Biophysical Chemistry, 46 (1993) 193–204
Elsevier Science Publishers B.V., Amsterdam

BIOCHE 01738

Relationships between intrinsic and induced curvature in DNAs: Theoretical prediction of nucleosome positioning

P. De Santis [a,*], M. Fuà [a], A. Palleschi [a] and M. Savino [b]

[a] *Dipartimento di Chimica,* [b] *Dipartimento di Genetica e Biologia Molecolare. Università di Roma "La Sapienza", P.le A.Moro, 5. 00185 Roma (Italy)*

(Received 24 June 1992; accepted in revised form 28 October 1992)

Abstract

Relationships between intrinsic and induced superstructures in DNAs are investigated using a theoretical method of sequence dependent curvature based on the integration of the slight deviations of the 16 different dinucleotide steps from the canonical B-DNA structure as evaluated by conformational energy calculations. The induced superstructures due either to circularization or to nucleosome formation are obtained from the intrinsic superstructures by localizing the minima of the distortion energy function evaluated adopting a simple harmonic model. The results in the case of circularization of a DNA fragment 169 bp are in good agreement with the experiments and offer a base of explanation to the fine features of the differential cleavage by DNAse I. The model works very satisfactorily in the prediction of the virtual positioning (rotational and translational) of nucleosomes along DNA in fairly good agreement with experimental results providing also a base for investigating their possible assembly in systems of high complexity as the minichromosomes.

Keywords: Intrinsic and induced curvatures; Circular DNA superstructures; Nucleosome theoretical prediction; Nucleosome assembly; Harmonic model of DNA deformations

1. Introduction

It is now generally accepted that DNA has an intrinsic ability to translate the deterministic fluctuations of base sequence in superstructural elements. The physical origin of this important DNA feature is still a matter of debate although some evidence has been accumulated of a major contribution of the nearest-neighbor differential interactions within the dinucleotide steps [1–9].

In fact many tracts of biosynthetic as well as of biologically relevant DNA are found to be intrinsically curved as monitored by the electrophoretic retardation and by the enhanced ability to circularization with respect to non-curved DNA tracts of equivalent length [10–21].

Beside the intrinsic curvature proteins binding to specific DNA regions often produce an induced curvature or amplify the intrinsic curvature of the examined DNA tracts. Both the actual and virtual DNA curvatures depend on the sequence and appear to be not independent functions although their relations are not yet fully clarified.

The present paper analyzes this point on the basis of our theoretical model of the sequence dependent curvature and adopting the principle that the deformation energy of the intrinsic superstructure necessary to obtain the induced su-

* To whom correspondence should be addressed.

perstructure should be minimum [22–23]. On such a basis we have explained some fine features of the differential cleavage pattern due to DNAse I on a 169 bp DNA tract after its circularization, obtained by Drew and Travers [24].

The extention of the model to the prediction of nucleosome positioning and phasing along DNAs was in very good agreement with the experimental results in different examined systems.

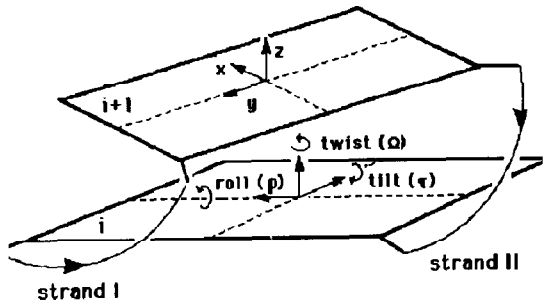

Fig. 1. Base pair orientational parameters.

2. Intrinsic and induced superstructures in DNA circularization

Circularization of a DNA tract is the result of stochastic and deterministic motions of the double helix: DNAs characterized by "in phase" curvatures have, in fact, a higher probability of circularization with respect to straight DNAs having the same length. This suggests that the final structure of the circular DNA is related to that of the corresponding linear form.

We have tried to simulate the circularization process starting from curved DNA tracts and using a conjugate gradient optimization method under the condition of minimum deformation energy, evaluated adopting a simple harmonic approximation.

Let $C(n)$ and $C'(n)$ be the intrinsic and induced curvature functions (per turn of the double helix) along the sequence, the deformation energy E was assumed to be proportional to $\langle |C'(n) - C(n)|^2 \rangle$, namely, to the average quadratic curvature deviation; where:

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

n, the sequence number; ν, the average B-DNA periodicity; g, the integration grid usually, a period of the double helix or its multiple; $d(s) = \rho - \mathrm{i}\tau$, the deviation of the dinucleotide step at position s of the sequence from the canonical B DNA in terms of the roll (ρ) and tilt (τ) angles and $w(s)$, the corresponding phase angle with respect to the first dinucleotide step evaluated by summing the relative twist angles (Ω) (see Fig. 1).

It is interesting to note that, if $d(s)$ is changed as $d(s)\exp(\mathrm{i}\omega_0) + d_0$, were d_0 and ω_0 are arbitrary values of roll, tilt and wedge direction, the modulus of curvature remains practically invariant when simultaneously changing the wedge angles and the directions of all the dinucleotide steps.

Therefore, since the gel electrophoretic anomalies due to DNA superstructure appear to be related to a quadratic form of the DNA curvature, it is hazardous to assigne physical values to the wedge angles and directions derived from the electrophoretic data. Nevertheless, it is possible to obtain from the retardation values of a representative number of multimeric sequential oligonucleotides an empirical wedge matrix as recently proposed by Bolshoy et al. [25], but without connections with the real structures of the dinucleotide steps.

In fact, a more realistic wedge matrix requires additional information about the local structure of at least two dinucleotide steps. Sources of such informations could be NMR and X-ray, as well as conformational energy calculations. It is, however, well known that X-ray structures are sensitive to packing forces and NMR data are influenced by dynamic effects which are in oligonucleotides plausibly different than in longer DNA tracts; and finally, the theoretical calculations can give only an approximate picture of the dinucleotide step structures.

Some years ago we tried to obtain a set of approximated local distortion parameters, roll and tilt angles, by localizing the minimum of the conformational energy of the different dinucleotide steps. These allowed a satisfactorily prediction of

the superstructures of synthetic and natural DNAs and their experimental manifestations; moreover the theoretical parameters showed a satisfactorily agreement with those obtained by X-ray crystal structure analysis of double helix dodecamers in spite of evident crystal packing effects [6,8,22]. On the contrary, the twist angles showed much smaller dispersion about the average values than for the X-ray data. This effect, however, does not influence sensitively the curvature which depends mainly on the average twist value. Therefore, as a first approximation, we adopted a constant periodicity around the double helical axis equal to the generally accepted average value of 10.4.

The theoretical roll and tilt angles were then assumed as the basic values for refining the local parameters of the dinucleotide steps and find a selfconsistent and realistic wedge matrix, by optimizing the agreement with a large set of electrophoretic data [4,22]. The refining procedure provided also to concerted changes of the angular parameters which implicitely take into account the other free energy contributions not considered in the conformational analysis, such as those involving solvent and counterions, as well as the effects deriving from the continuity of the polynucleotide chains including some averaging of eventual cooperative effects.

To do this, we investigated a simple physical

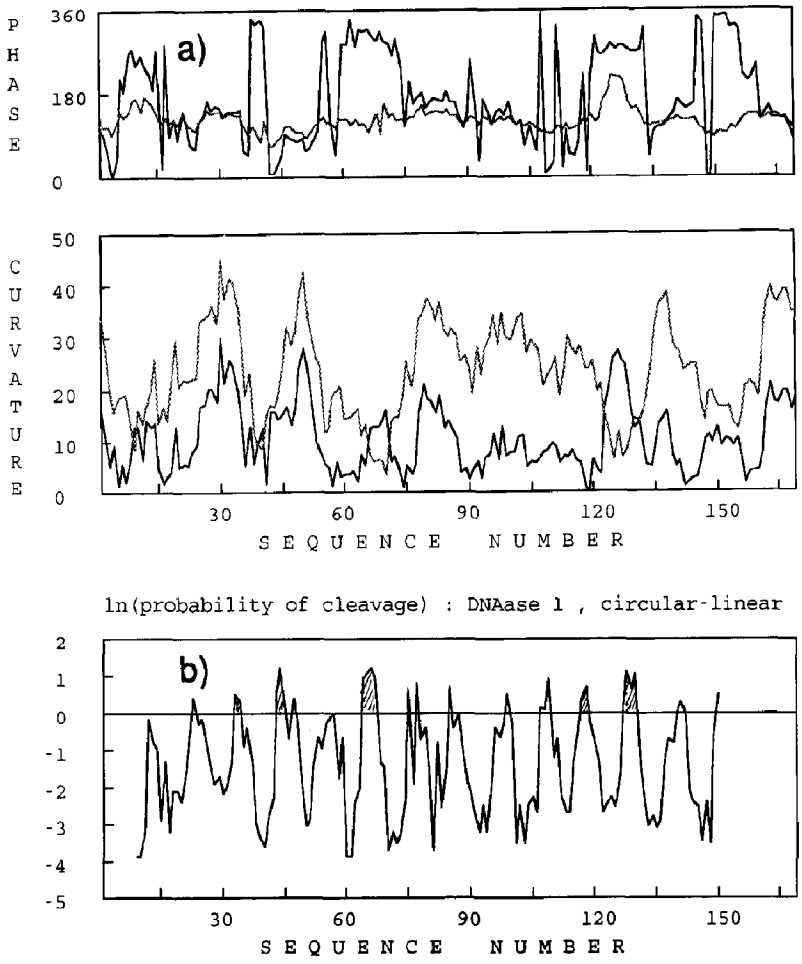

Fig. 2. (a) Differential curvature diagrams (modulus and phase) of the linear (dark trace) and circular (grey trace) DNA 169 bp investigated by Drew and Travers [24]. (b) The log of the differential (circular–linear) cleavage probability (averaged on both the strands) by DNAase I as obtained by rearranging the experimental data of Drew and Travers [24].

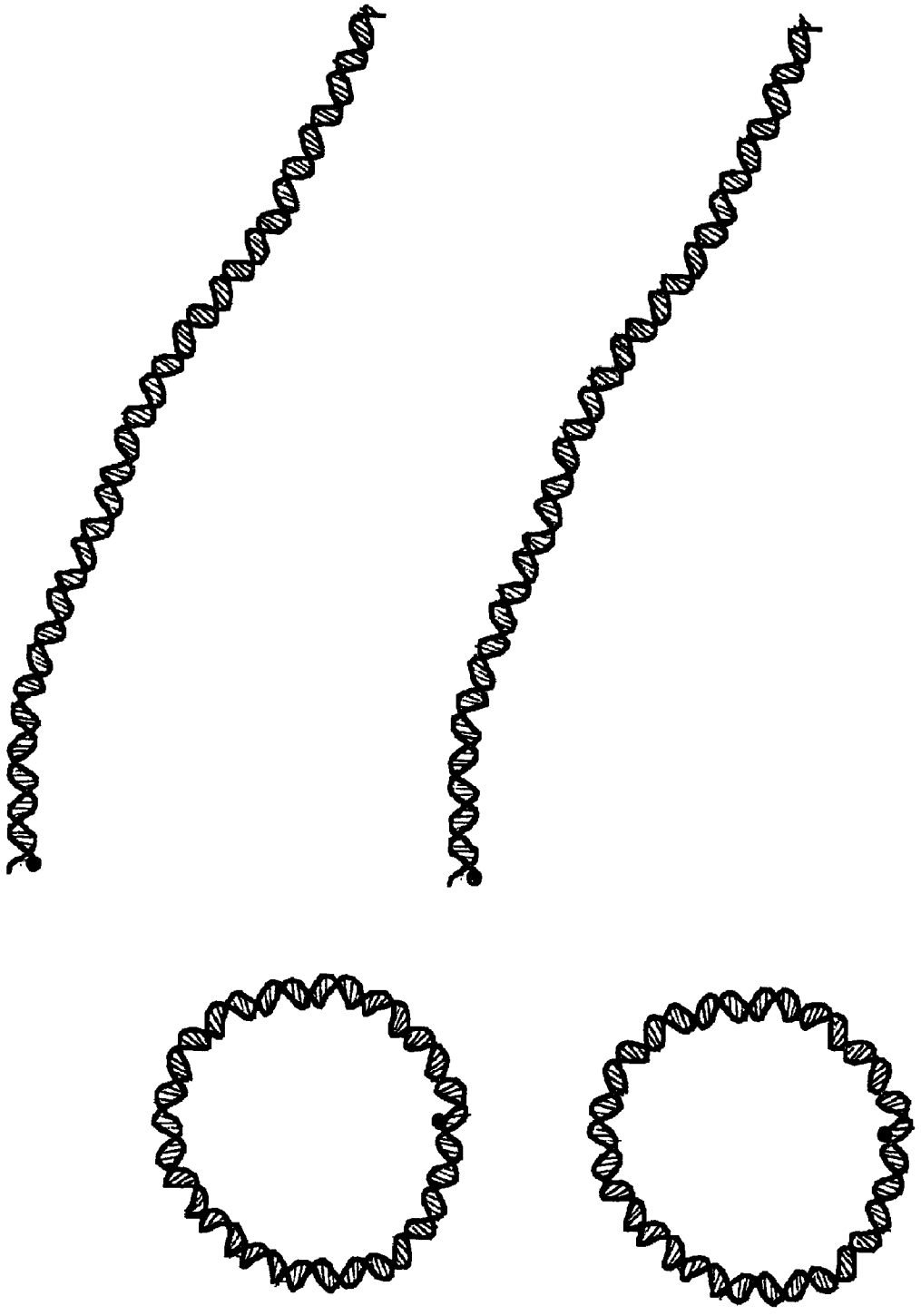

Fig. 3. Stereoprojections of the predicted linear and circular DNA 169 bp of Fig. 2.

model to explain the electrophoretic anomalies due to DNA curvature and obtained a quantitative correlation between retardation and curvature [4,22]. The model is based on the calculation of the central dispersion of curvature as evaluated by integrating the slight angular deviations from the canonical B-DNA structure of the dinucleotide steps in double stranded DNAs.

Thus, the roll, tilt and twist angles theoretically evaluated for the 16 different dinucleotide steps were refined and the results are reported below (the roll and tilt angles are conveniently represented in the complex plane as deviation vector $\boldsymbol{d} = (\rho, -\tau)$ in the following Hermitian matrix form), d and Ω are given 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	(2.0, 1.7)	(−2.5, −2.7)	(1.3, 0.6)	(−3.7, 0.0)

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

Thus, the intrinsic curvature $C(n)$ of a DNA tract can be obtained adopting these wedge matrices, whereas the induced curvature of the related circular form is the result of the variational method described before. Figure 2a illustrates the results obtained in the case of the DNA tract 169 bp investigated by Drew and Travers [24]: in particular the dark lines show the curvature diagrams (modulus and phase calculated with a grid of $g = 10$ bp) of the linear DNA and the grey lines the related diagrams of the circular form. It is interesting that in spite of the low curvature which characterizes such a DNA tract, as shown in the stereoprojection of Fig. 3, the same result was obtained for different cyclically permuted sequences, which strongly indicates that the curvature induced by circularization is determined by the intrinsic curvature of the linear DNA.

The curvature phase and the linking number of the circular DNA account very satisfactorily for the differential cleavage by DNAse I as found by Drew and Travers [24]. In fact Fig. 2b shows the cleavage probability averaged on both the strands (after a relative shift of 2 bp to take into account the topography of the minor groove where the DNAse I works) of the circular with respect the linear form: the typical trend mirrors the periodical deformation of the small groove which is generally contracted where it faces into the cycle.

It is very interesting, however, that in some sites of the sequence, the small groove results wider in the circular than in the linear form,as indicated by the crossing of the corresponding curvature functions in Fig. 2a. These features account fairly well for the strange increase of the cleavage probability in the circular form for those tracts (see Fig. 2b).

In order to reduce the calculation times, an alternative and new mathematical appraoch was attempted: it consists in finding the deformations of uniform cyclic DNA under the force field of the differential interactions along the sequence. Such deformations are given in the curvature complex representation by a Fourier series with variational amplitudes and phases. These are easily obtained by minimizing the deformation energy E with respect to the Fourier coefficients by using a steepest descent method. The amplitude of the first harmonic was set equal to zero in order to maintain fixed the cyclic form. Such a condition, in fact, ensures the persistence of the cyclization for small deformations of the structure: actually, only the first harmonic provides the concerted local curvature changes of the circular form which may affect its integrity; it corresponds in the linear approximation of the response theory to the integration of the effects of a periodical perturbation on a periodical superstructure which is non-zero only when the periods coincide. Furthermore, the amplitude of the zero order was fixed in order to ascertain the topological invariance of the circular DNA with its modulus equal to 360 divided the nearest integral number of DNA turns and its phase equal to the average value of the linear form.

Practically, the method requires iterative calculations of the harmonic components within a prefixed resolution power. However, by taking advantage of the theorem of the minimum

quadratic average deviation of two functions given in terms of Fourier series, and applying Parseval's formula, it is possible to solve analytically the problem of finding the most stable superstructure of a circular DNA in terms of that of the corresponding linear DNA.

Thus, adopting the harmonic approximation of the deformation energy, the curvature function of the cycle, $C'(n)$, results:

$$C'(n) = C(n) - A_0 - A_1 + A'_0 + A'_1$$

where A_0 and A_1 are the zero and the first harmonics, respectively, of the Fourier series of the intrinsic curvature, whereas A'_0 and A'_1 are the corresponding harmonics of the circular DNA; as defined before A'_1 is set equal to zero. We have tried such a method of minimum harmonic deformation energy for the case investigated by Drew and Travers and found exactly the same result as previously obtained just in a few seconds on a personal computer. The stereoprojections of the cyclic and the linear DNA are illustrated in Fig. 3.

The success of such a method prompted us to its extension to problems where the writhing of a chain is changed under the action of a force field, such as in the protein folding and for what this paper is concerned, in the nucleosome formation along DNA sequences.

3. Nucleosome positioning

The nucleosome positioning along a DNA sequence is defined by two parameters: the translation marking where the histone octamer dyad axis is placed and the orientation of DNA relative to the direction of curvature. Many authors agree that little is known about the translation, whereas the rotational parameter appears to involve structures where the A/T rich minor grooves face in towards the protein core.

Adopting the model of the minimum harmonic energy deformation, calculations were tried to investigate the nucleosome formation along DNAs. Recurrent tracts 145 bp of the DNA investigated by Drew and Travers [24] were in fact constrained to deform until a nucleosomal superstructure characterized by a regular superhelix of 1 and 3/4 turns with 43 Å of radius and 28 Å of pitch was obtained.

The most favourable nucleosomal structure, corresponding to the minimum deformation energy, accounts satisfactorily for the observed periodical differential cleavage of the reconstituted

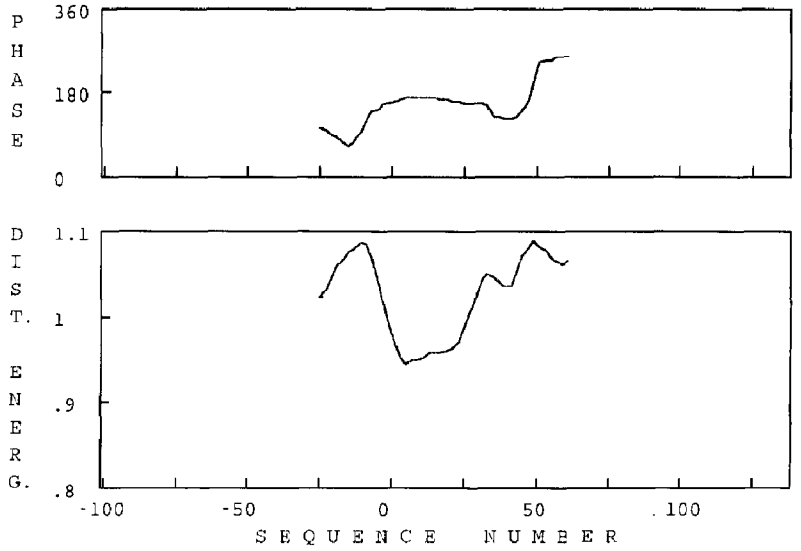

Fig. 4. Deformation energy profile of a DNA tract 240 bp of the *Xenopus* 5S ribosomal DNA gene and the corresponding phase of the nucleosome dyad axis.

nucleosome which appears to be coherent with that of the cycle as shown before.

Thus, the relatively small intrinsic curvature of this sequence seems to determine phasing and positioning of the nucleosomal superstructure. On the basis of this result, we have designed a computer program to predict the nucleosome positioning along different DNA sequences, by localizing the minima of the deformation energy function required to distort recurrent tracts 145 bp of the sequence in a nucleosomal superstructure. To do this the curvature of the nucleosomal tract was fixed with a modulus equal to 45 degrees and a regularly decreasing phase with a rate of -0.5 degrees per bp.

Figure 4 illustrates the minimum deformation energy profile obtained in the case of the DNA tract of the *Xenopus* 5S ribosomal RNA gene, 240 bp and the relative phase of the nucleosome dyad axis with respect the pseudodyad axis of the

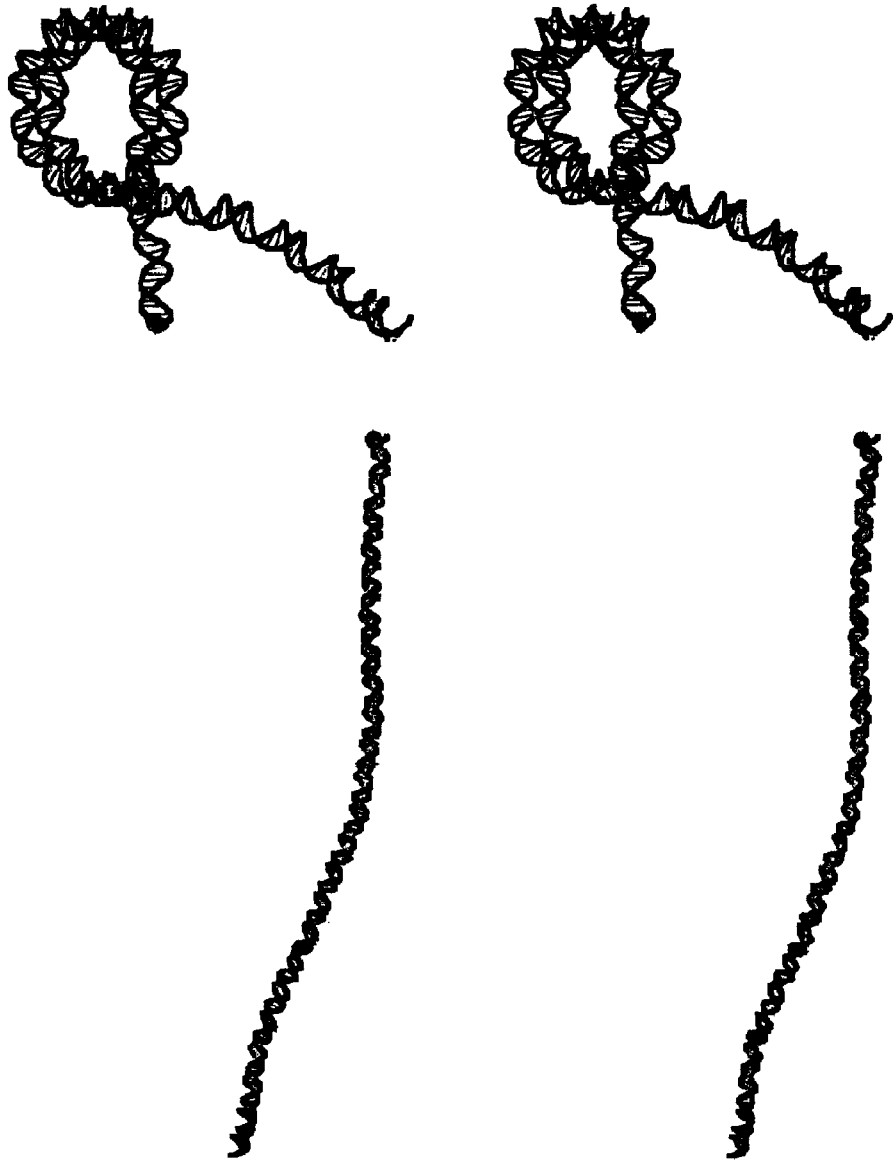

Fig. 5. Stereoprojection of the predicted intrinsic and induced nucleosomal superstructures of the DNA tract of Fig. 4.

first dinucleotide step. This DNA is of particular interest because the nucleosome has a precise localization as shown by Rhodes on the basis of DNAase I differential cleavage analysis [26]. The values of the distortion energy are given in terms of that of an equivalent tract of poly dA–poly dT which is well known to be unable to form nucleosome [27]. The minimum corresponds to the predicted virtual translation positioning of the nucleosome dyad axis in good agreement with the experimental localization [26].

Figure 5 shows the stereoprojections of the predicted superstructures of the linear and the corresponding nucleosomal form. Single nucleosomes were also succesfully localized on some other DNA tracts investigated in different laboratories.

Particularly interesting is the prediction of the nucleosome positioning in the case of the Alu-repeat of ribosomal DNA in fruitflies which was experimentally confirmed by Mirzabekov [28]. The deformation energy diagrams and the stereoprojections of the intrinsic and nucleosomal DNA superstructures are shown in Figs. 6 and 7.

Finally the preferential position of a nucleosome was successfully predicted in a tract of *Crithidia* mitochondrial DNA and localized at the center of the DNA loop with the same phase in agreement with the experimental data [29].

In order to analyse the sensitivity of the results to changes of the model of the nucleosome superstructure we investigated the distortion energy function when a DNA tract 145 bp was constrained to follow two spires 3/4 of turn of a flat superhelix interconnected by a straight DNA tract corresponding to an open key model of nucleosome [30,31]. This model is a reduced representation of the average intrinsic curvature distribution along the sequence as found for the set of 177 nucleosomal DNAs investigated by Satchwell et al. [31] and Boffelli et al. [32]. The results obtained adopting this model were very similar.

The possibility of finding the virtual positioning of nucleosomes along DNA sequences by localizing the minima of the distortion energy function, evaluated at a rate of a nucleosome per second, provides a suitable theoretical method to predict the translation and the rotational parameters of a set of nucleosomes along DNAs as well as their phase relations.

Thus, we extended the model to the more complex case of the formation of many nucleo-

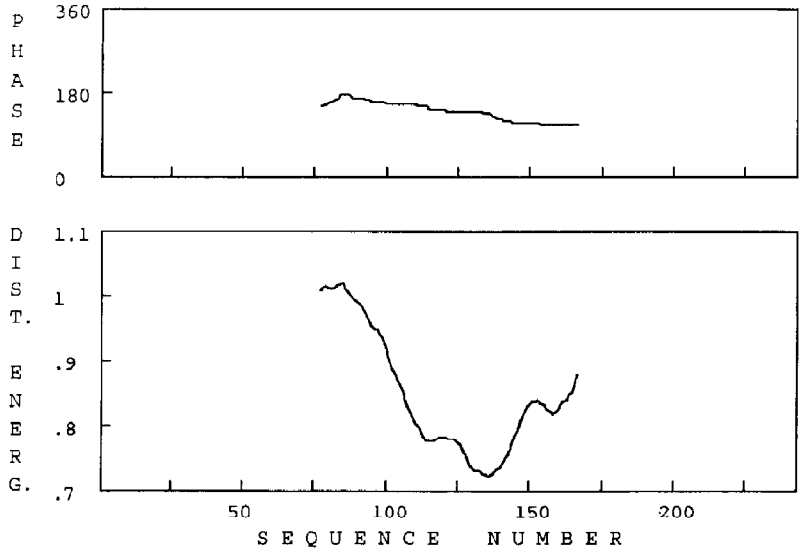

Fig. 6. Deformation energy profile of the Alu repeat of ribosomal DNA in fruitflies and the corresponding phase of the nucleosome dyad axis.

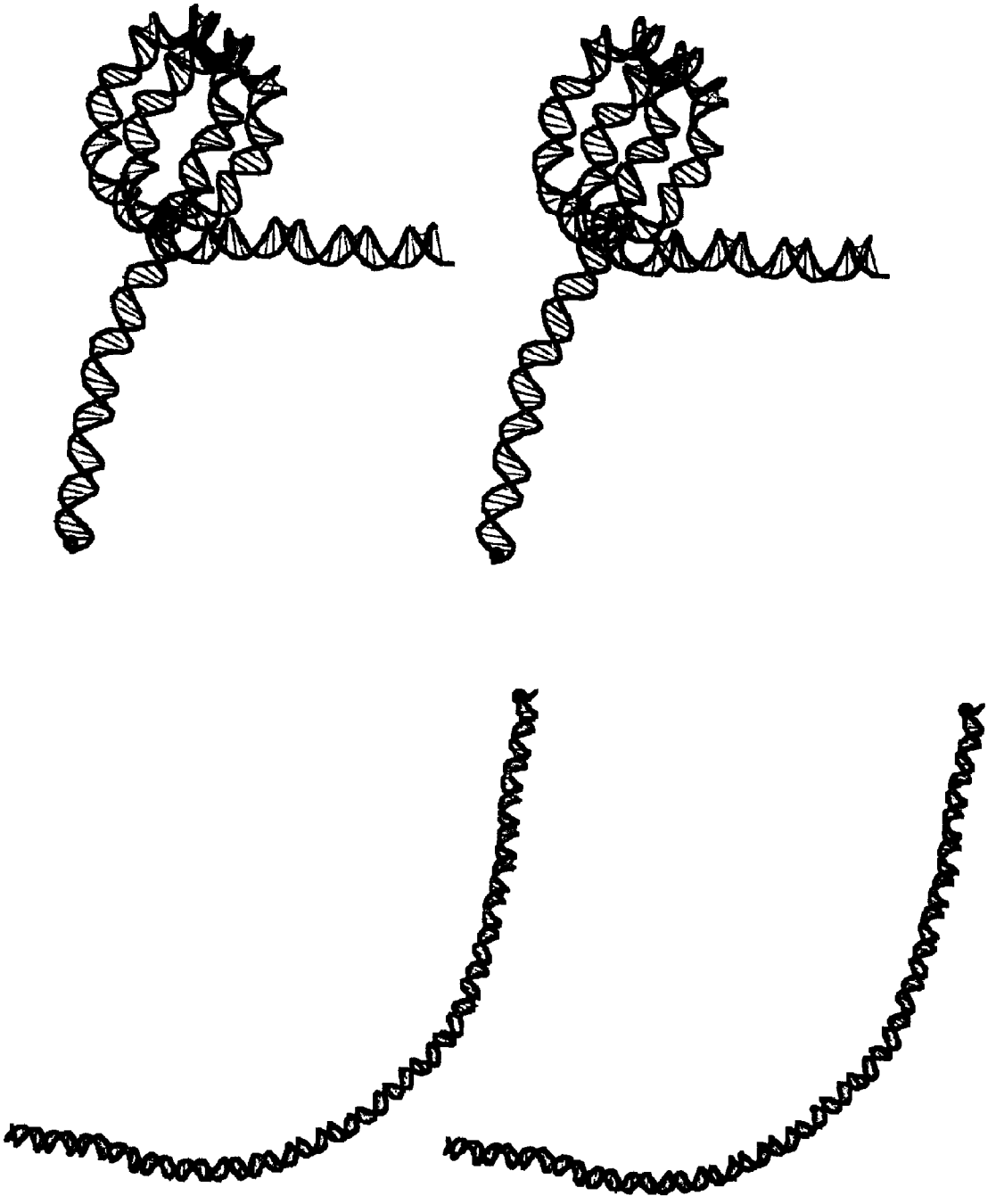

Fig. 7. Stereoprojection of the predicted intrinsic and induced nucleosomal superstructures of the DNA tract of Fig. 6.

somes on long tracts of DNA. Figure 8 illustrates the case of a tract of 800 bp about the replication origin of SV40 recently investigated by Ambrose et al. [33]. The minima of the deformation energy obtained, using the open key model, represent the virtual positions of the nucleosome dyad axis along the DNA sequence.

The obvious exclusion condition of nucleosomes at distances less than 160 bp, allows two optimal assemblies of four nucleosomes. These

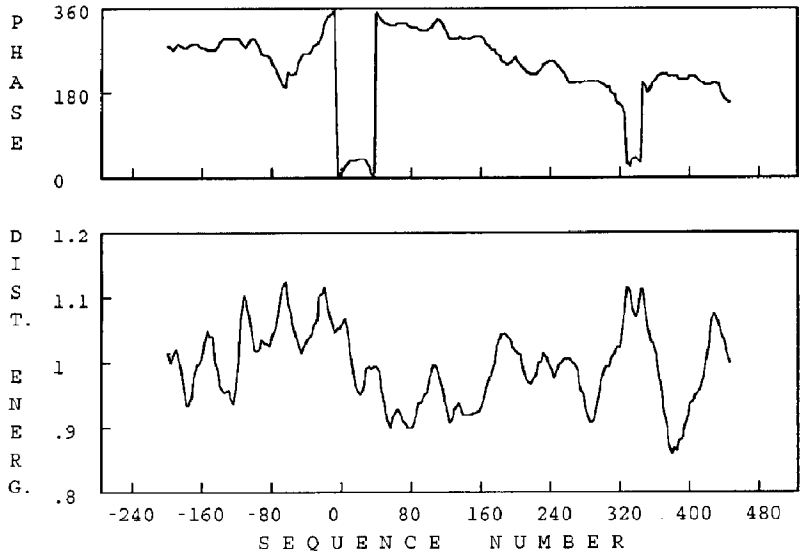

Fig. 8. Deformation energy profile of a DNA tract (−275, 550) about the replication region of SV40 and the corresponding phase of the nucleosome dyad axis. The minima represent the virtual positions of nucleosomes along the sequence.

appear closely related with the experimental clone frequency obtained by Ambrose et al. [33] which represents the differential occurrence of nucleosomes along that DNA and shown in Fig. 9 as a distribution function of the nucleosome dyad axis positions obtained by smoothing the experimental histogram. In fact, the deepest minimum at $n = 385$ corresponds very closely to the maximum clone frequency and this favours the occurrence of a second nucleosome at $n = 190$. Further, the precise position of the nucleosome dyad axis at the sequence number 384–387 was very recently found for the *in vitro* reconstituted nucleosome in a DNA tract around that position by Powers and Bina [34].

The most favorable assembly of four nucleosomes is shown in Fig. 9 as a stereoprojection. Here, it is easy to observe some regularity in the resulting superstructure which can be interesting for the further nucleosome association in minichromosomes as well as for the changes between the possible nucleosome assemblies promoted by some regulative proteins during replication and transcription in the active chromatin. It is interesting to note that the results obtained by adopting the theoretical model illustrated in the

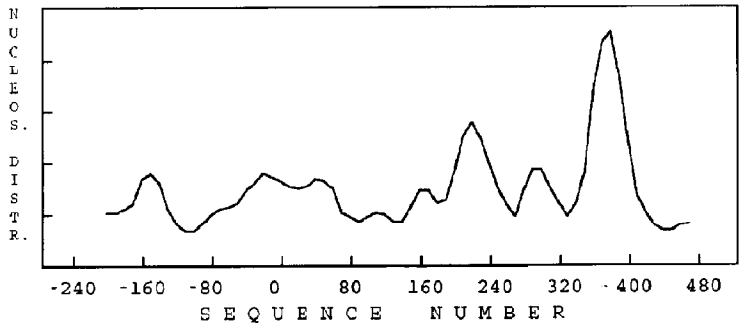

Fig. 9. Distribution of nucleosomes along the DNA tract (−275, 550) of SV40 as obtained by smoothing the histogram of the experimental data of Ambrose et al. [33].

present paper are practically the same we previously obtained [32] by localizing the maxima of correlation function between the curvature of recurrent tracts 145 bp of DNA and the average curvature function of the set of 177 nucleosomal DNAs investigated by Satchwell et al. [31].

Contrary to the cases of localization of a single nucleosome in relatively short tracts of DNA, where the distortion energy function contained only a minimum whose position is relatively insensitive to the change of the structural model of nucleosome as discussed before, the distorsion energy profile of a long tract of DNA appears to be sensitively influenced by the model adopted. The open key model seems to be in better agreement with the experimental data. This point requires further analysis: it probably is in relation with the mechanisms of nucleosome formation and its dynamics.

4. Concluding remarks

The implication of DNA curvature in mechanisms which govern biological processes such as replication, transcription and chromatin organization, appears to be a complex problem involving static and large scale dynamic effects and certainly requires further investigations.

However, the theoretical model we adopted, which is based on the integration of the local deviations from the canonical B DNA structure due to differential interactions along the sequence as evaluated by conformational energy calculations, appears to be suitable for translating the sequence fluctuations in superstructural elements of DNA. It allowed the prediction of superstructures of a large set of biosynthetic periodical DNAs as well as of biologically relevant tracts of natural DNAs and their electrophoretic manifestations in excellent agreement with the experiments [4,8,22,35]. Beside this reliability it shows a surprisingly high sensitivity as demonstrated by the good results obtained in the prediction of the electrophoretic changes observed in a tract of SV40 DNA after extensive point mutations [36].

This paper illustrates the application of the model to investigate the relations between the intrinsic DNA superstructures and those induced by cyclization as well as by nucleosome formation. The results obtained adopting a simple harmonic model to evaluate the distortion energy are very promising; they allow the theoretical prediction of fine effects of nuclease activity along DNA sequences as well as the *a priori* positioning and phasing of nucleosomes. This indicates that the intrinsic superstructures of DNA are the main determinant of the induced superstructures so providing the possibility to predict the virtual recognition sites of regulative proteins.

The important ability of DNA to filter and productively integrate the slightly different stereochemistry of the nucleotide steps giving rise to sequence dependent superstructures, is not yet quite evaluated but it is easy to predict that it

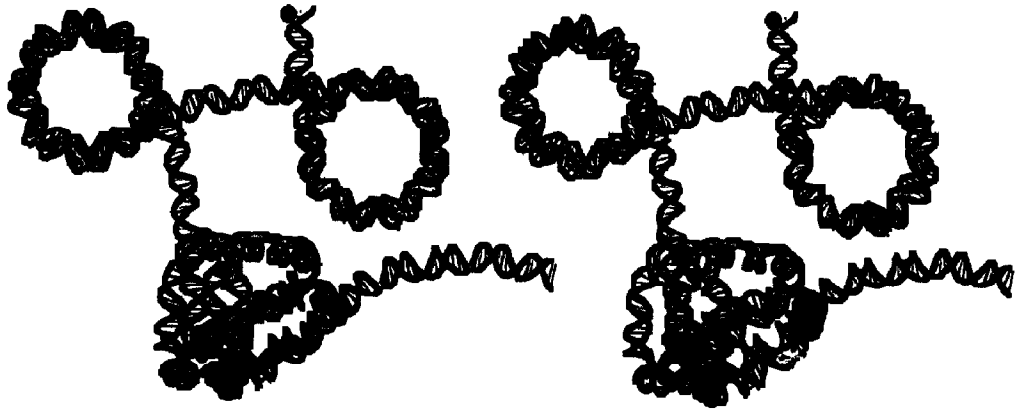

Fig. 10. Stereoprojection of the predicted most stable assembly of 4 nucleosomes along the −275, +550 tract of SV40.

could get an insight into the molecular mechanisms of management of the still latent genetic information encoded in DNA.

Acknowledgements

The financial support of CNR Progetto Finalizzato Chimica Fine II and of Istituto Pasteur-Fondazione Cenci Bolognetti is gratefully acknowledged.

References

1 E.N. Trifonov and J.L. Sussman, Proc. Natl. Acad. Sci. (USA), 77 (1980) 3816.
2 E.N. Trifonov, Nucleic Acids Res. 8 (1980) 4041.
3 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 (1986) p. 31.
4 P. De Santis, S. Morosetti, A. Palleschi, M. Savino and Scipioni A., in: Biological and artificial intelligence system, eds. E. Clementi and S. Chin (Plenum Press, New York) (1988) p. 143.
5 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Biophys. Chem. 32 (1988) 305.
6 P. De Santis, G. Gallo, A. Palleschi, M. Savino and A. Scipioni, J. Mol. Liquids, 41 (1989) 291.
7 S. Cacchione, P. De Santis D. Foti, A. Palleschi and M. Savino, Biochemistry, 28 (1989) 8706.
8 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Biochemistry 29 (1990) 9269.
9 L.E. Ulanovsky and E.N. Trifonoy, Nature 326 (1987) 720.
10 H.S. Koo, H.M. Wu and D.M. Crothers, Nature 320 (1986) 501.
11 P.J. Hagerman, Nature 321 (1986) 449.
12 W. Ross and A. Landy, J. Mol. Biol. 1576 (1982) 523.
13 K. Zahn and F.R. Blattner, Nature 317 (1985) 451.
14 R.R. Koepsel and S.A. Khan, Science 233 (1986) 1316.
15 T.T. Stenzel, P. Patel and D. Bastia, Cell 49 (1987) 709.
16 S. Deb, A.L. DeLucia, A. Koff, S. Toni and P. Tegtmayer, Mol. Cell. Biol. 6 (1986) 4578.
17 M. Snyder, A.R. Buchman and R.W. Davis, Nature 324 (1986) 87.
18 R.L. Gourse, H.A. de Boer and M. Nomura, Cell 44 (1986) 197.
19 L. Bracco, D. Kotlarz, A. Kolb, S. Diekmann and H. Buc, EMBO J. 8 (1989) 4289.
20 C.F. Mc Allister and E.C. Achberger, J. Biol. Chem. 263 (1988) 11743.
21 F. Rojo, A. Zaballos and M. Salas, J. Mol. Biol. 211 (1990) 713.
22 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Int. J. Quantum Chem. 42 (1992) 1409.
23 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Nucleic Acids Symp. Ser. 25 (1991) 83.
24 H.R. Drew and A.A. Travers, J. Mol. Biol. 186 (1985) 773.
25 A. Bolshoy, P. McNamara, R.E. Harrington and E.N. Trifonov, Proc. Natl. Acad. Sci. USA 88 (1991) 2312.
26 D. Rhodes, EMBO J. 4 (1985) 3473.
27 D. Rhodes, Nucleic Acids Res. 6 (1979) 1805.
28 A. Mirzabekov, Personal communication, 1992.
29 G. Costanzo, E. Di Mauro, G. Salina and R. Negri, J. Mol. Biol. 216 (1990) 363.
30 A.A. Travers and A. Klug, Phil. Trans. R. Soc. (London) 317 (1987) 537.
31 S.C. Satchwell, H.R. Drew and A.A. Travers, J. Mol. Biol. 191 (1986) 659.
32 D. Boffelli, P. De Santis, A. Palleschi and M. Savino, Biophys. Chem. 39 (1991) 127.
33 C. Ambrose, A. Rajadhyaksha, H. Cowman and M. Bina, J. Mol. Biol. 209 (1989) 255.
34 J.H. Powers and M. Bina, J. Mol. Biol 221 (1991) 795.
35 D. Boffelli, P. De Santis, A. Palleschi, G. Risuleo and M. Savino, FEBS Lett. 300 (1992) 175.
36 P. De Santis, A. Palleschi and M. Savino, Biophys. Chem. 42 (1992) 147.