The concept of sequence-dependent deformational anisotropy of DNA proposed earlier is further elaborated and a computational procedure is developed for the sequence-directed mapping of the nucleosomes along chromatin DNA nucleotide sequences. The deformational anisotropy is found to be nonuniform along the molecule of the nucleosomal DNA, suggesting that the DNA superhelix in the nucleosome is slightly oval rather than circular in projection. The number of superhelical turns in the nucleosome core particle is estimated to be 2.0 ± 0.2. Preliminary mapping of the nucleosomes in various chromatin DNA sequences yields the distribution of linker lengths which shows several minima separated by about 10 base-pairs. This is explained by sterical exclusion effects due to overlapping of the nucleosomes in space when some specific linker lengths are chosen. The mapping procedure described is tested by comparing its results with all the most accurate experimental mapping data reported so far. The comparison demonstrates that the exact positions of all the nucleosomes appear to be determined exclusively by the nucleotide sequences.

INTRODUCTION

A distinctive feature of chromatin is its repeat structure. The repeat unit consists of the nucleosome core particle (histone octamer and 145 base-pairs of the nucleosomal DNA) and linker DNA of variable size, up to about 100 base-pairs (see reviews:1,2). The linker region is presumably occupied by histone H1 and nonhistone proteins. Mild micrococcal nuclease digestion of chromatin yields double-stranded fragments of DNA, multiples of the repeat length - nucleosomal plus linker DNA. The repeat is usually considered to be a constant, though different for different species due to variable length of the linker.

An immediate question arises whether the histone octamers of the nucleosomes are attached to specific positions (sequences) along the chromatin DNA and, if so, what are the mechanisms which ensure the specific placing of the nucleosomes. According to the known experimental evidence the distribution of the histone octamers along the sequences seems to be of dual character demonstrating both random and sequence-specific behaviour. 3,4
A novel approach to the problem has been proposed based on the recently discovered weak periodicity of the chromatin nucleotide sequences with the period equal to the pitch of DNA. One possible explanation of this periodicity involves the hypothesis that adjacent base-pairs in the DNA molecule are not always parallel due to slight differences in stereochemistry of different combinations of neighboring base-pairs. Some of the combinations might serve as "wedges" inclining the DNA axis. The wedges repeated at distances multiple of the helical repeat of DNA should result in unidirectional curving of the DNA molecule. Such curved regions of the DNA might be energetically preferable for smooth bending in the nucleosome. It is noteworthy that the "wedge" might arise actually only under bending stress. The terms "wedge" and "curving" are used here for the sake of simplicity though there might be some regions of the DNA molecule which are actually straight but preferentially bendable in certain directions.

The periodicity of the dinucleotides found in the eukaryotic DNA might well be of a different nature, not related to the bendability of DNA. Nevertheless, here and further on we are using the deformational terminology thus favoring the hypothesis.

The sixteen different dinucleotides periodically repeating along the sequences have been found to form a very specific complementarily symmetrical pattern. The pattern has been interpreted as a reflection of nonequivalence of the sixteen different kinds of wedges. In particular, these different wedge-like combinations of the base-pairs seem to incline the DNA axis in different directions. The phase relationships found (matrix of bendability) might mean that the chromatin DNA sequences are specifically biased to orient all the wedges preferentially in a certain direction thus bringing about unidirectional curving of some portions of the molecule. Any nucleotide sequence in which the dinucleotides are distributed like in the matrix of bendability, therefore, should belong to the curved (deformationally anisotropic) piece of DNA molecule.

The purpose of the present work is a further elaboration of the concept of deformational anisotropy of DNA with the final target to develop the nucleosomal DNA bendability probe for locating the histone octamer binding sites in chromatin DNA. The mapping probe obtained allows for some very specific conclusions about the nucleosome geometry and higher order structure of chromatin.

The experimental mapping of the nucleosomes is usually based on combination of chromatin digestion by micrococcal nuclease, DNase I, exonuclease III and restriction enzymes. To achieve high accuracy of determination of
the positions of the nucleosomes along the DNA molecule one has to combine these different kinds of nuclease digestion. The accuracy of ±5 or less base-pairs was obtained in the experimental mapping of the nucleosomes in the green monkey a-satellite chromatin,\textsuperscript{8,9} rat-satellite chromatin,\textsuperscript{10} in reconstituted complexes involving fragments of 5S rRNA of \textit{X. laevis}, of the plasmid pBR322\textsuperscript{11} and of the lac-region of \textit{E. coli}.\textsuperscript{12} Our computational mapping data are compared with these most accurate experimental ones and the results are discussed below.

RESULTS AND DISCUSSION

\textbf{Matrix of bendability}

The iterative procedure of derivation of the complementarily symmetrical periodical pattern of the dinucleotides typical for the chromatin nucleotide sequences (matrix of bendability) is described in detail elsewhere\textsuperscript{5,6} (see also Methods). Essentially, it is the pattern recognition algorithm designed to find whether there are any archetype sequences which have a tendency to be periodically repeated with the period equal to the helical repeat of the chromatin DNA. The pattern found is shown in Fig. 1 where the distributions of sixteen dinucleotides within the period are presented rather than the repeating archetype sequences themselves. One can read the typical nucleotide sequences of the curved portions of the DNA molecule from this set of functions by choosing most frequently encountered dinucleotides for the corresponding positions along the period-long piece of the molecule. The sequences presented on the top of Fig. 1 are examples of many possible sequences. The distributions shown in Fig. 1 correspond to relative contributions of each of the sixteen dinucleotides to different positions within the averaged repeating pattern. It is the similarity to the pattern which is actually periodically repeating rather than any particular short sequence.

To derive the pattern only one of two strands of DNA is considered. Since the pattern found is complementarily symmetrical, the other strand (read also in 5' to 3' direction) is characterized by the same repeating pattern.

\textbf{Bendability function}

The matrix obtained can be used to locate unidirectionally curved pieces of a DNA molecule as soon as its nucleotide sequence becomes known. For this purpose one should compare short pieces of nucleotide sequences (12 bases in this case, see Fig. 1) with the matrix of bendability estimating the degree of similarity of these short pieces to the matrix. The similarity would be maximal if every particular dinucleotide of the 12 base long sequence is in register with the maximum of the distribution of the same dinucleotide within
Fig. 1: Matrix of bendability. The lines of the matrix are presented as relative occurrences of corresponding dinucleotides at different positions within the 10.5 base period (positions 0.75 to 11.25 bases). Only oscillating components are shown. The accuracy of calculation is limited by the total length of the analyzed sequences and equals about 0.005. The sequences typical for curved pieces of DNA are shown on the top of the figure.

the 10.5 base period, as given by the set of functions in Fig. 1. (Note that the sequence of 12 bases contains 11 dinucleotides covering 11 interbase distances.) As the measure of the similarity we have chosen the correlation coeff-
cient $C_b$ (see Methods for exact definition) which equals 1 or -1 for most anisotropic pieces depending on the direction of their curving. Pieces of intermediate anisotropy are characterized by coefficients $C_b$ of smaller absolute values.

Thus, to locate the curved pieces of a long DNA molecule the coefficient $C_b$ should be calculated for all positions of the nucleotide sequence by moving the matrix along in steps of one base each. As a result the function $C_b(N)$ is generated where $N$ is the nucleotide number of the long sequence analyzed. It is more convenient, however, to use the function $F_b(N) = C_b(N-5)$ which places the dinucleotide number $N$ in the center of the matrix. The function $F_b(N)$ is called the "bendability function".

An example of the bendability function calculated for two repeats of green monkey $\alpha$-satellite DNA sequence is presented on the top of Fig. 2. The function exhibits multiple maxima every one of which corresponds to the middle point of unidirectionally curved 12 base-pair piece of DNA which correlates well and is in phase with the bendability matrix.

Are there any longer pieces of DNA which are unidirectionally curved? What would then be the bendability function of such a piece? Consider, first, an ideally curved portion of DNA built of one of the ideal sequences like the ones shown on the top of Fig. 1. If several 10.5 base long curved pieces were positioned tandemly, the corresponding bendability function would show several maxima centered at the midpoints of ideal pieces and separated by 10.5 bases. The half-period shift from these positions would result in anticorrelation with the matrix. Thus, the bendability function of such an ideally curved portion of the DNA molecule should be an ideal periodical function, with the period 10.5 bases. More specifically, it should be a perfect sine (cosine) wave.

Indeed, if the correlation with the bendability matrix is the measure of curvature of the DNA axis in some specific direction, then the translation of the matrix along the sequence is equivalent to rotation of the curved piece around the DNA axis. The projection of the vector of curvature in the direction specified by the matrix would then oscillate sinusoidally, with period equal to the helical repeat of DNA.

The function $F_b(N)$ shown in Fig. 2 looks as if it does not have any such periodical regions. This only means, however, that there are no ideally curved pieces. There might be some regions with a more complicated trajectory of the DNA axis which are, nevertheless, on the average effectively curved. This, too, would be advantageous for the smooth bending of DNA in the nucleosome. To find such pieces one has to decompose the bendability function $F_b(N)$ within the
Fig. 2: Steps of location of curved portions of DNA in the green monkey α-satellite chromatin. The region of the bendability function $F_b$ between positions 153 (first repeat) and 126 (second repeat) has the biggest periodical component shown in the middle of the figure which is reflected by the maximum at position 54 in the mapping function $F_m$. The functions $F_b$ and $F_m$ are aligned to share the same coordinates shown at the bottom. The nucleotides of the sequence are numbered according to Ref.13.

specified interval of length $M$ into two components:

$$F_b(N) = a \cos(2\pi N / 10.5) + g(N),$$

where $g(N)$ is the nonperiodical component and $a$ is the amplitude of the periodical one (see Methods). The cosine wave which spans $M$ nucleotides can be considered as a bendability probe for location of $M$ nucleotide-long unidirectionally curved pieces for which the function $F_b(N)$ is known.

**Mapping function**

Calculating the amplitude $a(N)$ of the periodical component of the bendability function for all positions along the sequence, one could find the positions of the sequence which correspond to the most curved pieces of DNA.
In the middle of Fig. 2 the cosine wave is shown, spanning $M = 145$ bases, which corresponds to the periodical component of the bendability function $F_b(N)$ in the region from position 153 of the first repeat to position 126 of the second one. This part of the repeating sequence according to our calculations belongs to the most curved portion of the satellite DNA.

For the mapping of such curved pieces the function $F_m(N) = a(N-M/2)$ is more convenient. It places the $N$-th dinucleotide in the middle of the bendability probe of size $M$. The function $F_m(N)$ is called the "mapping function". An example of this function is given in Fig. 2, bottom. Here the same 145 nucleotide-long cosine probe is used for calculation. The function exhibits many maxima of which the one at position 54 is the highest. Therefore, the position 54 should correspond to the midpoint of the nucleosomal DNA of the $\alpha$-satellite chromatin (see also below).

An alternative way to locate the most bendable portions of the DNA molecule is to calculate the correlation coefficient $C_m$ between the bendability function $F_b(N)$ and the bendability probes (see Methods). This approach is more appropriate when noncosine bendability probes are considered (see below). Both approaches, however, lead to practically identical results.

**Mapping of nucleosomes**

As soon as the nucleosomal DNA bendability probe is chosen, one can try to map the nucleosomes along the chromatin DNA sequences. This would then provide predictions about the locations of nucleosomes which can be checked experimentally. In the computational mapping procedure used in this work, the biggest maximum of the mapping function is found first and the first putative nucleosome is placed, thus, excluding about 290 bases where no other nucleosome could be centered. Then the next strongest maximum is found and new 290 bases are excluded. This procedure is repeated many times until no more room is left to locate another nucleosome.

23 different chromatin DNA sequences have been subjected to this mapping procedure (see Methods). As a result 250 putative nucleosomes are located. In a few cases the computational mapping can be quantitatively compared with experimental mapping data, as described in the last sections.

**Nucleosomal DNA bendability probe**

For correct mapping of the nucleosomes the length of the DNA molecule smoothly bent in this particle has to be known. There are two possibilities: approximately 145 base-pairs which correspond to DNA folded in the nucleosome core particle, or 166 base-pairs of the "chromatosome". We have chosen the
shorter length of the bendability probe since this length of DNA is involved in the nucleosome structure in any case.

As we have found (data not shown), the period of the bendability probe is not very important as soon as it is sufficiently close to the pitch of DNA in the nucleosome. The pitch has been estimated to be \(10.3 - 10.4\) base-pairs per turn, very close to the period \(10.5 \pm 0.2\) bases discovered in the chromatin nucleotide sequences. The period 10.5 bases was chosen for the probe, the figure being more convenient for calculations.

The bendability probe has to be symmetrical to reflect the symmetry of the nucleosomal DNA and the nucleosome core particle. Therefore, the cosine probe is chosen rather than the sine wave. The midpoint of the cosine probe could be taken either positive (+ cos) or negative (- cos). For the given matrix of bendability (Fig. 1) the latter possibility is proven to be more appropriate since the -cos probe results in a better fit of the computational and experimental mapping data.

Thus, the first nucleosomal DNA bendability probe is chosen to be a -cosine wave with the period 10.5 bases, spanning 145 dinucleotides (146 nucleotides). The probe, \(P_0\), is shown in Fig. 2. For all the calculations discussed below the mapping function \(F_m(N)\) was based on the correlation \(C_m\) between the bendability function and the probe. Hence, the amplitude of the probe is irrelevant only its shape is of importance.

The 250 "nucleosomes" found by the mapping procedure with the probe \(P_0\) are described by 250 corresponding bendability functions \(F_b(N)\), \(1 < N < 145\). One could expect that the average bendability function of these "nucleosomes" would be very close to the probe \(P_0\) due to accumulation of the periodical components when the functions are summed. It is not exactly the case, however, as seen in Fig. 3A. The average bendability function appears as a symmetrically modulated wave rather than the original nonmodulated cosine.

What could the modulation mean? One possible explanation is that the nucleosomal DNA, although unidirectionally curved, has curvatures (bendabilities) varying along the molecule. The projection of the superhelix of the nucleosomal DNA on the plane perpendicular to the axis of the superhelix might be slightly oval rather than circular. That is, some positions of the nucleosomal DNA might be more important than the others in terms of the unidirectional bendability. Hence, the periodical average bendability function of the nucleosomal DNA should be modulated. The original bendability probe, therefore, has to be modified to reflect this feature.

The modification of the original probe \(P_0\) was accomplished by sequential
Fig. 3: Nucleosomal DNA bendability. A - the average nucleosomal DNA bendability function calculated by mapping with the cosine probe \( P_0 \) (see Fig. 2). B - the bendability probe \( P_1 \) obtained by the trial-and-error modification of the cosine probe maximizing its correlation with the chromatin nucleotide sequence. C - the average nucleosomal DNA bendability function calculated by mapping with the probe \( P_1 \). The amplitudes of the functions in A and C correspond to the nucleosomes of average "strength". For the ideal strongest nucleosome the main maxima would be of the amplitude 1.0.

Adding low amplitude random functions to the function \( P_0 \). These slightly modified probes were tested on the same set of chromatin DNA sequences. A modification was considered successful if it resulted in the increase of the average correlation coefficient \( C_m \) of the probe with the bendability functions of the "nucleosomes". This trial-and-error modification procedure after sequential addition of 40 low amplitude random functions and symmetrization resulted in the probe \( P_1 \) shown in Fig. 3B.

The probe \( P_1 \) also exhibits some modulation as expected. The average bendability function obtained after the mapping of nucleosomes of the same set of sequences using the probe \( P_1 \) is shown in Fig. 3C. Here the modulation is clearly pronounced. Considering the modulation of the bendability as reflection of variable curvature of the DNA trajectory in the nucleosome, one can estimate the length in base-pairs of one full superhelical turn of DNA in the particle. The sites of maximal curvature, i.e. the maxima of the modula-
Fig. 4: Distribution of the linker lengths between the nucleosomes located by the mapping procedure using the bendability probe P\textsubscript{1}. Only short linkers are presented.

Using the probe P\textsubscript{1} we have calculated the maps of the nucleosomes for all 23 chromatin DNA sequences analyzed (data not shown). The maps provide the linker lengths as well. The distribution of these lengths is shown in Fig. 4. The continuous curve is the result of smoothing the distribution by the running average of 5 adjacent positions. The distribution clearly shows at least three maxima separated by about 10.5 bases, thus indicating that some linker lengths are avoided, if not excluded. Specifically, the lengths up to about 5 base-pairs, from 10 to 15 and from 20 to 25 base-pairs are avoided. We consider this peculiar distribution as a reflection of the steric hindrances which the neighboring nucleosomes impose on each other at certain linker lengths. Let us take, for example, two hypothetical stacked core particles with zero linker length. The DNA path in this dimer is a continuous regular superhelix of about...
3.5 - 4.0 superhelical turns. How will the DNA path change if in place of the zero linker at the junction between two nucleosomal DNA's a linker of length, say 3 base-pairs, is inserted? It will result, first, in the translation of one nucleosome relative to the other by about 10 Å in the direction tangential to the DNA axis at the junction. Second, one of the two nucleosomal super-helices has to be rotated around the DNA axis at the junction point by about $34^\circ \times 3 = 102^\circ$. If the superhelix of the nucleosomal DNA and the double helix of DNA itself have opposite handedness (which is believed to be the case) then the rotation has to result in interpenetration of the nucleosomes which is of course impossible. That is, the linker length 3 is sterically forbidden, as well as lengths 2, 4 and 5 base-pairs all being shorter than the half-repeat of the DNA double helix. Similarly, the lengths about 10-15 and 20-25 should be excluded. For linker lengths higher than the size of the nucleosome (about 35 base-pairs) this "bumping" effect should vanish. The distribution of the linker lengths shown in Fig. 4 is in good agreement with computer calculations of possible DNA paths and the steric exclusion effects in the chromatin. Similar discrete distributions of the linker lengths with the period close to 10 bases have been observed experimentally as well. It was interpreted as $10n$ series in one case and as $10(n + 0.5)$ series in the others. If, however, the accuracy of the experiments is taken into account, both of these experimental series would fit as well to the formula $10.4(n + 0.75)$ which describes fairly well the positions of minima and maxima in the distribution calculated in this work.

This non-trivial result obtained by computer analysis of the chromatin DNA sequences inspires us with confidence that the nucleosome mapping procedure based on the sequence-dependent deformational anisotropy of DNA is essentially correct. Comparisons of the computational and experimental nucleosome mapping data are discussed in the following sections.

**Green monkey α-satellite chromatin**

The tandemly repeating DNA sequence of this satellite DNA was established by Rosenberg et al. Its repeat is 172 base-pairs long which is sufficient to adopt one nucleosome per repeat. The mapping function $F_m$ calculated for this repeating sequence is presented in Fig. 5. It has one major maximum, at position $54 \pm 1$. As discussed previously, this position should indicate where the nucleosome is centered. Recent experimental evidence shows that this satellite chromatin indeed has one nucleosome per every repeat. Accurate location of preferential position of the nucleosome within the repeating sequence has been obtained for Eco Rl subset of α-chromatin which has the nucleotide
sequence practically identical with the basic \(\alpha\)-satellite sequence.\(^9\) By combining results of digestion of this chromatin by restriction enzymes Eco RI, Hae III, Hind III and Mbo II, the right end of the nucleosomal DNA has been found\(^9\) to coincide with major micrococcal nuclease cutting site at position 122 ± 2 (126 ± 5 according to Ref.8). The center of the nucleosome resides, therefore, at position about 51, to compare with our position 54.

**Rat satellite chromatin**

The nucleotide sequence of this tandemly repeating satellite DNA was determined by Pech et al.\(^{20}\) The corresponding mapping function calculated for the sequence is shown in Fig. 6. This function displays many maxima of which the ones at positions A1 (345), B1 (242), A2 (159) and B2 (68) are most prominent. Two alternative ways of placing the nucleosomes in the rat satellite chromatin are possible according to computational mapping: two nucleosomes centered at A1 and A2, or two nucleosomes with centers at B1 and B2. Experimental mapping by combining results of digestion of this chromatin by restriction enzymes Hae III, Hind III and exonuclease III indicates that the nucleosomes in this chromatin are centered at positions 347 and 163 (173) ± 5 bases.\(^{10}\) Within the error bars these locations coincide with positions A1 and A2 found by the computational mapping procedure. After this work was completed, two new nucleosome positions were found experimentally in the rat satellite chromatin (T. Igo-Kemenes and H. Seligman, personal communication). Their coordinates correspond to the other two predicted positions, B1 and B2.
Fig. 6: Rat satellite chromatin mapping function. The nucleotides are numbered according to Ref.20. Some restriction sites are indicated. Small arrow A1 indicates position of the major maximum. If the nucleosome is centered at this position then one more nucleosome can be located at local maximum A2, others being spatially excluded. An alternative map is indicated by the arrows B1 and B2. Big arrows on the top correspond to experimentally found positions.10 Their width indicates the accuracy of the experimental estimates.

Reconstituted nucleosome, fragment Fnu-EcoRI (186 base-pairs) of X.lasvis 5S rRNA gene

The nucleotide sequence of the 5S rRNA gene of X.lasvis has been established by Miller et al.21 The Fnu-EcoRI 186 base-pair long fragment of the gene was found recently to bind the histone octamer in a sequence-specific manner leaving short ends of free DNA.11 The mapping function corresponding to the sequence is presented in Fig. 7. It spans only the middle part of the sequence since our mapping procedure is applicable only to the cases when the histone octamer contacts with full size nucleosomal DNA (145 base-pairs). The biggest maximum of the mapping function is found at the position 96g. Experimentally, the center of the nucleosome reconstituted from the histones and these 186 base-pair long fragments of DNA is found by Dnase I and exonuclease III digestion to be located at position 95g ± 2.11

Reconstituted nucleosome, fragment Fnu-EcoRI (254 base-pairs) of plasmid pBR322

The mapping function calculated for this fragment has one major maximum at position 4235 (Fig.8). According to exonuclease III and DNase I digestion data for the nucleosome reconstituted on this fragment, the midpoint of the
Fig. 7: Nucleosome reconstituted on *X. laevis* 5S rRNA gene fragment. The nucleotides are numbered according to Miller *et al.* The 5S rRNA gene spans positions 1g - 120g. Small arrow indicates position of the major maximum of the mapping function. Bigger arrow points to the position of the midpoint of the nucleosome found experimentally.\textsuperscript{11}

nucleosomal DNA in the reconstitute is also found at the position 4235±3.\textsuperscript{11}

*Reconstituted nucleosomes, 203 base-pair E.coli lac control region*

The mapping function calculated for the nucleotide sequence of this fragment has several comparatively small maxima (Fig. 9) of which the maximum A (position 78) and B (position 108) are the highest. Two alternative binding positions are found experimentally using this fragment for reconstitution with

Fig. 8: Nucleosome reconstituted on pBR322 DNA fragment. The bases of the Fnu-EcoRl fragment are numbered according to Sutcliffe. The mapping function has the main maximum at position 4235 (thin arrow). Experimentally found midpoint of the nucleosome in the histone-DNA reconstituted complex is shown by the arrow on the top. The width of the arrow equals to the experimental error.\textsuperscript{11}
Fig. 9: Two possible locations of the nucleosomes on reconstituted complex of histones with E. coli lac region DNA fragment. The nucleotide sequence used for the calculation of the mapping function $F_m$ was taken from Refs. 12, 23 and 24. Computed (this work) and experimentally found positions of the nucleosomes are indicated by thin and thick arrows, respectively.

histones. These positions (76±5 and 108±5) are obtained from the exonuclease III digestion data and from partial sequencing of the free DNA ends. In this case, too, the positions predicted from the nucleotide sequence analysis fit well to the experimentally found positions of the nucleosomes.

Sequence-specificity of the nucleosomes

The results show that for all nine nucleosomes experimentally mapped in natural and reconstituted chromatin, the corresponding mapping functions have their major maxima centered not farther than four bases from the midpoints of the nucleosomes. Since the discrepancies observed are well within the experimental error, the mapping procedure proposed appears to be quite sufficient to locate the nucleosomes exclusively on the basis of the nucleotide sequences. In other words, the positions of the histone octamers bound to the DNA molecules appear to be completely sequence-dependent.

The mapping procedure described is based on the specific distribution of the dinucleotides characteristic for the nucleosomal DNA sequences (matrix of bendability). It provides, therefore, a typical example of distributional recognition where the nucleotide sequence as such is not important for the recognition, but rather some sequence-dependent structural or physical features of the DNA molecule. In this case the “wedges” provided by certain combinations of adjacent base-pairs appear to be essential.

The fact that the mapping procedure derived from the analysis of chromatin DNA sequences is able to locate histone-octamer binding regions in prokaryotic DNA (see above examples) might give rise to some confusion. The prokaryotic sequences should not contain any signal related to chromatin structure. One has
to take into account, however, that even random sequences contain such "signal" regions recognized by the mapping procedure as "nucleosomes". But the amount of such regions and the amplitudes of their periodical components, on average, are smaller than in chromatin DNA sequences. The axis of a DNA molecule with random nucleotide sequence should appear as a limited random walk trajectory with many curved portions.

The amplitudes of the maxima of the mapping functions vary between 0.1 and 0.5 reflecting the similarity of the nucleosomes found to the standard (average) nucleosome. The amplitudes might also reflect the relative physical "strength", i.e. stability of nucleosomes. Unfortunately, the stability as well as the curvature of the DNA axis cannot be estimated from the amplitudes of the mapping peaks since the absolute wedge angles between different combinations of adjacent base-pairs are not yet known.

Inspection of the curves presented in Figs. 5-9 reveals that the major maxima are always surrounded by somewhat smaller ones which, in theory, might also correspond to the nucleosomes occupying these alternative positions. Only in one case (Fig.6) this expectation is supported experimentally. Another example of this kind is provided by the alternative locations of the Bam I nucleosome of SV 40[26,6] separated by about ten base-pairs one from another. There might exist some additional features of the sequences, not only the 10.5 base periodicity, which discriminate between several alternative positions. Further, more profound analysis of the nucleosomal DNA sequences is necessary to find these features and to sharpen the existing algorithm for the nucleotide sequence-directed mapping of the nucleosomes.

The fit of the computed positions of the nucleosomes to the experimentally found ones provides strong support to the original idea about sequence-dependent deformational anisotropy of DNA and its relation to chromatin structure[7,5,6]. The existence of the hypothetical intrinsic curvature of DNA molecules has been recently demonstrated by electric dichroism studies[27,28]. The nucleosomal DNA appears to be curved indeed[27] as well as certain non-chromatin DNA in which case some sequence periodicity is also found[28].

METHODS

Computation of the matrix of bendability (Fig.1) was accomplished essentially as described earlier[5,6] with modifications as follows. First, the iteration procedure used in the present work was designed to generate a pattern which is both periodical and complementarily symmetrical. Second, only oscillating components are taken from every line of the matrix of bendability and
every line is normalized dividing it by the total sum in the column rather than by the average in the line. As a result the amplitudes of the functions (lines) of the final bendability matrix are proportional to the absolute contributions of the dinucleotide to the oscillating component of the matrix of bendability.

The chromatin DNA nucleotide sequences used for the calculations are the following: SV 40 genome, polyoma virus genome, ovalbumin gene, fibroin gene, histones H2B and H3 genes, BKV virus genome, X. laevis rRNA genes, yeast 2μ plasmid, mouse immunoglobulin μ gene, and μ(γ2b) recombination region.

The correlation coefficient \( C_{b} \) which describes the similarity of particular 11 dinucleotide (12 nucleotide) long sequence to the matrix of bendability is calculated as follows:

\[
C_{b} = \left( \sum_{i=1}^{11} F(D_{i}) \right) / \sum_{i=1}^{11} \max_{i}(F)
\]

where \( D_{i} \) is the \( i \)-th dinucleotide of the short sequence and \( F(D_{i}) \) is the corresponding value of the normalized distribution function for the dinucleotide in the \( i \)-th column of the bendability matrix. The denominator is the sum of 11 maximal values of the distribution functions for 11 columns of the matrix. The number of the columns (11) of the bendability matrix is the closest integer to the chosen period of chromatin DNA sequences, 10.5 bases. Any number of columns could be taken depending on the size of the short sequence analyzed. In any case, however, this should be the same 10.5 base period-wide basic matrix repeated a number of times, not necessarily integer.

The bendability functions \( F_{b}(N) \) are calculated as described in the main text. The amplitude \( a \) of periodical component of the function \( F_{b}(N) \) is calculated as follows:

\[
a = \left( \sum_{k=1}^{M} F_{b}(k) \cos(2\pi k/10.5) \right) / \sum_{k=1}^{M} \cos^{2}(2\pi k/10.5)
\]

Here \( M \) is the length of the DNA sequence for which the average curvature is calculated. The correlation coefficient \( C_{m} \) for the function of bendability \( F_{b}(N) \) and the probe is calculated as follows:

\[
C_{m} = \left( \sum_{k=1}^{M} F_{b}(k) \cos(2\pi k/10.5) \right) / \sqrt{\sum_{k=1}^{M} F_{b}^{2}(k) \sum_{k=1}^{M} \cos^{2}(2\pi k/10.5)}
\]

For the trial-and-error modification of the periodical nucleosomal DNA bendability probe random functions with white spectrum were computer-generated.
The amplitudes of the functions added to the probe were adjusted to be less than 0.1 of the amplitude of the probe. Two independent sets of nucleotide sequences were subjected to this procedure using two sets of random functions, independent as well. The modulation of the probe obtained is found to be practically the same in both cases. The average pattern is shown in Fig. 3B.

For mapping of the nucleosomes the sequences listed above as well as the following additional nucleotide sequences were used: 70K heat shock gene of drosophila, β-globin genes of rabbit, mouse, and man, human γ-globin, δ-globin, ε-globin genes, yeast actin gene, yeast rRNA genes, human insulin gene, mouse immunoglobulin γ1 gene, κ light chain gene, and γ2b heavy chain gene.

For comparison of the computational mapping data with experimental ones the probe P₁ (modulated cosine probe Fig. 3B) was used for derivation of mapping functions.

The authors are grateful to G. Bolshoi for excellent technical assistance, to K. van Holde, H. Eisenberg and L. Ulanovsky for valuable discussions.

REFERENCES