco*RV and *Eco*RI. **c.** Intermittent-field PFGE. *X.laevis* genomic DNA digested with *SalI*, *XhoI* and *Eco*RV. Detailed PFGE running conditions are given in materials and methods. *Hind*III-digested λ DNA (0.5 to 23.1 kb) and *S.cerevisiae* chromosomes (240 to 2200 kb) are used as molecular weight markers.

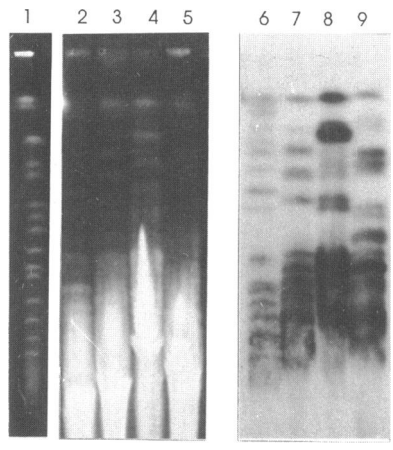

Figure 2. Individual restriction length polymorphism of satellite 1 arrays. lane 1, PFGE molecular weight marker (*S.cerevisiae* chromosomes); lane 2 to 5, *X.laevis* genomic DNA prepared from different individuals and digested with *SalI*, *XhoI* and *Eco*RV; lane 6 to 9, Southern blot hybridization with a satellite 1 probe. Running conditions used are the standard conditions described in materials and methods, except that 60 s pulses for 15 h, followed by 90 s pulses for 8 h were used. The size of *S.cerevisiae* chromosomes are given in figure 1b. It should be noticed that the individual of lane 4/8 is the same one as in figure 1.

selected restriction enzymes can lead to the isolation of large arrays of tandemly-repeated DNA [24, 25]. This approach was used here to investigate the higher-order organization of satellite 1 DNA.

Frequent-cutter restriction enzymes not likely to cut within satellite 1 DNA were selected by computer analysis. In theory, these enzymes should cleave only flanking sequences, releasing tandem arrays intact. DNA plugs prepared from *X.laevis* blood cells were digested with *SalI*, *XhoI* and *Eco*RV and were submitted to pulsed-field electrophoresis. Different types of running conditions were successively used in order to resolve restriction fragments ranging from 750 bp (the size of a satellite 1 monomer) up to 2 Mb.

Figure 1a shows a pulsed-field electrophoresis performed in running conditions suitable for sorting small restriction DNA fragments. Southern blotting and hybridization with a satellite 1 DNA probe reveals that almost all the satellite 1 repeats are found as restriction fragments sizing at least 100 kb. When longer running periods are used, multiple high-molecular weight arrays of satellite 1, ranging from 100 kb to 2 Mb, are resolved (figure 1b). Additional restriction fragments are detected when endonucleases cutting rarely within satellite 1 DNA are used (figure 1b, lane 2). In contrast, identical patterns are obtained when genomic DNA plugs are digested with different restriction enzymes having no cutting sites within satellite 1 DNA (data not shown). Consequently, we assume that the high molecular weight restriction fragments hybridizing to the probe consist of continuous stretches of satellite 1 tandem repeats. Finally, intermitent-field electrophoresis [26] was used to estimate the size of the clusters found at limit mobility in standard running conditions. Figure 1c shows that the actual size of the largest array is 1600 kb in the individual tested. Three clusters that co-migrate in standard PFGE running conditions are also clearly resolved.

In order to investigate the individual polymormism of satellite 1 DNA in *Xenopus*, genomic DNA plugs prepared from blood cells of different individuals were digested with the same set of restriction enzymes and submitted to PFGE. Figure 2 shows that the clusters of satellite 1 are marked by a high degree of restriction length polymorphism, as already reported for other tandemly repeated sequences [1, 27]. It should be noticed that some of the high-molecular weight DNA fragments observed after ethidium bromide staining are not positive to the satellite 1 probe. These restriction fragments correspond to large clusters of ribosomal DNA and of oocyte-type 5S DNA that are co-purified here. A wide range of variability is also observed for these tandemly-repeated DNA families [24 and unpublished observations].

In conclusion, PFGE analyses revealed that satellite 1 DNA is tandemly organized to form large arrays (200−2000 kb) that are marked by restriction length polymorphism. The total number of clusters per individual was estimated. This yielded an estimate of approximately 20,000 to 25,000 copies of satellite 1 per haploid genome. This value is in good agreement with the estimates previously obtained by quantitative analysis [4, 6].

Intra-array organization of satellite 1 variants

Several bands positive to satellite 1 DNA probe are detected in gel when genomic DNA is digested with an enzyme that has only one cutting site per repeating unit. Indeed, sequence variants are generally observed in addition to the major 750 bp repeating unit. These variants are multimers resulting from the absence of one or several cutting sites and 1050 bp repeats which derive from the major 750 bp repeats by a single insertion event [4, 6].

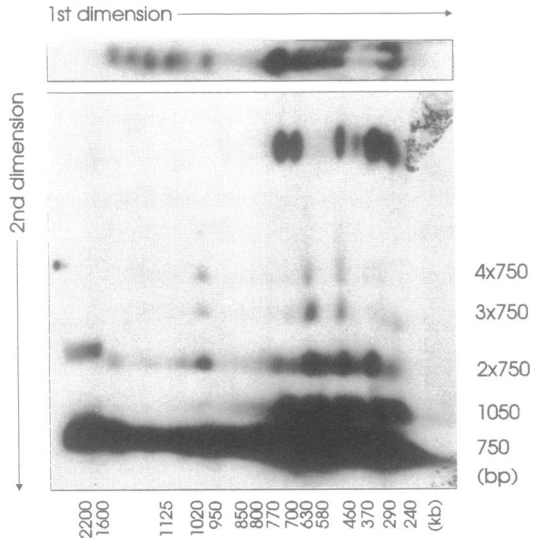

Figure 3. Two dimensional gel electrophoresis analysis of the intra-array organization of satellite 1 DNA sequence variants. Genomic DNA was digested with *Bam*H1, an endonuclease having no cutting site within satellite 1 sequence, and was submitted to PFGE with standard running conditions (see materials & methods). The gel is then dissected into individual lanes and one gel slab is digested with *Hind*III, an enzyme presenting one cutting site per repeating unit. The gel slice is then reoriented perpendicular to the initial direction of electrophoresis, submitted to a conventional gel electrophoresis, blotted and hybridized with a satellite 1 probe. A duplicate lane from the first dimension is used as reference for the size of the arrays. Molecular weight markers are *S.cerevisiae* chromosomes (1st dimension) and *Hind*III-digested λ DNA (2nd dimension)

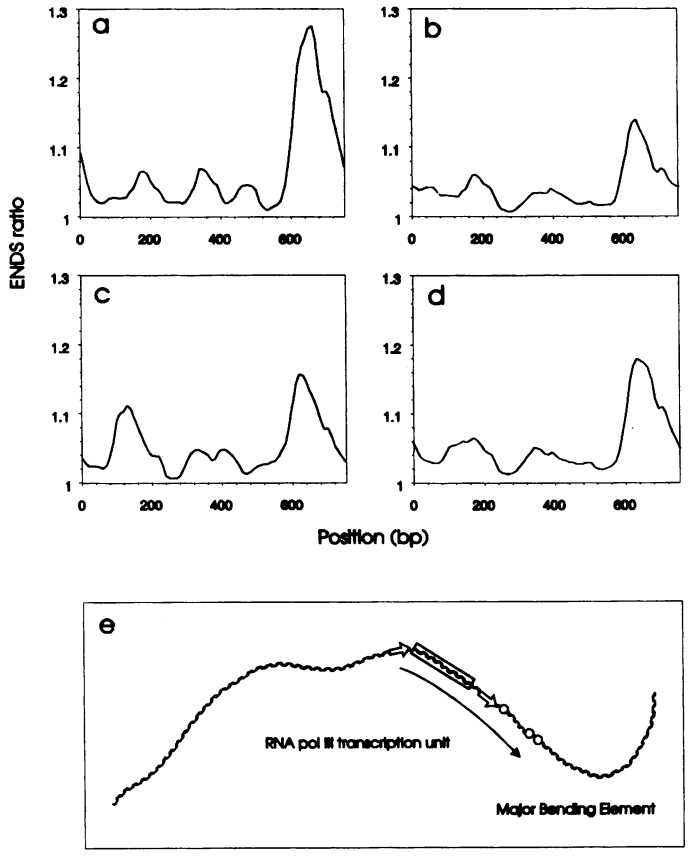

Figure 4. Theoretical maps of curvature calculated for satellite 1 DNA with three different models of DNA bending. **a.** Trifonov dinucleotide wedge model—1991 [21]. **b.** de Santis model [22]. **c.** Crothers junction model [23]. **d.** Average curvature map calculated from the three models **e.** Two-dimensional projection of the three dimensional path of a satellite 1 DNA monomer (*Hin*dIII–*Hin*dIII) calculated with the dinucleotide wedge model. The boxed region corresponds to the tRNA-related region as reported by Nagahashi et al. [8] and Meyerhof et al. [7]. Direct repeats are indicated by open arrows and transcription termination sites are indicated by open circles. The three-dimensional path of the molecule was computed as described in materials and methods. Local curvature (ENDS ratio) is calculated for 120 bp segments centered at positions each of 10 bp. Nucleotide numbering refers to the distance from the 5′ terminal *Hin*dIII site.

Here, we have used a two-dimensional gel electrophoresis technique based on the method of Warburton & Willard [19] to investigate the distribution of these satellite 1 sequence variants within the arrays. In figure 3, sequence variants are seen as spots above the broad signal corresponding to the major 750 bp repeat. Their position and distribution can be determined according to the duplicate lane from the first dimension. For instance, several spots corresponding to the 1050 bp repeats are detected. The distribution of these sequence variants is clearly non-random. The 1050 bp repeats are found within arrays ranging from 200 to 800 kb but not within larger arrays. Since the hybridization intensity is significantly weaker that the major 750 bp signal, we assume that these variants are only parts of larger composite arrays. However, it is to be noticed that the intra-array arrangement of these repeats is not accessible to analysis here.

In contrast, the distribution of multimers of the major 750 bp repeat is highly informative. Dimers, that are the interspersed form of this sequence variants are found in variable amount within the arrays. Three- and four-mers are also observed within some of the clusters. In addition to these spots, six sharp bands positive to the probe are found at limit mobility. It is likely that these large segments with no *Hind*III cutting sites correspond to sequence variants that have spread locally to form intra-array homogeneous domains. It should be noticed that similar results are obtained when different individuals are tested or when different restriction enzymes are used (data not shown).

Computer modelling analysis of satellite 1 DNA

DNA bending was shown to be relevant for gene function in more than one hundred eukaryotic and prokaryotic biological systems [28] and a stable curvature of the helix axis is now likely to be a conserved structural feature of satellite DNAs [10]. To assess whether satellite 1 DNA is also intrinsically curved, we have analysed its structure by computer modelling.

Since the discovery of intrinsically bent DNA in kinetoplast DNA of *L.tarentolae* [29], several theoretical models have been developed to predict the spatial path of the helix axis from its primary nucleotide sequence [28]. These models are generally classified into two groups, depending on whether or not the DNA remains in its normal B-conformation. The *wedge models* stipulate that a stable sequence-directed bending of the DNA axis can be seen as the result of the cumulative effect of small deflections associated with every dinucleotide [30]. In contrast, the *junction models* state that tracts of oligo(dA) of at least three nucleotides do not have a classical B-DNA conformation in physiological conditions. A deformation of the helix axis is then observed at the junctions between A-tracts and B-DNA [31].

In this report, we have used the dinucleotide wedge model of Trifonov [20, 21], the model of de Santis [22] and the junction model of Crothers [23] to predict the three-dimensional path of satellite 1 DNA. Computer programs based on the algorithm of

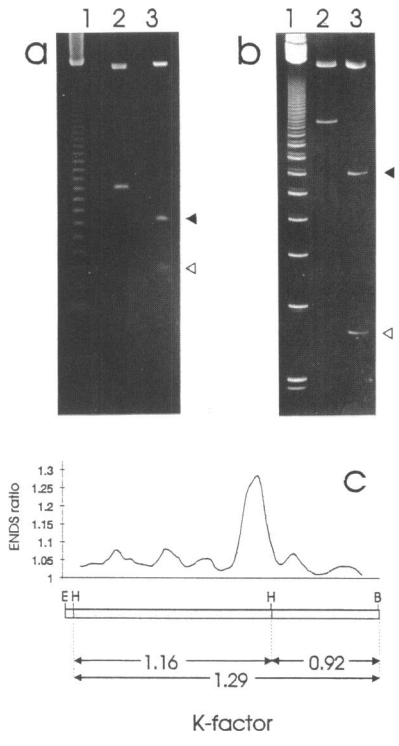

Figure 5. Gel electrophoresis analysis of satellite 1 DNA curvature. **a.** 1% agarose-TBE gel electrophoresis. lane 1: molecular weight marker (123-bp DNA fragment ladder). lane 2: pE190 digested with *Eco*RI and *Bam*HI. lane 3: pE190 digested with *Hin*dIII and *Bam*HI. **b.** 7% polyacrylamide-TBE gel electrophoresis. Lanes are identical to (a). The size in base pairs of selected molecular weigth markers is indicated. **c.** Curvature map of the *Eco*RI−*Bam*HI restriction fragment of pE190 corresponding to the 742 bp satellite 1 DNA sequence cloned at the HindIII site of pBR322. K factor values (observed length/expected length) are given for each restriction fragment tested. Abreviations for restriction enzymes sites are as follow: B: *Bam*HI, E: *Eco*RI, H: *Hin*dIII.

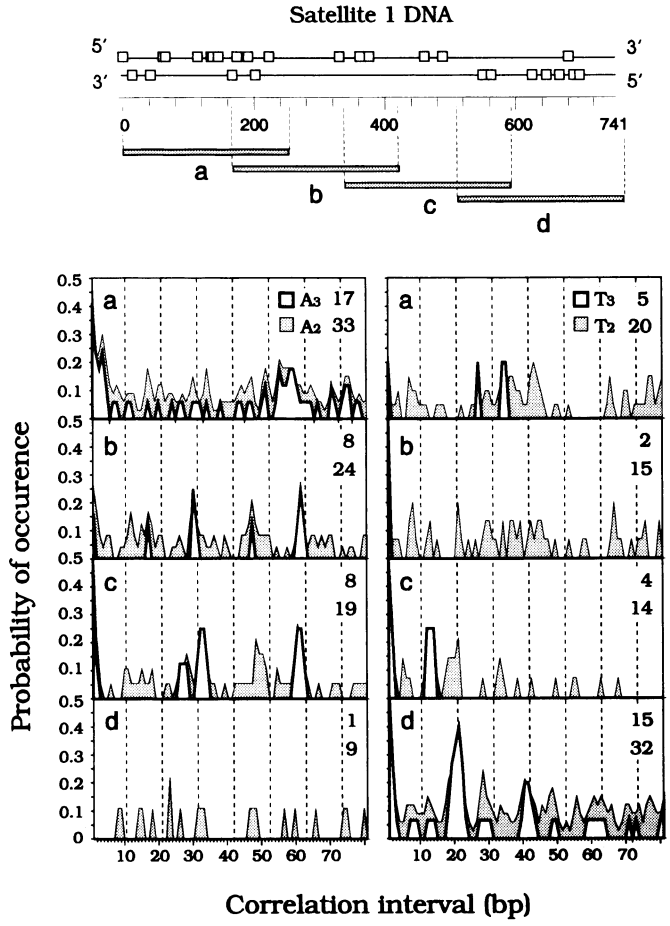

Figure 6. Positional autocorrelation analysis of A_3/T_3 tracts and of ApA/TpT dinucleotides in satellite 1 DNA. Four segments of 250 bp (a,b,c,d) moving along the 5′-3′ strand of satellite 1 DNA [4] by 170 bp steps were analysed. The total number of A3/T3 tracts and AA/TT dinucleotides present in the window analysed in each panel is indicated. Dotted lines represent multiples of a 10.5 bp periodicity. The distribution of $d(A.T)_{n\geq 3}$ tracts (open boxes) along both strands of the molecule is shown.

Tung and Harvey [32] were developed using the nomenclature defined at the EMBO workshop on DNA Curvature and Bending [33]. The secondary structure of satellite 1 DNA was calculated with each model and the variation of DNA curvature along the molecule (ENDS ratio) was plotted as reported by Eckdahl & Anderson [34].

A good agreement is observed between the curvature maps calculated with the different theoretical models (figure 4). The dinucleotide wedge model (figure 4a) shows two overlapping peaks at position 660 and 710 bp. These peaks correspond to two adjacent bending element of 78° and 66°. A similar structure is predicted by the model of de Santis (figure 4b) and the junction model (figure 4c). Several minor peaks are observed in the regions 150−200 and 350−450. However, with the exception of the peak predicted by the junction model at position 150, these bending elements are not significantly higher than the basal level of DNA curvature. Figure 4d shows a mean curvature map calculated from the three theoretical models. This map clearly indicates that the most prominent structural feature of satellite 1 are two strong bending elements located close to the *Hin*dIII cutting-site of the repeating unit.

Figure 4e shows a projection of the three dimensional path of satellite 1 DNA calculated with the wedge model of Trifonov. This scheme shows that there is a cooperative effect of the two major curves of satellite 1 DNA, leading to an overall bending of the helix axis of approximately 120°. A curvature map calculated with a larger windows (250 bp) spanning the two bending elements confirmed that the cumulative effect of these curves generate a strong bending element of 117° (Ends ratio=1.92) centered around position 675 bp (data not shown).

The RNA polymerase III transcription unit is represented in figure 4d. It is obvious that the major bending element is not related to the RNA-mediated region of satellite 1 DNA. Moreover, it should be noticed that in contrast to mammalian SINEs in which the amplified region roughly corresponds to the transcriptional unit, the retrotransposed region of satellite 1 DNA represents only 22% of the repeating unit.

Analysis of satellite 1 DNA curvature in polyacrylamide gel

Bent DNA shows an anomalously low mobility in polyacrylamide gels. This altered electrophoretic migration is not observed when the samples are electrophoresed in agarose gels [35]. In order to verify experimentally the theoretical results obtained by computer modelling, the electrophoretic behavior of satellite 1 DNA was been investigated by gel electrophoresis. Different restriction fragments prepared from the plasmid pE190 [4] which

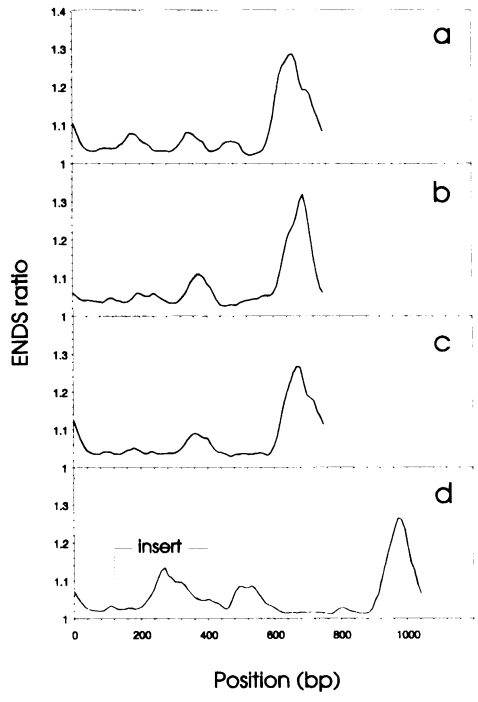

Figure 7. Curvature maps calculated for four different members of the satellite 1 DNA family. **a.** Satellite 1 DNA [4]. **b.** OAX DNA [5]. **c.** RHM2 [6]. **d.** RHM5 [6]. Computer modelling was performed with the Trifonov dinucleotide wedge model [21] as reported in figure 4.

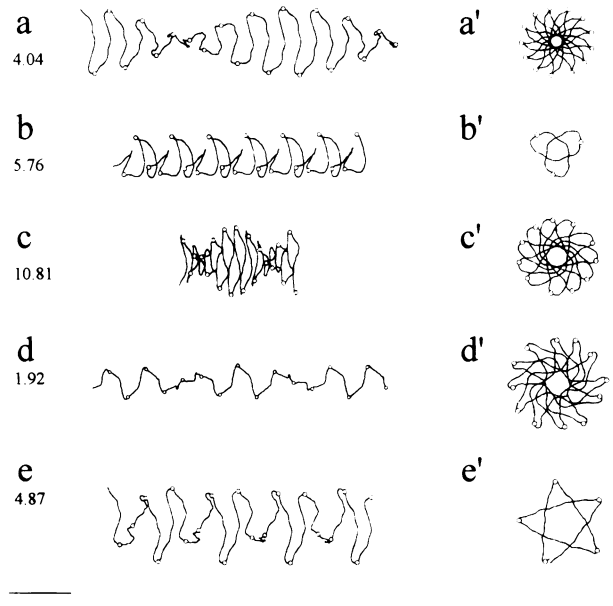

Figure 8. Computer-generated molecular models of homogeneous tandem arrays of satellite 1 sequence variants. **a.** satellite 1 DNA (20 repeating units—14820 bp); **b.** OAX DNA (20 repeating units—15040 bp); **c.** RHM2 (20 repeating units—14900 bp); **d.** RHM5 (16 repeating units—16592 bp); **e.** satellite 1 Δ ApA 158 (20 repeating units—14780 bp). This sequence was constructed from satellite 1 DNA by deleting 2 bp at postion 158. The three dimensional path of the helix axis is calculated with the dinucleotide wedge model [21]. For each sequence, a longitudinal view (a, b, c, d, e) and an axial view (a', b', c', d', e') of the array is shown. The spots correspond to points where the repeat starts again. The ENDS ratio, corresponding to the ratio of the curvilinear to the end-to-end distance between the two extreme base pairs of the array, are also indicated. Bar is 500 bp, except for d., where it represents 1200 bp.

correspond to a monomeric repeat of satellite 1 DNA cloned at the *Hin*dIII site of pBR322, were analysed and their electrophoretic profiles are shown in figure 5. The 742-bp *Hin*dIII restriction fragment (satellite 1 DNA) exhibits an anomalous electrophoretic behavior in polyacrylamide gel but not in agarose gel, confirming that this sequence is intrinsically curved. Moreover, we observed that the deviation from the expected gel mobility is higher for a *Eco*RI−*Bam*HI restriction fragment containing 346 bp of pBR322 DNA at the 3'-end of satellite 1 sequence. The mobility of curved DNA in polyacrylamide gels is highly influenced by the precise location of the bend relative to the ends of the fragment and the electrophoretic retardation is maximum when the bending element is located close to the center of the molecule [36]. Since the *Hin*dIII−*Bam*HI pBR322 fragment is not intinsically curved, we assume that the main contribution of this sequence to the electrophoretic retardation of the molecule is to move the relative position of the bending element of satellite 1 DNA toward the center of the molecule. These results confirm that satellite 1 DNA contains a strong bending element located close to the 3'-end of the sequence.

Nucleotide sequence elements responsible for the bending of satellite 1 DNA

It is generally assumed that the repetition of adenine tracts [31] or ApA dinucleotide [30] with a spacing close to the helix pitch is mainly responsible for DNA curvature. The *Xenopus* RHM2 DNA, like several other satellite DNA sequences, is statistically enriched in runs of three or more adenine residues [10]. Here, we observed that there are 27 d(A.T)n tracts in satellite 1 DNA with an average length of 3.7 nucleotides. In order to address the question of whether these sequence motifs are responsible for satellite 1 DNA curvature, we have made a positional autocorrelation analysis of the distribution of ApA dinucleotides and d(A)3 tracts along this molecule. Figure 6 shows the analysis of four overlapping segments of 250 bp spanning the 750 bp satellite 1 DNA sequence. The occurence of A tracts was analysed separately for the two strands and revealed a strong heterogeneity in the distribution of these sequence elements. Remarkably, all but one of the 16 A3 tracts found in segment d are located on the same strand. This segment, which contains the major bending elements predicted by the theoretical models of DNA curvature, also shows a clear periodicity in the occurence of these tracts. Unusually, the average periods are approximately of 21 bp, a value corresponding exactly to two helix turns in B-DNA. No clear periodical arrangement is observed in the other regions. Therefore, we assume that the intrinsic curvature of satellite 1 DNA in is mainly due to the occurence of d(A.T) tracts in phase with the helix pitch that are located on the same strand in a region ranging from nucleotide 590 to nucleotide 740.

Bent DNA is a conserved structural feature of the satellite 1 DNA family

In order to address the question of whether bent DNA is a general feature of satellite 1 DNA, we have established a curvature map of all the members of this repeated DNA family for which the nucleotide sequence is available. The sequences analysed include three monomers of 750 bp (Satellite 1 DNA, OAX DNA and RHM2) and one 1037 bp unit (RHM5). The three 750 bp repeats share about 95% of homology. The 1037 bp unit differs from

RHM2 by a 287 nt insertion at position 115, the remaining portion showing a 84.3% homology [7]. The result of this analysis is reported in figure 7. Remarkably, strong bending elements are observed in the curvature maps of the four repeats. These bending element are not located in the regions of highest variability defined by Lam & Carroll [4]. In spite of the 5 to 15% divergence found between the repeats, the magnitude and the relative position of the overall curve appeared strikingly similar. As far as we can judge from the analysis of four different repetition units, this bending element is therefore a highly conserved structural feature of the satellite 1 DNA family.

Three-dimensional structure of tandem arrays of satellite 1 DNA

In order to analyse the structure of tandem arrays of satellite 1 DNA, we have constructed homogeneous arrays of different sequence variants and the three-dimensional shape of these sequence blocks was calculated by computer modelling.

Figure 8 shows a two-dimensional projection of the spatial path of these arrays as predicted by the Trifonov wedge model. All these structures show a clear superhelix shape, with each bending element located outwards from the helix axis. It is to be noticed that both major and minor bending elements from adjacent repeating units are phased with the helical pitch of DNA and generate an overall bending of nearly 180°. However, although these arrays consist of repeating units presenting very similar curvature maps (figure 7), highly polymorphic structures are observed for the arrays. Some of them complete one superhelical turn in about 2.2 kb while others require 9.6 kb (or 13 repeating units). Composite sequences such as the motif...$(741)_4(1000)(741)_2(1000)(741)_2(1000)(741)_4$... characterized by Lam and Carroll [4] were also modellised and appeared as highly folded structure with no obvious superhelical shape (data not shown).

To assess the effect of sequence variation on the structure of satellite 1 arrays, we have constructed sequence variants of satellite 1 DNA by deleting 1 to 5 nucleotides in the regions of greatest variability pointed out by Lam and Carroll [4]. Figure 8e shows a construction corresponding to a deletion of a dinucleotide ApA at position 158 of satellite 1 DNA. This single deletion event affects one of the minor bending elements of satellite 1 DNA but also changes the number of helical turns in the molecule. This rotation has little effect on individual repeats. However, as in the phasing analysis of Zinkel and Crothers [42], it modifies the relative orientation of two adjacent curved repeating units and therefore leads to dramatic changes in the shape and the periodicity of the superhelical structure. Periodicities varying from 13 to 3 repeating units were observed for satellite 1 DNA arrays by deleting 1 to 5 bp at position 158 (data not shown).

DISCUSSION

Much attention has been devoted through the past few years to the hypothetical biological function of the *X.laevis* satellite 1 DNA. This reiterated sequence was initially classified among complex satellite DNAs. However, new details concerning its evolutionary origin have been recently reported, suggesting that satellite 1 DNA is closely related to SINEs [2]. This unusual type of repetitive DNA is transcribed quite efficiently *in vivo* and *in vitro* [5, 7, 38]. However, it is worthwhile to mention that the transcripts of satellite 1 DNA (or OAX DNA) present no obvious secondary structure which might provide a clue to their function in the cell. Moreover, it seems that only a limited set of the chromosomal repeating units are effectively transcribed *in vivo* and this endogenous transcription is apparently regulated by other sequences [38]. Finally, no relatives of *X.laevis* satellite 1 are found in the closely related species *X.mulleri* and *X.borealis*, suggesting that this sequence is subject to no selective pressure [4, 6]. The genetic function of satellite 1 DNA in the genome of *Xenopus laevis* is therefore still under debate.

In this report, we have investigated another potential role for satellite 1 DNA based on its long-range organization and on its intrinsic structural features rather than on hypothetical RNA coding capabilities. Indeed, it has been proposed recently that intrinsically bent DNA may be a conserved feature of satellite DNAs [10] and several lines of evidence indicate that this unusual structure of the helix may influence chromatin organization at different levels. Nucleosome positioning is mainly determined by the anisotropic flexibility of DNA [12], but also by stable deflections of the helix axis [13]. Therefore, the reiteration of a bent motif in satellite DNA is likely to influence significantly the positioning of nucleosome core particles, and consequently the overall structure of the chromatin fiber. Moreover, it has been shown that repetitive sequence motifs with unusual helical DNA conformations are able to bind specific non-histone nuclear proteins [37, 18]. Since tandemly-repeated sequences have the ability to develop higher-order sequence structures on the chromosomes, it is assumed that sequence blocks of satellite DNA may attract specific nuclear proteins and therefore 'code' for a locus-specific folding of the chromatin fiber [14, 15]. This model is exemplified by a satellite DNA of the mouse genome. This centromeric satellite DNA was shown to be intrinsically curved [9]. The treatment of mouse cells with drugs such as distamycin A or Hoechst 33258 that eliminate DNA bending results in a strong decondensation of centromeric heterochromatin, suggesting that DNA curvature is directly involved in chromatin condensation processes.

In order to test whether bent DNA is also a conserved structural feature of satellite 1 DNA that may account for its role in the nucleus, its secondary structure was investigated by computer modelling and PAGE analysis. Three independant models of DNA curvature—the Trifonov wedge model [20, 21], the Crothers junction model [23] and the model of de Santis [22]—were used. These three models were selected in light of recent reports showing that they are viable tools for predicting the three-dimensional path of a DNA molecule from its primary sequence [34, 39–41]. It should be noticed that although these models yield fairly similar results, their parameters were derived from experiments that yielded to different sets of angles. Indeed, although it has been shown that A-tracts are mainly responsible for DNA stable curvature, it is still unclear where this bending precisely occurs [43]. The overall bending of the molecule can therefore be seen as the cumulative effect of small deflections of the helix axis at junctions between these A-tracts and B-DNA (junction models). Alternatively, this global curvature can also be the result of the vectorial sum of wedge angles associated with all ApA dinucleotide constituting these A-tracts (wedge models). However, it is worth pointing out these models are not necessarily incompatible and both may be understood in terms of X-ray crystal data [31,43]. The same overall bending of a DNA molecule can be generated from different sets of roll angles, and this bending was clearly shown by Calladine and Drew to be proportionnal to the first component of the Fourier analysis of the roll angles, at a period equal to the helical repeat of the DNA [44]. Therefore, although wedge and junction models use

different sets of Eulerian angles, they generally appear to fit the experimental reasonably well.

Computer modelling analyses revealed that satellite 1 DNA contains a stable curve of nearly 120° centered around position 675 bp. Recurring runs of adenine in phase with the helix pitch were found on the same DNA strand in this region and good agreement was observed between the curvature maps calculated with different theoretical models. Moreover, the existence of an intrinsic curvature was verified experimentally by polyacrylamide gel analysis, thereby confirming theoretical analyses. Finally, we also observed that this bending element is highly conserved among all the members of satellite 1 DNA family that are accessible to analysis. These results suggest that a stable curvature is an important structural feature of satellite 1 DNA.

Previous reports have suggested that satellite 1 DNA is tandemly arranged [4–6]. However the long-range organization of this reiterated sequence on the chromosomes has not been reported so far. Here, pulsed-field gel electrophoresis analyses revealed that the repeating units are tandemly arranged to form large arrays comprising up to 2500 repeats. Using different PFGE running conditions, we observed that almost all the repeating units are organized within arrays larger than 100 kb. Therefore, although satellite 1 DNA contains an internal domain related to interspersed sequences, it displays a long-range organization characteristic to complex satellite DNAs.

The *chromatin folding code* model [14, 15] stipulates that the locus-specific folding of the chromatin fiber in tandemly-repeated sequences is coded by locus-specific sequence blocks on the chromosome. There is an inherent potential for satellite DNA to develop locus-specific higher order sequence structures. Indeed, new variants continuously arise throughout the arrays of satellite DNA and sometimes spread locally to form intra-array domains of sequence variation. Such patterns of intermediate spread and fixation were pointed out here for satellite 1 DNA using PFGE 2D electrophoresis. They are likely to be the result of genetic mechanisms acting continuously to maintain sequence homogeneity and to spread variation throughout the arrays [1]. Here, the secondary structure of such homogeneous arrays of satellite 1 sequence variants was analysed by computer modelling. These arrays displayed very polymorphic super-helical structures, although they are made of monomers sharing a high degree of sequence homology and presenting very similar curvature maps. Moreover, we observed that a low level of sequence variation (2 bp) is sufficient to affect significantly the secondary structure of these tandem arrays. As it is shown in figure 8 (a and e), the deletion of two nucleotides in satellite 1 has little effect on the overall shape of the individual repeating units. However, it modifies the relative position of the adjacent bending elements and therefore affects strongly the super-helical periodicity of the array. These observations illustrate how the curving potential of a bent satellite DNA is amplified when this sequence occurs in tandem arrays and are consistent with to the *chromatin folding code* model. They might help us to understand better how locus-specific patterns of chromatin condensation are 'coded' by the secondary structure of the corresponding higher-order sequence blocks.

REFERENCES

1. Brutlag,D.L. (1980) *Ann. Rev. Genet.*, **14**, 121–144.
2. Okada,N. (1991) *Curr. Opinion Genet. Dev.* **1**, 498–504.
3. Martin,S.L. (1991) *Curr. Opinion Genet. Dev.* **1**, 505–508.
4. Lam,B.S. and Carroll,D. (1983) *J.Mol.Biol.* **165**, 567–585.
5. Ackerman,E.J. (1983) *EMBO J.*, **2**, 1417–1422.
6. Meyerhof,W., Tappeser,B., Korge,E. and Knöchel,W. (1983) *Nucleic Acids Res.* **11**, 6997–7009.
7. Meyerhof,W., Wittig,B., Tappeser,B. and Knöchel,W. (1987) *Eur. J. Biochem.*, **164**, 287–293.
8. Nagahashi,S., Endoh,H., Suzuki,Y. and Okada,N. (1991) *J. Mol. Biol.*, **222**, 391–404.
9. Radic,M.Z, Lundgren,K. and Hamkalo,B.A. (1987) *Cell*, **50**, 1101–1108.
10. Martinez-Balbas,A., Rodriguez-Campos,A., Garcia-Ramirez,M., Sainz,J., Carrera,P., Aymami,J. and Azorin,F. (1990) *Biochemistry*, **29**, 2342–2348.
11. Carrera,P., Martinez-Balbas,A., Portugal,J. and Azorin,F. (1991) *Nucleic Acids Res.* **19**, 5639–5644.
12. Travers,A.A. (1987) *Trends Biochem. Sci.* **12**, 108–112.
13. Hsieh,C.H. and Griffith,J.D. (1988) *Cell*, **52**, 535–544.
14. Vogt,P. (1990) *Hum. Genet.* **84**, 301–336.
15. Vogt,P. (1992) *Chromosoma* **101**, 585–589.
16. Strauss,F. and Varshavsky,A. (1984) *Cell* **37**, 889–901.
17. Neuer-Nitsche,B., Lu,X. and Werner,D. (1988) *Nucleic Acids Res.* **16**, 8351–8360.
18. Hibino,Y., Nakamura,K., Asano,S. and Sugano,N. (1992) *Biochem. Biophys. Res. Commun.* **184**, 853–858.
19. Warburton,P.E. and Willard.H.F. (1990) *J. Mol. Biol.* **216**, 3–16.
20. Ulanovsky,L.E. and Trifonov,E.N. (1987) *Nature* **326**, 720–722.
21. Bolshoy,A., McNamara,P., Harrington,R.E. and Trifonov,E.N. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 2312–2316.
22. De Santis,P., Palleschi,A., Savino,M. and Scipioni,A. (1988) *Biophys. Chem.* **32**, 305–317.
23. Koo,H. and Crothers,D.M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1763–1767.
24. Marilley,M., Pasero,P. and Got,C. (1992) *Exp. Cell Res.* **202**, 87–97.
25. Pasero,P. and Marilley,M. (1993) *Mol. Gen. Genet.* **236**, 448–452.
26. Turmel.C., Brassard.E., Slater.G.W. and Noolandi.J. (1990) *Nucleic Acids Res.* **18**, 569–575.
27. Willard,H.F. (1991) *Curr. Opinion Genet. Dev.* **1**, 509–514.
28. Hagerman,P.J. (1992) *Biochem. Biophys. Acta* **1131**, 125–132.
29. Marini,J.C., Levene,S.D., Crothers,D.M. and Englund,P.T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7664–7668.
30. Trifonov,E.N. (1991) *Trends Biochem. Sci.* **16**, 467–470.
31. Crothers,D.M., Haran,T.E. and Nadeau,J.G (1990) *J. Biol. Chem.* **265**, 7093–7096.
32. Tung,C.S. and Harvey,S.C. (1986) *Nucleic Acids Res.* **14**, 381–387.
33. Dickerson,R.E., Bansal,M., Calladine,C.R., Diekman,S., Hunter,W.N., Lavery,R., Nelson,H.C.M., Olson,W.K. Saenger,W., Shakked,Z., Sklenar,H., Soumpassis,D.M., Tung,C.S. von Kitzing,E., Wang,A.H.J. and Zhurkin,V.B. (1989) *J. Mol. Biol.* **205**, 787–791.
34. Eckdahl,T.T. and Anderson,J.N. (1987) *Nucleic Acids Res.* **15**, 8530–8545.
35. Diekmann,S. (1992) *Methods in Enzymology* **212**, 30–46.
36. Wu,H.M. and Crothers,D.M. (1984) *Nature*, **308**, 509–513.
37. Avila,J., Montejo de Garcini,E., Wandosell,F., Villasante,A., Sogo,J.M. and Villanueva,N. (1983). *EMBO J.*, **2**, 1229–1234.
38. Jamrich,M., Warrior,R., Steele,R. and Gall, J.G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3364–3367.
39. Eckdahl,T.T. and Anderson,J.N. (1990) *Nucleic Acids Res.* **18**, 1609–1612.
40. Muzard,G., Théveny,B. and Révet,B. (1990) *EMBO J.* **9**, 1289–1298.
41. Boffelli,D., de Santis,P., Palleschi,A., Risuleo,G. and Savino,M. (1992) *FEBS*, **300**, 175–178.
42. Zinkel,S.S. and Crothers,D.M. (1987) *Nature* **328**, 178–181.
43. Goodsell,D.S., Kopka,M.L., Cascio,D. and Dickerson,R.E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2930–2934.
44. Calladine,C.R. and Drew,H.R. (92) in Understanding DNA. Academic Press.
45. Vollrath.D. and Davis.R.W. (1987) *Nucleic Acids Res.* **15**, 7865–7876.
46. Marilley,M. and Gassend-Bonnet,G. (1989) *Exp. Cell res.* **180**, 475–489.
47. Chu,G., Vollrath,D. and Davis,W.R. (1986) *Science* **234**, 1582–1585.
48. Kabsch,W., Sander,C. and Trifonov,E.N. (1982) *Nucleic Acids Res.* **10**, 1097–1104.