Biophysical Chemistry, 39 (1991) 127–136

BIOCHE 01532

The curvature vector in nucleosomal DNAs and theoretical prediction of nucleosome positioning

D. Boffelli [a], P. De Santis [a], A. Palleschi [b] and M. Savino [a]

[a] *Dipartimento di Chimica, Dipartimento di Genetica e Biologia Molecolare, Università di Roma 'La Sapienza', P.le A. Moro 5, 00185 Roma and* [b] *Dipartimento di Chimica, Università della Basilicata, via N. Sauro 85, 85100 Potenza, Italy*

Received 17 April 1990
Revised manuscript received 3 August 1990
Accepted 16 August 1990

Curvature vector; Nucleosome; DNA

Using our model for predicting DNA superstructures from the sequence, the average distribution of the phases of curvature along the sequences of the set of the 177 nucleosomal DNAs investigated by Satchwell et al. (J. Mol. Biol. 191 (1986) 659) was calculated. The diagram obtained shows very significant features which allow the visualization of the intrinsic nucleosomal superstructure characterized by two quasi-parallel tracts of a flat left-handed superhelical turn connected by a left-handed inflection in a perpendicular direction; such a superstructure appears to be closely related to the nucleosome model of Travers and Klug (Phil. Trans. R. Soc. Lond. 317 (1987) 537). The nucleosomal curvature phase diagram was then adopted as a sensitive determinant for the nucleosome virtual positioning in DNAs via correlation function, obtaining a good agreement with the experimental mapping of SV40 regulatory region as recently investigated by Ambrose et al. (J. Mol. Biol. 209 (1989) 255). This analysis shows also the presence of a constant phase relation between the virtual nucleosome positions which suggests its possible implication in the nucleosome condensation in chromatin.

1. Introduction

The relevance of the sequence-dependent curvature for the regulatory regions of DNAs is becoming a basic concept in understanding gene regulation.

Curved DNAs have in fact been found to be associated with the binding sites of many regulatory proteins, and shown to be involved in the mechanisms of recombination, replication and transcription [1–10].

The nucleosome is the association complex of DNA (~ 145 bp) with the histone octamer where the DNA axis is wrapped on the proteic core with an average curvature radius of about 43 Å. Whilst reconstitution of nucleosomal structures can be practically obtained with any DNA sequence, except some such as poly (A)/poly (T), a number of papers have shown the existence of preferential sites for DNA-histone association as well as the occurrence of nucleosome-free regions in chromatin [11–17].

The experimental evidence indicates in fact that nucleosome positioning along DNA sequence occurs with different affinity and that it is regulated by cooperative and competitive effects. It is clear that the histone octamer is capable of distorting the DNA helical axis in a flat, more or less regular supercoil and this ability is enhanced by certain base sequences either because of specific interactions between the amino acid residues and certain base-pairs, or because those regions of DNA have

Correspondence address: P. De Santis, Dipartimento di Chimica, Università di Roma 'La Sapienza', P.le A. Moro 5, 00185 Roma, Italy.

an intrinsic propensity to wrap over the protein core, due to a high bendability, as first suggested by Trifonov [18,19] and/or to a degree of intrinsic curvature.

Using our theoretical model for predicting DNA superstructures from the sequence [20–25], this paper investigates the relations between intrinsic curvature and nucleosome positioning, starting from the analysis of the 177 nucleosomal sequences obtained by Satchwell et al. [26]. The main result of this analysis is that the phase distribution of the curvature vector along the nucleosomal sequences is a sensitive determinant for locating the virtual nucleosome sites on DNAs.

2. The curvature features of 177 nucleosomal DNAs

We have previously shown [20–25] that it is possible to predict with a good approximation the superstructure of DNA tracts by integrating the theoretical deviations d, as obtained by conformational energy calculations, of the 16 different dinucleotide steps from the canonical B-DNA structure; these are represented by the following hermitian complex matrix (in degrees):

d	A	T	G	C
T	(8.0, 0.0)	(−5.4, −0.5)	(6.8, 0.4)	(2.0 −1.7)
A	(−5.4, 0.5)	(−7.3, 0.0)	(1.0, 1.6)	(−2.5, 2.7)
C	(6.8, −0.4)	(1.0, −1.6)	(4.6, 0.0)	(1.3, −0.6)
G	(2.4, 1.7)	(−2.5, −2.7)	(1.3, 0.6)	(3.7, 0.0)

the real and imaginary parts being the roll (ρ) and the tilt (changed in sign) ($-\tau$) angles, respectively, of the dinucleotide step: $d = \rho - i\tau = (\rho, -\tau)$. It is worth noting the dominant role of the roll angles which satisfactorily agree with the values that characterize the crystal structures of the double-helical dodecamers investigated by X-ray analysis [23,24].

The curvature at the sequence number n is given, per turn of B-DNA, in modulus and phase by the complex function:

$$C(n) = \nu^0/g \sum_{s=n-g/2}^{n+g/2} \mathrm{d}(s) \exp(2\pi i s/\nu^0)$$

where ν^0 is the average periodicity of B-DNA 10.4, g the integration grid and $\mathrm{d}(s)$ the local deviation pertinent to the dinucleotide step at the sequence number s.

The modulus of curvature $[C(n) \cdot C^*(n)]^{1/2}$ represents the change of the helical axis after a turn of the double helix; it is equal to $34/R(n)$ (Å) where $R(n)$ is the curvature radius in Å. The phase of curvature is given by $\tan^{-1}(Im(C(n))/Re(C(n)))$ and represents the direction of the curvature vector with respect to the pseudo dyad axis of the first dinucleotide step of the sequence; following the recent standard convention the positive direction is that centrifugal.

We demonstrated in previous papers that this formulation is equivalent to the Taylor series first order of the matrix transformations representing the angular changes of the local helical axes.

In order to reduce the noise it was found convenient to adopt a value of g corresponding to three turns of B-DNA (31 bp; s changing accordingly between $n-15$ and $n+15$ sequence numbers).

It is interesting to note that the same modulus of curvature is obtained using the sequence of one or the other strand in the correct 5′-3′ direction, whereas the phase must be changed in sign in order to have the same dyad axis as origin of the phase angles.

Such a representation of the curvature satisfactorily accounted for the experimental evidence of all the multimeric oligonucleotides as well as for the tracts of natural DNAs investigated by different authors as illustrated in our previous papers (refs 2,3,6,10,27–34 and M. Plhol et al., personal communication). A very critical test of the sensitivity of the model is the good correlation obtained in predicting the electrophoretic retardation changes in a tract of SV40 subjected to point mutations [35] as investigated by Milton et al. [36].

Recently we introduced slight changes in the local periodicity of B-DNA [25] in order to improve the correlation diagram between the theoretical and experimental electrophoretic retardation factors of 450 multimeric oligonucleotides differing in periodicity (9–21 bp), sequence and in molecular complexity (50–200 bp). The following symmetrical matrix Ω reports the local twist an-

gles for each dinucleotide step:

Ω	A	T	G	C
T	34.6	35.9	34.5	35.8
A	35.9	35.0	35.6	34.6
C	34.5	35.6	33.7	33.0
G	35.8	34.6	33.0	33.3

The formula of curvature now becomes:

$$C(n) = \nu^0/g \sum_{s=n-g/2}^{n+g/2} \mathrm{d}(s)\, \exp(is\Omega(s))$$

Using this model, the curvature diagrams of the 177 nucleosomal DNAs were evaluated and the diagram of the average curvature modulus obtained. This is shown in fig. 1, and is characterized by a variability of only 20% along the sequence and a large dispersion ($\sigma = 3.09$), indicating that the modulus of curvature is a rather weak determinant for nucleosome positioning along a DNA sequence.

In contrast, the average distribution of the curvature phases along the sequence of the 177 nucleosomal DNA evaluated by assigning the modulus of curvature as the statistical weight (in fact the phase is better determined as greater than the corresponding modulus) shows very significant features, as can be seen in fig. 2. This diagram was obtained from the corresponding histogram after a smoothing procedure followed by the self-consistent optimization of the phasing of the 177 sequences within the limited shift of ± 2 bp from the original register of Satchwell et al. (see table 1). In particular, the Fourier transform of the curvature-weighted phase histogram was calculated and multiplied with a Gaussian smoothing function in order to obtain through the inverse Fourier transform a resolution power of 2 bp along the sequence and 40° along the phase angle. Then the correlation of this function and each curvature function (phase and modulus) of the 177 nucleosomal DNAs was calculated summing up the products of the two functions for all the sequence positions. The calculation was repeated shifting the sequence of ± 1 and ± 2 bp. The whole calculation was then iterated adopting the phasing of the 177 nucleosomal DNA characterized by the best correlations until achieving self-consistence; it was obtained in four cycles. The resulting diagram is shown in fig. 3; the range of n is between 15 and 127 because of the integration grid of 31 bp used as already reported and the tailoring of the bp number at 142, corresponding to the minimum bp number of the set of Satchwell et al. which ranges between 146 and 142. Finally, the origin of the phase angles was shifted at the pseudo dyad axis of the central 71st dinucleotide step by subtracting to the phase angles $360° \cdot 71/\nu^0$.

The diagram is characterized by the presence of a pseudo center of inversion located at half sequence and by a practical invariance of the phase except for the central tract, where the phase splits about 90°.

In order to take into account that both the strands of DNA are involved in nucleosome formation, the diagram of fig. 2 was correctly symmetrized with respect to its inversion center at the 71st step (in fact the complementary strands have the same modulus and the phase changed in sign). This corresponds to the presence of a dyad axis in the resulting DNA superstructure coincident to the pseudo dyad of the 71st dinucleotide step (see fig. 3).

It is interesting to note that if only the contribution of AA and TT is considered, a very similar phase diagram will be obtained, as illustrated in fig. 4. This diagram is identical to the angular distribution of AA and TT steps around the DNA axis in the nucleosomal DNAs; the dominant value of approx. 0° of the curvature phase indicates that the majority of AA and TT steps faces in the opposite direction of the curvature vector (which points in the centrifugal direction according to the standard convention) given the negative sign of the pertinent deviation vectors: thus, in the case of AA and TT steps, the pseudo dyad axes of such dinucleotides have the propensity to be coplanar with the nucleosome dyad axis. This agrees with the results reported by Satchwell et al. [26], which indicated that the AA and TT sequences face in towards the nucleosome core.

However, the curvature phase diagram of the remnant sequences, calculated without the contributions of the AA and TT dinucleotide steps, also showed a similar pattern (see fig. 5). Thus, the

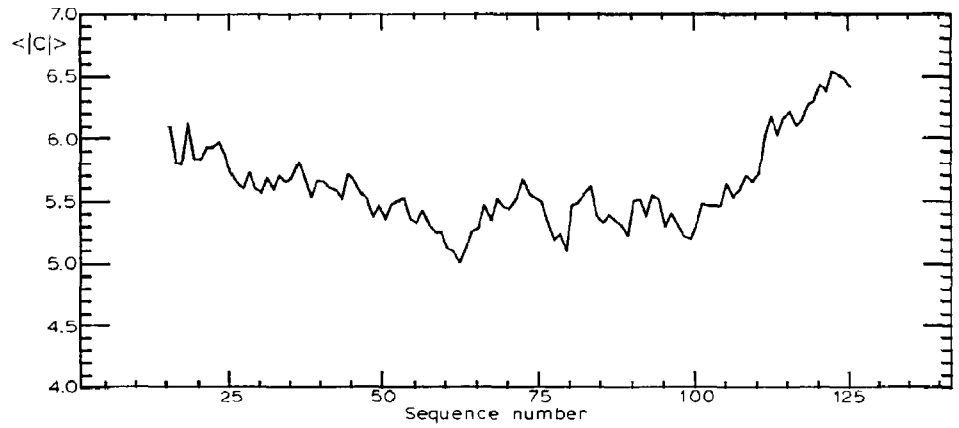

Fig. 1. Average curvature modulus distribution of the set of the 177 nucleosomal sequences investigated by Satchwell et al. [26] (standard deviation = 3.0).

diagram of the curvature phase distribution obtained in this paper shows that not only AA and TT steps, but also the majority of the dinucleotide steps cooperates in determining the curvature of the nucleosomal DNAs.

Adopting the maximum of the phase distribution for each dinucleotide step along the sequence as the corresponding curvature vector and using the pertinent matrix transformations, a model of the intrinsic virtual superstructure of nucleosomal DNA was obtained (fig. 6a and b); it is represented by the segmental chain of the local helical axis of recurrent turns of double helix, calculated using the values of the curvature phase distribution (the maximum value was set equal to 8°, corresponding to the maximum deviation of a dinucleotide step) as bond angles, the phase increments as torsional angles and a constant bond

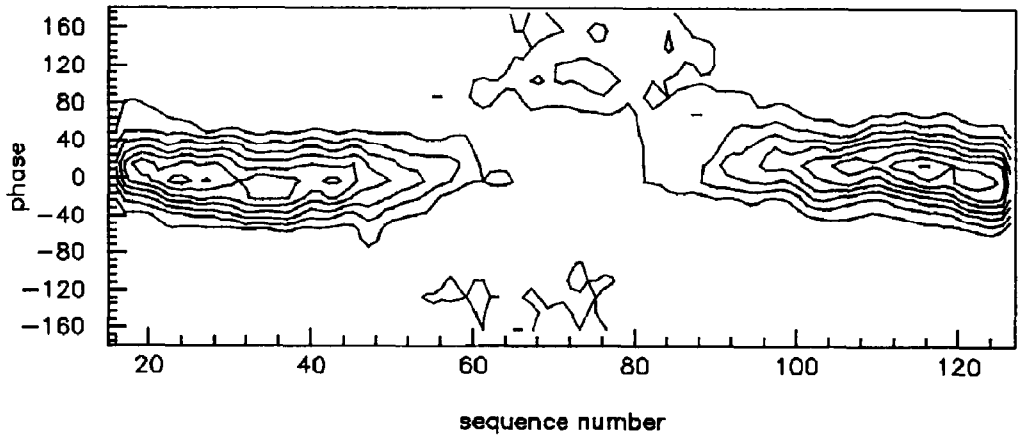

Fig. 2. Curvature phase distribution of the set of the 177 nucleosomal DNAs. Eight contour lines drawn at regular intervals from 20 to 100% of the whole range.

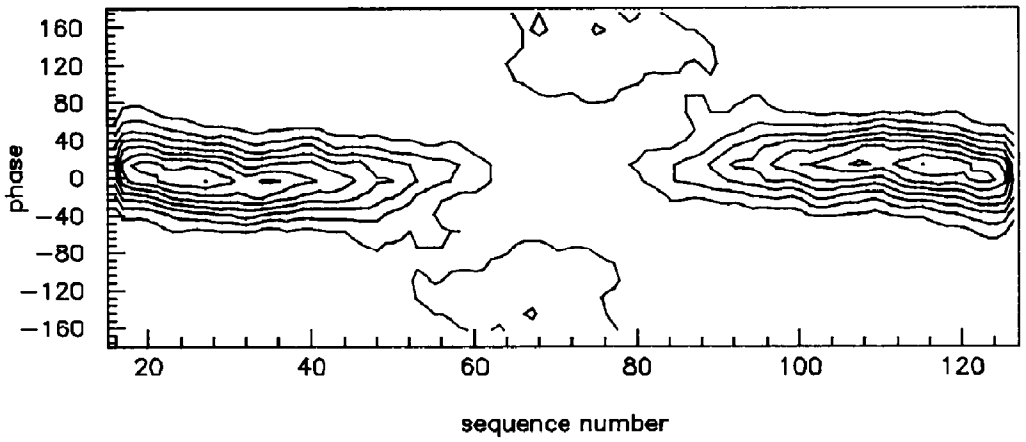

Fig. 3. Curvature phase distribution of the set of the 177 nucleosomal sequences and their complementary ones. Contour lines as in fig. 2.

length of 3.4 Å. Such an operation saves the phases namely the directions of curvature but enhances the values of curvature in proportion with the phase distribution; it should mimic the action of the histones in amplifying the intrinsic curvature as experimentally found for DNA-binding regulatory proteins [3,4,10,33,38–41]. In the case of nucleosomal DNAs, the histone cohesive forces should also provide the reversible packing of the two symmetric halves of the superstructure, as suggested by Dieterich et al. [42].

The model shows some characteristic features of that proposed by Travers and Klug [37] on the basis of different kinds of considerations and experimental evidence. It is characterized in fact by the presence of a dyad axis and three curved

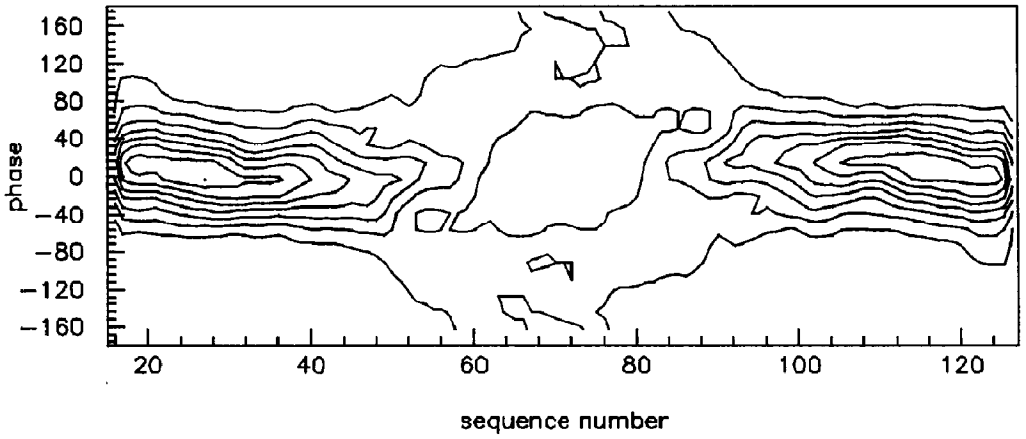

Fig. 4. Average phase distribution of AA and TT dinucleotide steps pseudo dyad axes of the set of the 177 nucleosomal sequences and their complementary ones. Contour lines as in fig. 2.

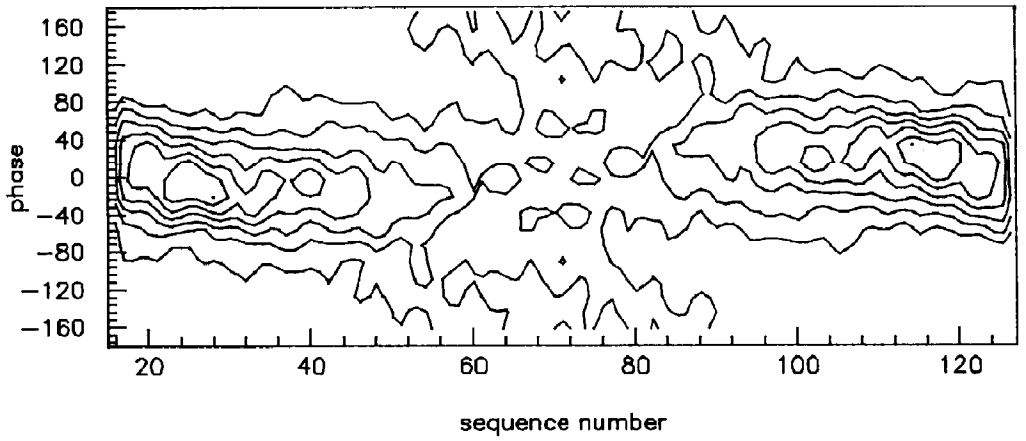

Fig. 5. Curvature phase distribution due to all but AA and TT dinucleotide steps of the set of the 177 nucleosomal sequences and their complementary ones. Seven contour lines drawn at regular intervals from 25 to 100% of the whole range.

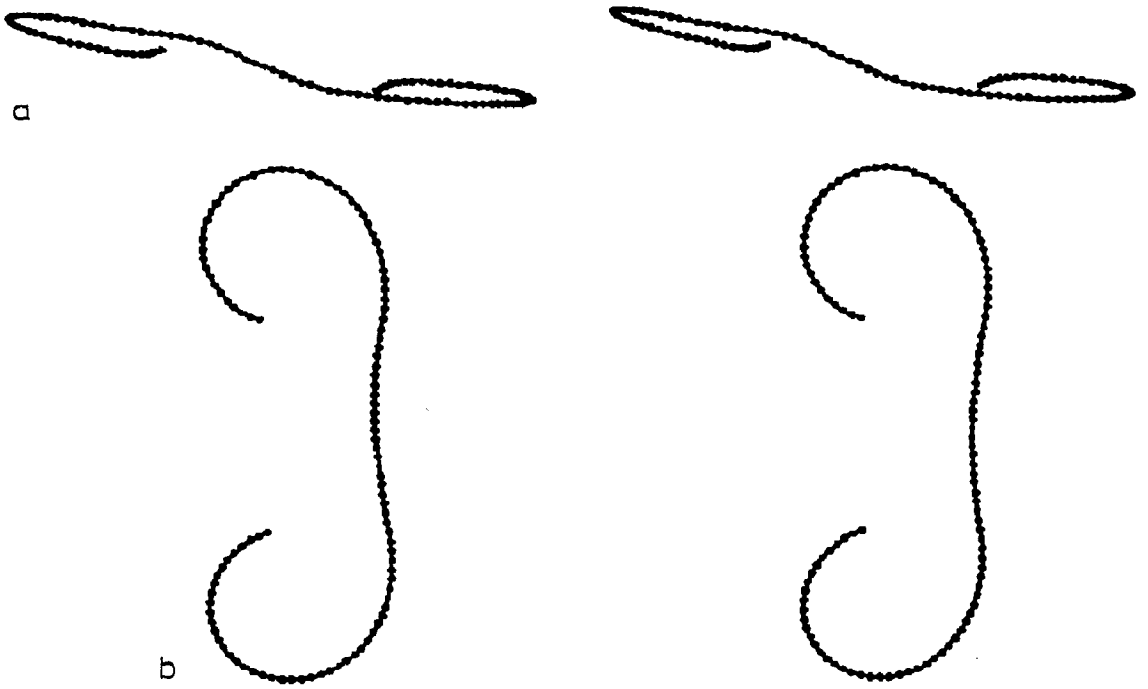

Fig. 6. Stereoprojections along (a) and perpendicular (b) to the dyad axis of the intrinsic nucleosomal DNA superstructure obtained from the maximum pathway of the curvature phase diagram of fig. 3.

regions: two quasi-coplanar with a slight left-handed pitch and one central with a curvature in a plane nearly perpendicular, characterized by a left-handed inflection (the Travers 'jog' [37]).

On the basis of these results we can assume the curvature phase diagram of fig. 3 as a good determinant for the virtual nucleosome positioning by localizing the maxima of its correlation function with the corresponding curvature diagrams of DNAs, as depicted in fig. 7. Such a function is easily calculated as the average product of the phase curvature diagram with the corresponding curvature diagram of an identically sized tract recurrent along the DNA sequence. The value of correlation is assigned to the central dinucleotide step of the DNA tract; therefore, the maxima of the correlation function indicate the virtual positions of the dyad axis of the nucleosome.

3. Nucleosome positioning in SV40 regulatory region

Adopting as determinant function the nucleosomal curvature phase diagram of fig. 3, but shifting the origin of the function at its average value (~ 30% of the variability range), we calculated the correlation factors (the covariance) of the single sequences of the 177 nucleosomal DNAs. All the 177 nucleosomal DNA, except very few cases, showed a positive correlation, as illustrated in fig. 8.

Thus, we tried to obtain the virtual positions of the nucleosomes in the case of the SV40 regulatory region (–275,650 sequence number) recently investigated by Ambrose et al. [43]; the diagram of the modulus of curvature is shown in fig. 9 and the resulting correlation diagram is shown in fig.

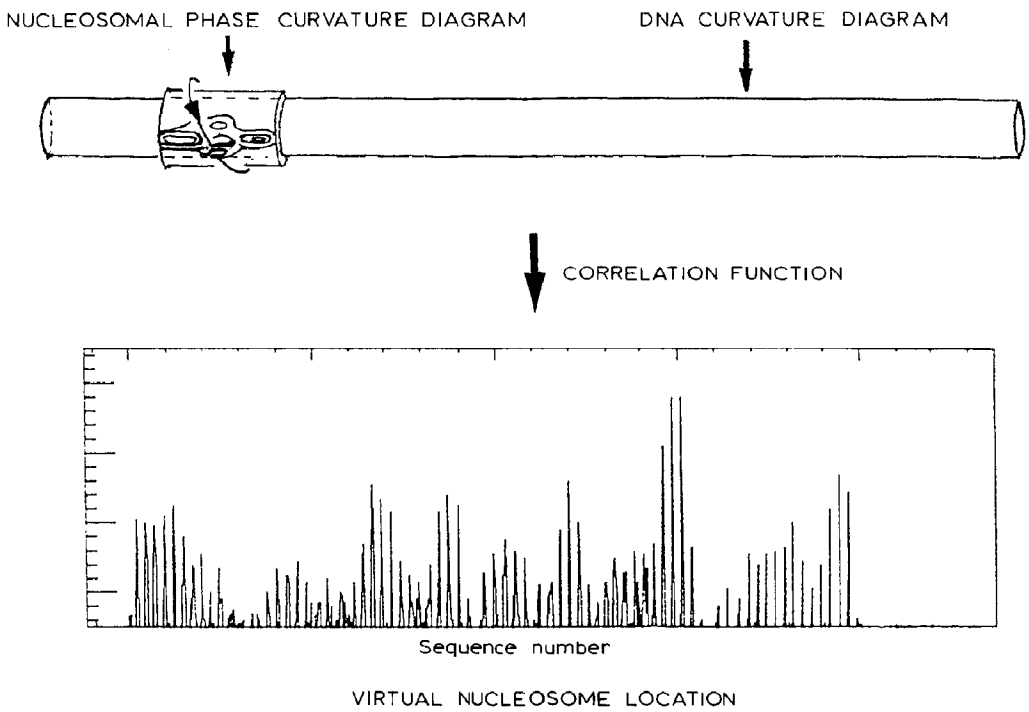

Fig. 7. Pictorial drawing of the derivation of the correlation function between the nucleosomal curvature phase distribution of fig. 3 and curvature diagram of a DNA fragment.

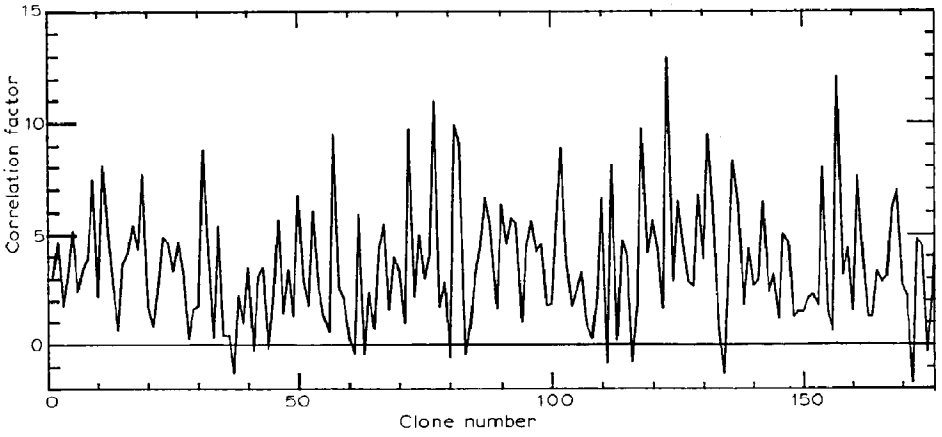

Fig. 8. Correlation factors of the set of the 177 nucleosomal DNAs with the curvature phase diagram of fig. 3 with the origin shifted at the average value.

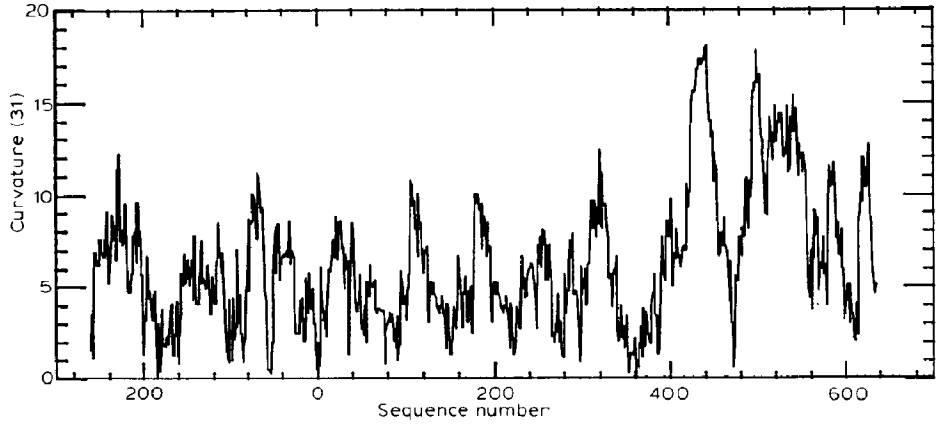

Fig. 9. Curvature modulus profile of SV40 regulatory region (−275,650): integration grid = 31 bp.

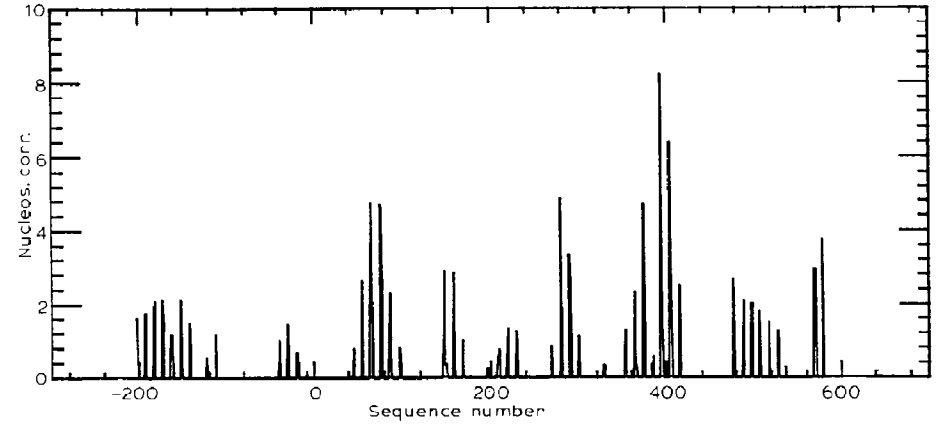

Fig. 10. Correlation function and nucleosome positioning along the SV40 regulatory region (−275,650).

Table 1

Sequence shifts for optimal phasing of the Satchwell et al. nucleosomal DNAs

0	2	0	1	1	1	−2	1	−1	1	10
2	0	2	1	2	1	2	−2	0	−2	20
2	0	2	2	1	1	−1	2	1	−1	30
−1	0	−1	0	−2	1	−2	1	−2	0	40
2	1	−1	2	2	−2	−2	−2	0	2	50
0	2	−1	−2	2	−2	1	−1	−2	2	60
2	0	−1	1	1	1	−1	0	0	−2	70
2	0	−1	−1	−2	−2	0	−1	−2	−2	80
−2	2	2	−2	−2	1	1	0	−2	0	90
1	−2	1	1	1	−2	0	0	1	−1	100
−2	2	0	1	1	−2	−2	−2	−2	1	110
2	0	−2	1	0	−2	1	2	−2	1	120
1	2	1	−2	2	0	1	−2	2	0	130
0	2	2	2	−1	2	2	2	1	1	140
1	2	2	2	−2	1	1	−2	2	1	150
−2	0	0	2	1	−1	0	−2	0	0	160
−1	−1	0	2	2	−1	0	0	0	0	170
−1	2	2	−1	−2	−2	−2				177

10. In the correlation calculations the curvature phase pattern of fig. 3 was used with the origin shifted at its average value plus the standard deviation (50% of the variability range), in order to localize the nucleosome with the highest affinity. As can be observed on comparison with the curvature diagram of fig. 9, high curvature is not a sufficient condition for nucleosome location.

The apparent feature of the correlation diagram is the recurrent periodicity corresponding to the helical symmetry of the B-DNA structure, which indicates the possibility of a limited nucleosome sliding along the same cylindrical sector of DNA. The maxima represent possible locations of the nucleosome dyad axis along the sequence and satisfactorily account for the experimental nucleosome map as obtained by smoothing, via Fourier transform, the original clone distribution of Ambrose et al. [43] (see fig. 11). In fact, the highest group of maxima of correlation coincides with the maximum of the clone abundance which indicates the preferential occurrence of nucleosome in that DNA region.

It should be noted that the maxima of the correlation function as well as of the experimental map occur at intervals less than the extension of nucleosomal DNA (> 145 bp) and then represent only possible sites for histones. The actual nucleosome distribution is therefore conditioned by both saturation and exclusion effects. This can explain the fact that certain regions such as the origin of replication of SV40 appear nucleosome free in some experiments when in others they seem to be involved in nucleosome formation.

There are in fact different possible patterns of nucleosome arrangements which are determined by the number of nucleosomes, the exclusion of nucleosome assemblies with distances less than 150–160 bp, and finally by the optimization of the

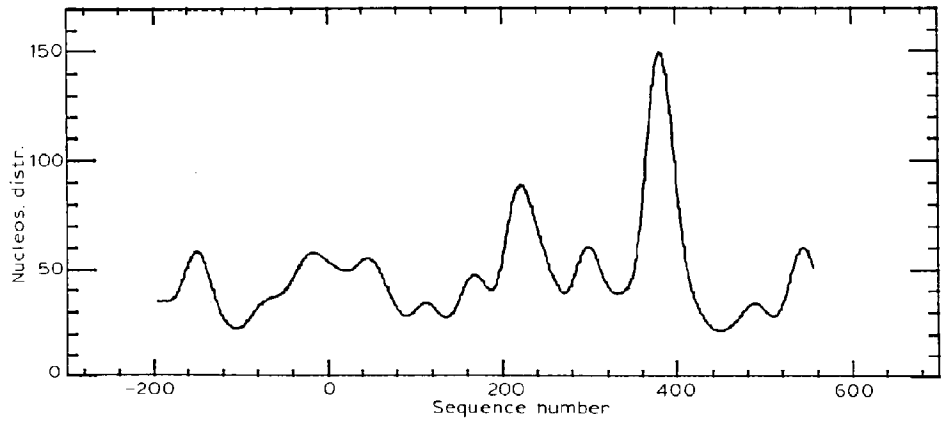

Fig. 11. Experimental nucleosome map of SV40 regulatory region (−275,650) obtained by smoothing the Ambrose et al. [43] clone frequencies.

global correlation with nucleosomal curvature features.

Finally, it could be of particular interest that the strong periodicity of 10.26 bp (evaluated via Fourier transform), which characterizes the correlation function of the regulatory region of SV40 in fig. 10, is strictly coherent over the whole range of about 800 bp, indicating that the assembling nucleosomes are in phase, which could be an important feature for their condensation in chromatin.

4. Conclusions

The results obtained are very promising for further applications of our model and eventually, its refinement; it should be noted that they are obtained by adopting the same successful model used in previous papers [20–25] in predicting the superstructures of DNAs and their experimental manifestations (electron microscopy visualization, gel electrophoresis, selective DNAse digestion, point mutations and X-ray structures of double-helical dodecamers).

Acknowledgments

We are grateful to Dr Sandra C. Satchwell for the listing of the 177 sequences. This research has been financially supported by MPI 40% funds.

References

1 W. Ross and A. Landy, J. Mol. Biol. 156 (1982) 523.
2 K. Zahn and F.R. Blattner, Nature 317 (1985) 451.
3 R.R. Koepsel and S.A. Khan, Science 233 (1986) 1316.
4 T.T. Stenzel, P. Patel and D. Bastia, Cell 49 (1987) 709.
5 S. Deb, A.L. DeLucia, A. Koff, S. Tsui and P. Tegtmayer, Mol. Cell. Biol. 6 (1986) 4578.
6 M. Snyder, A.R. Buchman and R.W. Davis, Nature 324 (1986) 87.
7 R.L. Gourse, H.A. de Boer and M. Homura, Cell 44 (1986) 197.
8 L. Bossi and D.M. Smith, Cell 39 (1984) 643.
9 C.F. McAllister and E.C. Achberger, J. Biol. Chem. 263 (1988) 11743.
10 F. Rojo, A. Zaballos and M. Salas, J. Mol. Biol. 211 (1990) 713.
11 D. Rodhes, EMBO J. 4 (1985) 3473.
12 F. Thoma, L.W. Bergman and R.T. Simpson, J. Mol. Biol. 177 (1985) 715.
13 R.T. Simpson and D.W. Stafford, Proc. Natl. Acad. Sci. U.S.A., 80 (1983) 51.
14 C.A. Edwards and R.A. Firtel, J. Mol. Biol. 180 (1984) 73.
15 T.E. Palen and T.R. Cech, Cell 36 (1984) 933.
16 N. Ramsay, G. Felsenfeld, B. Rushton and J.D. McGhee, EMBO J. 3 (1984) 2605.
17 H.R. Drew and A.A. Travers, J. Mol. Biol. 186 (1985) 773.
18 E.N. Trifonov and J.L. Susmann, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 3816.
19 E.N. Trifonov, Nucleic Acids Res. 8 (1980) 4041.
20 P. De Santis, S. Morosetti, A. Palleschi and M. Savino, in: Structures and dynamics of nucleic acids, proteins and membranes, eds. E. Clementi and S. Chin (Plenum, New York, 1986) vol. 31.
21 P. De Santis, S. Morosetti, A. Palleschi, M. Savino and A. Scipioni, in: Biological and artificial intelligence systems, eds E. Clementi and S. Chin (ESCON, 1988) vol. 143.
22 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Biophys. Chem. 32 (1988) 305.
23 P. De Santis, G. Gallo, A. Palleschi, M. Savino and A. Scipioni, J. Mol. Liquids, 41 (1989) 291.
24 S. Cacchione, P. De Santis, D. Foti, A. Palleschi and M. Savino, Biochemistry 28 (1989) 8706.
25 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Biochemistry 29 (1990) 9269.
26 S.C. Satchwell, H.R. Drew and A.A. Travers, J. Mol. Biol. 191 (1986) 659.
27 H.S. Koo, H.M. Wu and D.M. Crothers, Nature 320 (1986) 501.
28 H.S. Koo and D.M. Crothers, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 1763.
29 T.E. Haran and D.M. Crothers, Biochemistry 28 (1989) 2763.
30 P.J. Hagerman, Nature 321 (1986) 449.
31 S. Diekmann, EMBO J. 6 (1987) 4213.
32 S. Diekmann and E. von Kitzing, in: Structure & expression, eds W.K. Olson, M.H. Sarma, R.H. Sarma and M. Sundaralingam (Adenine Press, New York, 1988) vol. 1, p. 57.
33 H.M. Wu and D.M. Crothers, Nature 308 (1984) 509.
34 D.L. Milton and R.F. Gesteland, Nucleic Acids Res. 16 (1988) 3931.
35 P. DeSantis, A. Palleschi, M. Savino and A. Scipioni, J. Mol. Biol. (1990) submitted.
36 D.L. Milton, M.L. Casper and R.F. Gesteland, J. Mol. Biol. 213 (1990) 135.
37 A.A. Travers and A. Klug, Phil. Trans. R. Soc. Lond. 317 (1987) 537.
38 L.M. deVargas, S. Kim and A. Landy, Science 244 (1989) 1457.
39 D.J. Shuey and C.S. Parker, Nature 323 (1986) 459.
40 G. Kunke, H.J. Fritz and R. Ehring, EMBO J. 6 (1987) 507.
41 G.F. Hatfull, S.Z. Noble and N.G.F. Grindley, Cell 49 (1987) 103.
42 A.E. Dieterich, R. Axel and C.R. Cantor, J. Mol. Biol. 129 (1979) 587.
43 C. Ambrose, A. Rajadhyaksha, H. Lowman and M. Bina, J. Mol. Biol. 209 (1989) 255.