ELSEVIER

Biophysical Chemistry 95 (2002) 23–47

Biophysical Chemistry

www.elsevier.com/locate/bpc

From the sequence to the superstructural properties of DNAs

C. Anselmi, P. De Santis*, R. Paparcone, M. Savino, A. Scipioni

Dipartimento di Chimica, Università 'La Sapienza', P. le A. Moro 5, I-00185 Rome, Italy

Received 26 July 2001; received in revised form 31 October 2001; accepted 31 October 2001

Abstract

A theoretical model for predicting intrinsic and induced DNA superstructures as well as their thermodynamic properties is presented. Intrinsic sequence-dependent superstructures are evaluated by integrating local deviations from the canonical B-DNA of the different dinucleotide steps. Induced superstructures are obtained by adopting the principle of minimum deformation free energy, evaluated in the Fourier space, in the framework of first-order elasticity. Finally dinucleotide stacking energies and melting temperatures are considered to account for local flexibility. In fact the two scales are strongly correlated. The model works very satisfactorily in predicting the sequence-dependent effects on the DNA experimental behavior, such as the gel electrophoresis retardation, the writhe transitions in topologically constrained domains, the thermodynamic constants of circularization reactions as well as the nucleosome thermodynamic stability constants.

Keywords: DNA curvature; DNA flexibility; Electrophoretic retardation; Writhe transitions; DNA circularization; Nucleosome stability

1. Introduction

The reductionistic program of Molecular Biology represents the attempt to explain the immense variety of phenomena in living matter in terms of the physical chemistry laws. Proteins and nucleic acids are the most important basic elements of biological systems. Proteins are involved in a wide spectrum of biological functions as a result of their structural polymorphism and conformational versatility. This is due to the large chemical dispersion of the 20 different amino-acid side groups, which strongly modulate the conformational features of the polypeptide chain.

Many attempts have been made to predict the tertiary structures of proteins from their amino-acid sequence. In spite of the large progress in the study on the nature of intermolecular forces and the present power of computational systems and techniques, the full solution to this problem is still far from being found. On the contrary, the substantial homogeneity and the consequent conformational degeneracy of the nucleotide residues along DNA double helix, have recently allowed good progress in the knowledge of the molecular mechanisms that control the functional organization of the genome as well as the prediction of some

* Corresponding author. Tel.: +39-06-4991-3228; fax: +39-06-445-3827.

E-mail address: pasquale.desantis@uniroma1.it (P. De Santis).

0301-4622/02/$ - see front matter
PII: S0301-4622(01)00246-0

biologically relevant structural properties in terms of the sequence.

In fact, the perfect base pairing of the DNA double helix restricts the structural variance in the biological conditions to relatively slow modulations of the B-DNA structure. These modulations can produce superstructures when phased with the DNA helical periodicity. They are relevant for their physical–chemical properties and biological functions. In fact, many DNAs were found to be curved in biologically important tracts, such as protein-binding regions, transcription loops or nucleosomes. In addition, it is reasonable that the differential flexibility of the DNA chain could favor the structural deformations involved in these processes. This suggests that sequence-dependent curvature and flexibility play a relevant role in the management of the complex physical–chemical transformations in the living matter.

In the present paper some theoretical and experimental aspects of a statistical–thermodynamic model, advanced some years ago, are reported. It allows the prediction of the sequence-dependent DNA superstructures by integrating the local deviations from the canonical B-DNA structure, produced by the differential interactions between the 16 different dinucleotide steps. The framework of first-order elasticity theory is adopted to represent the DNA response to perturbing forces arising, for example, from protein binding, circularization, and, generally, from the environment.

2. DNA intrinsic curvature

The curvature of a space line is defined as the derivative, $\mathrm{d}\mathbf{t}/\mathrm{d}l$, of the tangent versor, **t**, along the line, l [1]. Its modulus is the inverse of the curvature radius and its direction is that of the main normal to the curve. Namely, curvature represents the bend of the local helical axes while the phase is its direction. Sequence-dependent DNA curvature can be conveniently represented by a complex function of the sequence number, n

$$\mathbf{c}(n,t)=\frac{1}{\nu}\sum_{n\text{th turn}}\left[\mathbf{d}(s)+\delta(s,t)\right]\exp\left(2\pi\mathrm{i}\frac{s}{\nu}\right)=\mathbf{c}(n)+\chi(n,t) \qquad (1)$$

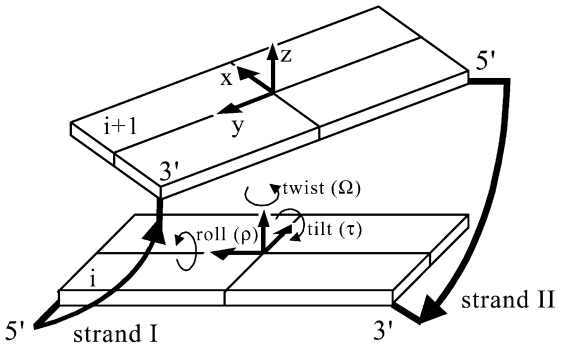

Fig. 1. Orientational parameters of the base pair average plane in a dinucleotide step.

$\mathbf{c}(n,t)$ is the curvature of the double helix per bp, calculated over recurrent turns and assigned to the central base pair of the nth turn, at the time t; $\mathbf{d}(s)=(\rho-i\tau)$ is the intrinsic deviation from the canonical B-DNA structure of the sth dinucleotide step in terms of its roll, ρ and tilt, τ, angles (Fig. 1); $\mathrm{d}(s,t)$ represents the local dynamic fluctuations of sth base pair curvature at the time t; ν is the helical average periodicity at the nth turn, evaluated from the local twist angles, Ω.

Summing up the local roll and tilt deviations over a turn defines the curvature of the helical axis that occurs only when a harmonic component of the distribution of the local deviations exists coherent with the helix periodicity. All the other contributions affect locally the DNA structure, but their effects cancel and do not produce superstructural modifications of the chain.

$\mathbf{c}(n,t)$ results equal to the sum of a static intrinsic contribution, $\mathbf{c}(n)$, and a dynamic contribution, $\chi(n,t)$, whose time average is obviously zero. Therefore the time average or the ensemble average curvature simply corresponds to the intrinsic curvature of the DNA sequence.

Roll, tilt and twist angles were first obtained by minimizing the conformational energy of the different dinucleotide steps and later refined to optimize the agreement with experimental data [2,3]. They are reported in degrees in Table 1.

As curvature is a complex function, it is represented in modulus, which quantifies the local angle of bending, and phase, namely the direction where

the bending points. As an example, Fig. 2 illustrates the intrinsic curvature diagram in modulus and phase of 423 bp tract of Kinetoplast DNA from *L. tarentolae*. It represents the first DNA where intrinsic curvature was experimentally detected by polyacrylamide gel electrophoresis retardation [4]. In agreement with the experimental evidence the diagram shows a region of marked curvature centered at $n=145$, which results in a rather sharp bend of the helical axis of approximately 90°.

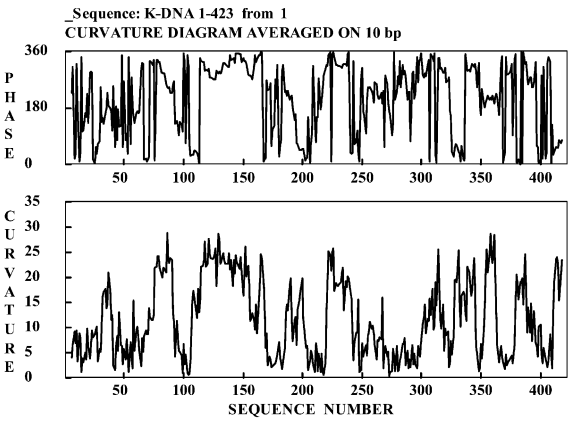

Fig. 2. Diagrams representing the curvature modulus and phase of the 217 bp DNA sequence containing the loop region of *C. fasciculata*. The modulus is reported in degrees per turn of double helix. The phase is reported in degrees.

3. The dispersion of curvature and the prediction of the gel electrophoresis retardation

At present, the most sensitive tool for detecting curvature in DNA tracts or localizing the main curvature sites, is represented by polyacrylamide gel electrophoresis. Curved DNAs are characterized by a lower electrophoretic mobility than straight DNAs with the same molecular mass. The retardation factor quantifies this property as the ratio between the apparent and the actual molecular sizes.

Wu and Crothers [5] measured electrophoretic mobility of a restriction fragment isolated from *L. tarentolae*. They localized bent regions along the sequence corresponding to regions of repeated A·T in phase with the B-DNA helix screw.

Later Koo et al. [6] studied the influence of phasing AA·TT stretches in homopolymeric DNA oligomers. They found that the phasing of the sequences coherent with the helical repeat is a sufficient condition to induce electrophoretic anomalies, whereas DNA with AA·TT stretches with different periodicities were not affected. They concluded that AA·TT stretches with the proper phasing are able to induce an intrinsic curvature in DNA chains.

Hagerman studied the problem of phasing by analyzing the gel retardation of several DNA containing AA·TT stretches with different polarity. He found that DNAs with sequences $d(CA_4T_4G)_n$ or $d(GA_4T_4C)_n$ show high retardation, whereas DNAs with sequences $d(CT_4A_4G)_n$ or $d(GT_4A_4C)_n$ have normal mobility. He concluded that also the dinucleotides different from AA·TT, play a relevant role in determining intrinsic bending [7,8].

Two models were advanced to explain the electrophoretic anomalies in physical–chemical terms. The first suggested the conformational transition at the boundaries between the canonical B-DNA and a different structure, characterizing AA·TT repeated stretches, as the origin of the retardation

Table 1
Roll, ρ, tilt, τ, and twist, Ω, angles in degrees for each dinucleotide step

	ρ, τ, Ω			
	A	T	G	C
T	8.0, 0.0, 34.5	−5.4, 0.5, 36.0	6.8, −0.4, 34.1	2.0, 1.7, 34.6
A	−5.4, −0.5, 36.0	−7.3, 0.0, 35.3	1.0, −1.6, 34.4	−2.5, −2.7, 33.7
C	6.8, 0.4, 34.1	1.0, 1.6, 34.4	4.6, 0.0, 33.5	1.3, 0.6, 33.1
G	2.0, −1.7, 34.6	−2.5, 2.7, 33.7	1.3, −0.6, 33.1	−3.7, 0.0, 33.3

[6,9]. This gives rise to bends, which integrate effectively when in phase with DNA periodicity.

Following the wedge hypothesis suggested by Trifonov [10], an alternative model based on the differential interactions between nearest-neighbor base pairs was advanced, which introduces structural perturbations on the double helix [2,3]. When coherent with the B-DNA helical symmetry, they are amplified on a superstructural level. As a matter of fact, the resulting curvature satisfactorily correlated with the retardation factors of a number of multimeric oligonucleotides.

Later, Bolshoy et al. [11] analyzed the gel retardation data relative to 54 synthetic DNA fragments not having AA·TT in their sequences, and found that also DNA containing other types of dinucleotides can show large electrophoretic anomalies, in agreement with the previous proposal [2,3].

Several authors have tried to rationalize and justify in terms of physical-chemical properties the relation between curvature and electrophoretic mobility. Most of the model can be considered extensions of the reptation theory [12].

The expression for electrophoresis mobility, μ, was derived by Lumpkin et al. [13]. The most interesting result was that μ appears to be proportional to the component of the end-to-end vector in the direction of the electric field and inversely proportional to the squared DNA contour length. The first term can be related to the persistence length of the DNA chain. However, this model does not account for electrophoretic anomalies.

The experimental evidence of curved DNAs [4] has lead to reconsider the concept of persistence length, trying to include not only the dynamical effects due to the chain motion, but also the static effect of intrinsic bents.

Actually, Trifonov et al. [14] introduced the concept of apparent persistence length. They stated that the presence of static wedges in the DNA chain, due to the occurrence of the different dinucleotides, reduces the measured value of the DNA persistence length, with respect to the value relative to the dynamic flexibility of the chain. They proposed that the apparent persistence length, P_{app}, is related to the static and dynamic contributions as $1/P_{app}=1/P_s+1/P_d$.

Later, Schellman and Harvey [15] proposed a mathematical method to analyze the static contributions to the persistence length and the dynamic contributions to curvature due to the DNA sequence.

On the contrary, Levene and Zimm [16] tried to explain DNA retardation including DNA elastic free energy. They simulated the behavior of DNA chain as it moves along the gel network and studied the DNA conformations in term of the first-order elastic free energy to distort the chain from its intrinsic structure. The results they obtained reproduced the dependence of the anomalous mobility of kinetoplast DNA fragments.

At present, it is clear that DNA gel retardation is affected by sequence-dependent DNA curvature and flexibility. Therefore a general model should account for both effects.

The electrophoretic retardation is defined as the ratio of the apparent and real molecular weights of a DNA chain. In a first approximation, it corresponds to the mobility ratio of a curved DNA and a straight DNA with the same molecular weight, indicated by μ_{str} and μ_{cur} respectively. If we adopt the simple thermodynamic hypothesis that DNA mobility depends on the activation free energy required to travel among the gel network, the electrophoresis retardation should be exponentially related to the activation free-energy difference necessary to distort the straight and curved DNA, respectively

$$R=\frac{\mu_{str}}{\mu_{cur}}=\exp\left(-\frac{\Delta G^{\ddagger}_{str}-\Delta G^{\ddagger}_{cur}}{RT}\right) \quad (2)$$

Practically it depends on the average energy necessary to straighten the curved DNA.

Consequently, the electrophoretic retardation should be a manifestation of the longer relaxation time required by the curved DNA to reach a plausibly straight structure, more suitable to travel among the gel network in the electric field direction.

Following the Landau–Lifshitz model of an elastic cylindrical bar deformation [1] and assuming a simple first-order elasticity model of DNA deformation, the straightening energy is proportional to σ^2, the second moment of the distribution

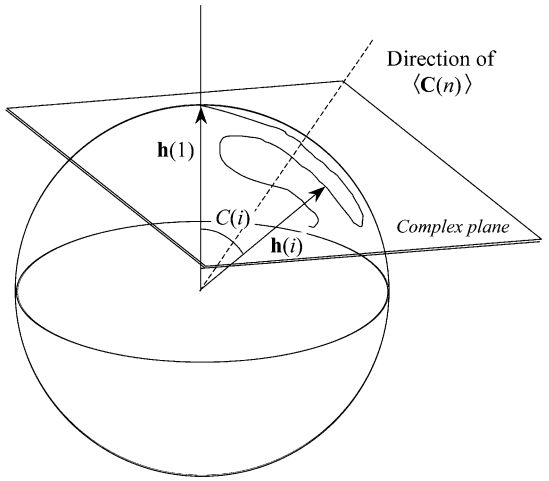

Fig. 3. Representation of the integral curvature on a sphere and a complex plane. The local axes of a DNA chain are translated at the center of the sphere so that their tips draw a curve on the surface. $\mathbf{h}(1)$ and $\mathbf{h}(i)$ represent the first and the ith local axis. The modulus of the integral curvature, $C(i)$, represents the angle between the ith local axis and the first. If DNA is slightly bent, the distribution of the local axes around their mean is narrow. They cover a small area of the spherical surface, which can be projected onto a complex plane as a consequence.

of the local axes of the DNA chain reported on a sphere. It corresponds to the mean square deviation along the sequence, in a certain time interval and with respect to its average value, of the integral curvature $\mathbf{C}(n,t)$, namely the angle between the first and the nth local helical axes

$$\sigma^2 = \left\langle \left| \mathbf{C}(n,t) - \left\langle \mathbf{C}(n,t) \right\rangle \right| \right\rangle^2 \tag{3}$$

In the case of little curvature, it is possible to project the tips of the vectors representing the local axes, lying on the surface of a sphere, on a plane (Fig. 3). Adopting a complex plane representation [see Eq. (1)], it follows

$$\mathbf{C}(n,t) = \mathbf{C}(n) + \Xi(n,t) \tag{4}$$

where, at the limit of a continuous chain, when it is possible to express terms as linear integrals along the chain, instead of summations:

$$\mathbf{C}(n) = \int_0^n \mathbf{c}(m)\mathrm{d}m \tag{5}$$

and $\Xi(n,t)$ is

$$\Xi(n,t) = \int_0^n \chi(m,t)\, dm \tag{6}$$

If we define static and dynamic contributions as

$$\sigma_s^2 = \left\langle \mathbf{C}(n) \middle| \mathbf{C}(n) \right\rangle - \left| \left\langle \mathbf{C}(n) \right\rangle \right|^2$$

and

$$\sigma_d^2 = \left\langle \Xi(n,t) \middle| \Xi(n,t) \right\rangle - \left| \langle \Xi(n,t) \rangle \right|^2$$

respectively, σ^2 becomes

$$\sigma^2 = \sigma_\mathrm{s}^2 + \sigma_\mathrm{d}^2 + 2Re\left\{ \left\langle \mathbf{C}(n) \middle| \Xi(n,t) \right\rangle - \left(\left\langle \mathbf{C}(n) \right\rangle \middle| \left\langle \Xi(n,t) \right\rangle \right) \right\} \tag{7}$$

The terms containing $\langle \Xi(n,t) \rangle$, which represents the integral dynamic fluctuations time average, equal to zero. Therefore, the mean square deviation, in the absence of a significant covariance between dynamic fluctuations and curvature, can be written as $\sigma^2 = \sigma_\mathrm{s}^2 + \sigma_\mathrm{d}^2$, and the dynamic contribution becomes $\sigma_\mathrm{d}^2 = \left\langle \Xi(n,t) \times \middle| \Xi(n,t) \right\rangle$.

The importance of the dynamic contribution on DNA mechanical properties was first advanced by Olson and coworkers [17].

The straightening energies for a bent and a standard straight DNA are respectively $\Delta G_\mathrm{cur}^{\ddagger} = k(\sigma_\mathrm{s}^2 + \sigma_\mathrm{d}^2)$ and $\Delta G_\mathrm{str}^{\ddagger} = k(\sigma_\mathrm{d}^{*2})$, being the static σ^2 of a straight DNA equal to zero by definition. k is a proportionality constant which depends on the experimental conditions, such as the temperature, the gel, the applied electric field as well as the DNA flexibility.

As a result, electrophoretic retardation can be represented as

$$R = \exp\left\{ \frac{k}{RT} \left[\sigma_\mathrm{s}^2 + (\sigma_\mathrm{d}^2 - \sigma_\mathrm{d}^{*2}) \right] \right\} \tag{8}$$

or, in the approximation of small curvatures

$$R = 1 + \frac{k}{RT} \left[\sigma_\mathrm{s}^2 + (\sigma_\mathrm{d}^2 - \sigma_\mathrm{d}^{*2}) \right] \tag{9}$$

In the case of bent DNAs the static term is overwhelming. Fig. 4 illustrates the correlation between the experimental retardation values and σ_s^2 in the case of approximately 450 sequences, investigated by various authors [6–9,18–21]. They are multimeric double helical oligomers or natural

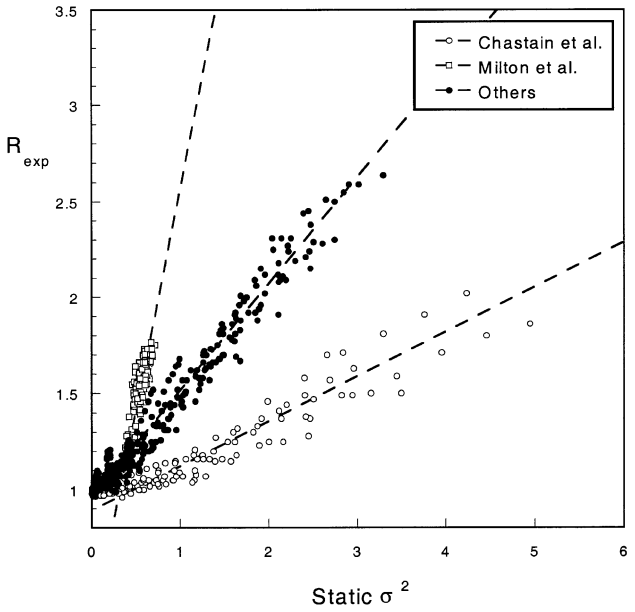

Fig. 4. Values of experimental retardation, R_{exp}, vs. the mean square deviation values, σ_s^2, for three groups of sequences. The one called 'others' contains data relative to sequences investigated by various authors [6–9,18]. The two remaining groups correspond to sequences studied by Chastain II et al. [20] and Chastain and Sinden [21], and by Milton et al. [19].

DNAs, different for sequence, periodicity and size (50–300 bp). We omitted some data corresponding to DNAs which are expected to form more than one superhelical turn, because they show deviations from linearity, probably due to more complicate interactions with the gel network, such as possible entanglements.

The points segregate into three different groups, probably due to the different experimental conditions in which data have been obtained. Anyway the correlation inside each group is very good. Among them, data from Milton et al. represent the most critical test of the model. In fact the results obtained in predicting the gel electrophoresis retardation changes due to extensive point mutations in a tract of 174 bp DNA of the simian virus 40 [19] are striking, proving that it is possible to evaluate the effect of only one base pair out 174 on the macroscopic behavior of the DNA molecule.

In the case of straight DNAs, where the static σ^2 is zero, it is important to evaluate σ_d^2. In the hypothesis that there is no correlation between dynamic fluctuations along the sequence

$$\sigma_d^2 = \frac{1}{N\Delta t}\int_0^N \int_0^{\Delta t} \Xi^*(n,t)\Xi(n,t)\mathrm{d}n\,\mathrm{d}t = \frac{1}{N\Delta t}\int_0^N \mathrm{d}n \int_0^n \mathrm{d}m \int_0^{\Delta t} |\chi(m,t)|^2 \mathrm{d}t \tag{10}$$

where N is the number of base pair of the sequence and Δt is a time interval long enough to consider statistical the behavior of the macromolecule.

Owing to the ergodic theorem, indicating $|\chi(m,t)|^2$ as χ^2 for simplicity, it results

$$\frac{1}{\Delta t}\int_0^{\Delta t} \chi^2 \mathrm{d}t = \frac{\int_0^\infty \chi^2 \exp\left(-\frac{b}{2RT}\chi^2\right)\chi\mathrm{d}\chi}{\int_0^\infty \exp\left(-\frac{b}{2RT}\chi^2\right)\chi\mathrm{d}\chi} = \frac{2RT}{b} \tag{11}$$

where b is the apparent elastic force constant of DNA. As a result, σ_d^2 can be written as

$$\sigma_d^2 = \frac{NRT}{b} \tag{12}$$

Dynamic terms are generally small and become relevant for straight DNAs only. It is worth noting that flexible DNAs have larger σ_d^2 than a standard DNA, whereas rigid DNAs have lower values. This means that higher retardation should result for flexible DNAs, whereas rigid DNAs have higher mobility, in agreement with the reptation theory [16].

However, the question about DNA flexibility is still controversial. As matter of fact, whilst the main determinants of DNA curvature are generally recognized, the evaluation of DNA flexibility in terms of sequence still remains an open problem. Thus the same dinucleotide steps are considered either flexible or alternatively rigid. In fact the dispersion of the orientational parameters of X-ray crystal structures of double helix oligonucleotides with different sequences, produces a scale of flexibility where the AA·TT step belongs to the rigid

Table 2
Rigidity parameters, f^{-1}, for each dinucleotide step, expressed in terms of the normalized melting temperatures [26] and stacking energies [28]

Dinucleotide step	T_{melt}/T^{*}_{melt}	E_{stk}/E^{*}_{stk}
AA·TT	0.945	0.703
AC·GT	1.070	1.323
AG·CT	0.956	0.780
AT	0.952	0.854
CG	0.997	1.124
GA·TC	1.037	1.230
GC	1.180	1.792
GG·CC	1.036	0.984
TA	0.894	0.615
TG·CA	0.945	0.790

class while GG·CC and GC·GC dinucleotides result to be more flexible [22,23].

In previous papers [24,25] we proposed a flexibility scale based on the double helix thermodynamic stability [26]. It shows predictive power for sequence-dependent nucleosome binding affinity. Recently, by analyzing images of DNA molecules, obtained by scanning force microscopy (SFM) [27], we found that dynamic fluctuations correlate to dinucleotide stacking energies [28]. The two scales, reported in Table 2, are correlated ($R=0.96$). This support the hypothesis that stacking energies are the most important determinant in stabilizing DNA double helix [29].

Consequently, the apparent elastic force constant of a dinucleotide, b, can be related to that of a standard DNA, b^*, by a rigidity parameter, $\langle f^{-1}\rangle$. It corresponds to the ratio of the dinucleotide stacking energies, averaged on the whole sequences, of the considered dinucleotide and that belonging to a standard DNA (Table 2). As a result $b=b^*\langle f^{-1}\rangle$, and

$$\sigma_d^2-\sigma_d^{*2}=\frac{NRT}{b^*}\left(\frac{1}{\langle f^{-1}\rangle}-1\right) \quad (13)$$

Therefore, from Eq. (9), for straight chains

$$R=1+\frac{k}{b^*}N\left(\frac{1}{\langle f^{-1}\rangle}-1\right) \quad (14)$$

Fig. 5 illustrates experimental data of electrophoretic retardations vs. $N(1/\langle f^{-1}\rangle-1)$ for some straight sequences studied by Chastain and Sinden [21]. It can be stated that there is an effective dependence of electrophoretic retardations on flexibility, here represented by the term $1/\langle f^{-1}\rangle$. This suggests the validity of both the described models and the adopted flexibility scale.

It is interesting to note that the value of the constant k obtained from Fig. 5 is 0.25, whereas the corresponding value from Fig. 4 for the data relative to Chastain's data is $k=0.21$. These values are very similar. In addition to this, it should be noted that the value for σ_d^2 in Eq. (12) was obtained considering fluctuations in a three-dimensional space, whereas it is reasonable to consider that the gel network partially decreases the dimensionality of the problem, as the DNA is restrained inside linear pathways. This corresponds to lower values for σ_d^2 and could take into account for the small deviation between the two estimated values of k.

4. Induced curvature as an elastic deformation of the intrinsic curvature

The analysis of DNA-protein interactions is an essential step towards a better understanding of

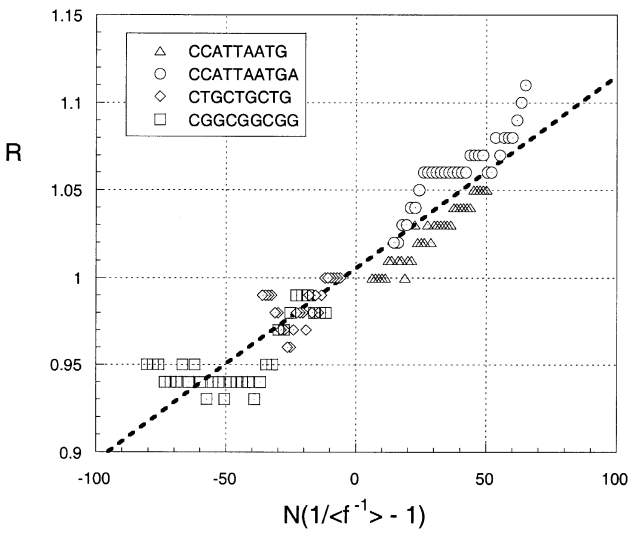

Fig. 5. Experimental data of electrophoretic retardation vs. $N(1/\langle f^{-1}\rangle-1)$ for some of the sequences studied by Chastain and Sinden [21].

molecular mechanisms of gene regulation. Proteins are capable of recognizing and amplifying as well as inducing curvature in biologically relevant DNA tracts. As a consequence of the intrinsic curvature, the relative amplitudes of the DNA grooves are proportionally and periodically changed with respect to the canonical B-DNA. This produces a differential electrostatic field along the curved DNA tracts, responsible for changes in hydration and affinity towards proteins.

Therefore, curvature could mediate the first step of the DNA-recognition process operated by proteins. Afterwards, the cooperation between short-range forces can result in an amplification of the curvature, as revealed in several cases by the permutation gel electrophoresis assays, performed in the presence and in the absence of proteins.

A similar mechanism holds for the nucleosome formation. It is noteworthy that binding of proteins such as histones, involved in chromatin organization, CAP and TBP, involved in the first step of transcription mechanism, or resolvase, involved in recombination, induces strong curvature on the corresponding DNA targets. To investigate the main determinants of the induced curvature, we adopted the principle of minimum deformation elastic free energy from the intrinsic curvature. On this basis we predicted the superstructures induced in DNA tracts by writhe transitions, circularization and nucleosome formation in good agreement with the experiments.

4.1. Catastrophic transition in topologically constrained DNA

Circular and looping DNAs are important in many fundamental biological processes; they represent topological domains where the local events and perturbations due to interactions with different kinds of molecules are integrated and lead to relevant superstructural transformations.

Since the DNA supercoiling discovery, great effort has been devoted to theoretical studies on circular and looping DNA represented as a deformable elastic filament, to investigate the superstructural changes that occur by increasing magnitude of linking numbers. Both analytical and numerical strategies were adopted to obtain the minimum energy structures of small circular filaments as a function of the linking difference. Several authors investigated the transitions of twist and writhe in topologically constrained circular DNAs [30–40].

A general conclusion of these researches was the existence of a catastrophic transition between the relaxed and the supercoiled forms of a circular DNA in response to twisting deformations; this conclusion can be also extended to looping DNAs.

More recently, some criticism was raised against mechanical models because the predicted transitions could have a physical meaning only at low temperature limits, where energy fluctuations and related entropy can be neglected. In order to overcome this problem, Gebe and Schurr [41] used Monte Carlo simulations to investigate the first transitions in writhe of a circular filament representing a 468 bp DNA. They adopted twisting and bending elastic parameters and an appropriate model of the Poisson–Boltzmann potential for the electrostatic contributions of a cylindrical worm-like DNA with a linear charge density. They obtained the mean free energy and internal energy changes as a function of writhe, W, for some fixed linking number differences, ΔL, ranging from zero to two turns. From these functions they easily calculated the ΔL value that induces the transformation between the circular and figure-8 forms; it resulted significantly lower and the transformation smoother than that predicted by simple mechanical models.

However, it is known that DNA circularization is strongly dependent on the sequence even for a DNA length of several hundreds of bp. Therefore, it is plausible that also the transition between relaxed and supercoiled forms, as a consequence of torsional stress induced by bound proteins, should be sequence directed. This can be relevant for the control of the gene expression that often depends upon bringing together DNA sequences separated by considerable distance along the double helix: e.g. supercoiled DNA template that facilitates loop formation in *lac* and *ara* repressors.

We successfully adopted a simple elastic model to approach the problem of the sequence-dependent writhe transformations [42].

Let $\mathbf{c}(n)$ be the curvature function of a circular relaxed DNA along the sequence number, n,

implicitly characterized by the linking number L and writhing number $W=0$, (the twisting number is obviously $L-W$ [43]). It can be conveniently given in terms of Fourier amplitudes, as:

$$\mathbf{c}(n)=\frac{1}{N}\sum_j \mathbf{A}(j)\exp\left(-\frac{2\pi \mathrm{i}jn}{N}\right) \tag{15}$$

where N is the bp number and

$$\mathbf{A}(j)=\sum_n \mathbf{c}(n)\exp\left(\frac{2\pi \mathrm{i}jn}{N}\right) \tag{16}$$

For the shift theorem [44], the Fourier amplitudes of the curvature function, representing the circular DNA after a uniform twisting along its axis are shifted in their frequencies by the twisting number, ΔT

$$\begin{aligned}\mathbf{A}'(j)&=\sum_n\left[\mathbf{c}(n)\exp\left(\frac{2\pi \mathrm{i}n\Delta T}{N}\right)\right]\exp\left(\frac{2\pi \mathrm{i}jn}{N}\right)\\&=\mathbf{A}(j+\Delta T)\end{aligned} \tag{17}$$

In fact, in our curvature representation, twisting around the local DNA axes is equivalent to multiply the curvature function by $\exp(2\pi \mathrm{i}n\Delta T/N)$ along the sequence number.

Therefore the series of $\mathbf{A}(j+\Delta T)$ represent a DNA superstructure with the same curvature moduli, but uniformly twisted along the sequence of $2\pi n\Delta T/N$ radians.

Such a transformation does not generally ensure the circularization integrity; nevertheless, to correctly seal the DNA strands it is sufficient to refine the modulus of the Fourier amplitudes (this preserves the twisting) that are directly involved in circularization. Such procedure is obtained adopting a steepest-descent algorithm that ensures a fast convergence.

At the end of this procedure we obtain a new curvature function (by inverse Fourier transform) which represents a circular DNA characterized by a writhing number $-W=\Delta T$, as the linking number, L, is constant in the whole process.

This was verified by direct writhing number calculation using the discretized form of Gauss' integral [45]

Table 3
Values of the effective amplitudes as a function of the writhing number for a hypothetical infinitely long DNA (the values for negative writhes is symmetrically related)

Writhing number	Moduli of the effective amplitudes (rad/π)		
	$A(-W)$	$A(2-W)$	$A(4-W)$
0	2.000	–	–
0.1	2.050	1.065	–
0.2	2.101	1.468	–
0.3	2.153	1.745	–
0.4	2.204	1.952	–
0.5	2.254	2.109	–
0.6	2.300	2.226	–
0.7	2.341	2.309	–
0.8	2.373	2.364	–
0.9	2.396	2.395	–
1	2.405	2.405	–
1.1	2.476	2.262	0.609
1.2	2.513	2.182	1.006
1.3	2.533	2.128	1.307
1.4	2.539	2.088	1.547
1.5	2.535	2.058	1.745
1.6	2.523	2.035	1.916
1.7	2.496	2.019	2.054
1.8	2.459	2.008	2.169
1.9	2.408	2.002	2.265
2	2.344	2.000	2.344

$$W=\frac{1}{4\pi}\sum_{i,j}\frac{\mathbf{l}_i\times\mathbf{l}_j\cdot\mathbf{r}_{ji}}{r_{ji}^3} \tag{18}$$

where $\mathbf{l}_i$ and $\mathbf{l}_j$ represent the unit elements of the DNA axis and $\mathbf{r}_{ji}$ their vectorial distance. For sake of clarity, in the whole paper we adopted the italic fonts to represent the moduli of the complex quantities indicated by bold fonts.

It is interesting that practically a few Fourier modes are able to determine supercoiled structures when the shift theorem is considered. In fact, it should be noted that for $W=0$ the effective amplitude is that with frequency zero, corresponding to W. In the case of $W=1$, the effective amplitudes are those with frequency -1 and 1, namely $-W$ and $2-W$. For $W=2$, they are those with frequency -2, 0 and 2, that is $-W$, $2-W$ and $4-W$.

In general, the effective amplitudes seem to be those with frequency $j-W$, when j is even. On the contrary, amplitudes with frequency $j-W$, with j

odd, perturb the symmetry of the structure and should remain equal to zero.

For example, Table 3 reports the limit values assumed by the effective amplitudes for a hypothetical infinitely long DNA chain, for which excluded volume effects are negligible.

It is worth noting that, following Parseval's equality [46], we have to change the minimum number of Fourier amplitudes to select the ground-state conformation of a circular DNA for each writhing number. It ensures the minimum square average deviation between two functions in terms of the sum of the corresponding squared Fourier amplitude differences

$$\left\langle \Delta c(n)^2 \right\rangle = \frac{1}{N^2} \sum_j \Delta A(j-W)^2 \tag{19}$$

If we maintain all the amplitudes invariant, namely equal to the values characterizing the intrinsic curvature, except for the moduli of the amplitudes necessary to the circularization at a certain writhing number, we obtain the ground-state bending (first order) elastic energy (the condition of minimum energy requires in fact, the invariance of the phases):

$$E^{\circ}_{\mathrm{ben}} = N\frac{b}{2}\left\langle \Delta c(n)^2 \right\rangle = \frac{b}{2N} \sum_k \Delta A(k-W)^2 \tag{20}$$

where the sum extends over the k amplitudes determined by the geometrical constraints, as already discussed.

Let $\mathbf{B}(j)$ be the Fourier amplitudes necessary to represent the twist along the sequence and t the apparent twisting force constant. The corresponding twisting energy is

$$E^{\circ}_{\mathrm{tws}} = \frac{t}{2N} \sum_j \Delta B(j)^2 \tag{21}$$

However, the only geometrically constrained term is $\Delta\mathbf{B}(0)$, because it corresponds to 2π times the twisting number difference between the linear and circular forms. Being $\Delta T = \Delta L - W$, where ΔL is the linking number difference between the circular and the linear intrinsic form, it follows

$$E^{\circ}_{\mathrm{tws}} = \frac{t}{2N}(2\pi\Delta T)^2 = \frac{2\pi^2 t}{N}(\Delta L - W)^2 \tag{22}$$

All the remaining energy contributions arising from water molecules and counterions interactions are considered to be invariant for all the writhes. This is justified by the fact that writhe transformations induce only slight changes at the local molecular level, at least for a low supercoiling density; therefore negligible energy contributions are involved. This is not true for macro-counterions as proteins, whose interactions extend over many dinucleotide steps.

Consequently, in our model, writhe changes and dynamic fluctuations practically occur in a constant external force field and are driven only by an internal elastic potential, at least for the first transitions in writhe.

Adopting such a strategy we can easily calculate the bending and twisting (first order) elastic energy of the ground states for different writhes and ΔL. Such calculations were first performed for a worm-like 468 bp DNA, adopting the same bending and twisting force constants of Gebe and Schurr [41] in order to try a comparison with their Monte Carlo simulations.

The results are very significant, as illustrated in Fig. 6, where the Monte Carlo mean internal energy is compared with the theoretical ground-state elastic energy. The two sets of values differ for a strictly constant energy contribution, that clearly indicates a practical equivalence of both the electrostatic and solvent interactions vs. the writhing number as well as of the energy fluctuations.

In addition, the good agreement between the two sets of data proves the validity of our model in finding the ground-state curvature for a given DNA at a certain writhing. This is necessary to study superstructural transformations toward circular forms, as will be shown below.

4.2. Sequence-dependent circularization propensity

Circularization of a DNA tract is the result of stochastic and deterministic motions of the double helix, so that DNAs characterized by in phase bends have a higher probability of circularization than straight DNAs.

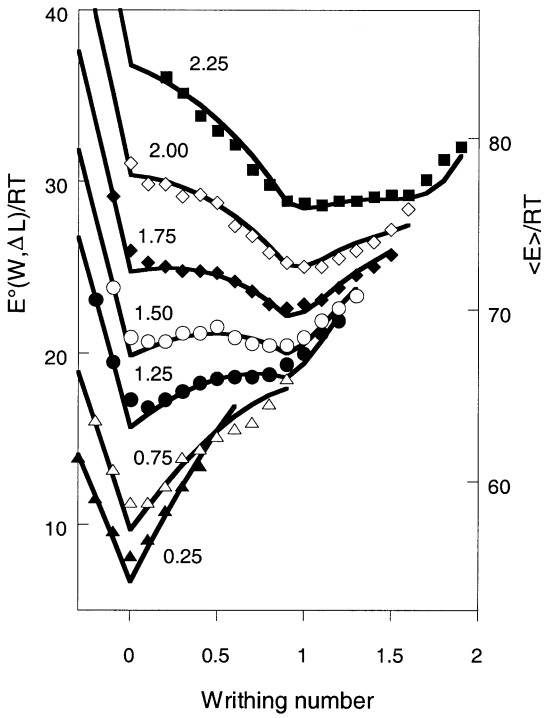

Fig. 6. Ground-state first order elastic energy, $E°$, (solid lines) of a uniform 468 bp DNA vs. the writhing number, W. Numbers close to the lines indicates the different linking number differences, ΔL. The same bending and twisting force constants of Gebe and Schurr [41] (corresponding to a persistence length $P=500$ Å and a torsional rigidity $C=2.0\times10^{-19}$ erg cm) were adopted for an immediate comparison with the mean internal energy, $\langle E\rangle$ (different symbols) of their Monte Carlo simulations.

The ring closure probability is conveniently represented by the J factor. First defined by Jacobson and Stockmayer [47], it corresponds to the ratio of the equilibrium constant for cyclization, K_{circ}, and that for bimolecular association, K_{dim}, expressed as

$$\mathrm{L}\rightleftarrows\mathrm{C}\quad K_{circ}=\frac{[C]}{[L]} \tag{23}$$

and

$$2\mathrm{L}\rightleftarrows\mathrm{D}\quad K_{dim}=\frac{[D]}{[L]^2} \tag{24}$$

where [C], [L] and [D] indicate the concentrations of the linear, circular and dimeric DNAs. It follows that

$$J=\frac{K_{circ}}{K_{dim}}=\frac{[C][L]}{[D]} \tag{25}$$

According to Jacobson and Stockmayer, in the case of a long and flexible DNA molecule with complementary cohesive ends, where the orientations of the terminal chain elements are not correlated, the Gibbs free-energy variation for circularization, $\Delta G°_{circ}$, simply depends on $W(0)\delta V$, which represents the probability of finding both the chain terminals in the same infinitesimal volume δV.

$$\Delta G°_{circ}=\Delta G^{*}-RT\ln\left(W(0)\delta V\right) \tag{26}$$

In the same way, the free-energy variation for dimerization can be expressed as

$$\Delta G°_{dim}=\Delta G^{*}-RT\ln(\mathcal{N}\delta V) \tag{27}$$

where $\mathcal{N}$ is Avogadro's number.

The term ΔG^{*} accounts for all the interactions among chain terminal elements and it is the same for both the equations. Therefore, it is possible to obtain the ring closure probability as

$$J=\exp\left(-\frac{\Delta G°_{circ}-\Delta G°_{dim}}{RT}\right)=\frac{W(0)}{\mathcal{N}} \tag{28}$$

R is the gas constant, whereas $W(0)$ corresponds to the spatial probability density, $W(r)$, where r is the end-to-end distance of the chain, evaluated at $r=0$.

Assuming that the DNA can be modeled as a Gaussian chain, composed by N elements of length l, $W(0)$ can be easily calculated as

$$J=\frac{1}{\mathcal{N}}\left(\frac{3N}{2\pi Nl^2}\right)^{3/2} \tag{29}$$

The Jacobson–Stockmayer theory well agrees with experiments performed on long DNA molecules, as demonstrated by the experiments of Wang and Davidson [48,49]. On the contrary, in the case of short and stiff fragments, the assumptions of

Gaussian statistics and no correlation between the orientations of the ends of the chain do not hold. In this case, they would seem to be more similar to 'rod-like' than to 'freely-jointed' chains [50].

It would be useful to follow Kratky and Porod [51] and to model DNA as a worm-like chain, as proposed by Yamakawa and Stockmayer [52]. They also accounted for excluded volume effects, which become important for short chains.

Later, Flory and co-workers expressed the cyclization and dimerization probabilities in an elaborated form of the Jacobson-Stockmayer theory [53].

They considered the probability that a DNA will cyclize depending on the simultaneous fulfillment of three conditions: (1) the ends of the chain must be located at a distance such that the new bond can be formed; (2) the angle between the helix axes at the ends of the chain must be within the range allowed by the bendability of the DNA; (3) the torsion angle between the ends must be such that the helix backbone is continuous at the newly created bond.

It is possible to define $\Gamma_r(\gamma)$ as the conditional probability density for the scalar product γ of the vectors parallel to the helix axes at the chain termini, at a specified value of the end-to-end distance r. Similarly, $\Phi_{r,\gamma}(\tau)$ is the conditional probability density for the torsional angle τ, between the vectors which define the helical orientation of the chain ends, with r and γ, specified. Therefore, J can be written as follows:

$$J = \frac{4\pi\, W(0)\Gamma_0(1)\Phi_{0,1}(\tau_0)}{\mathscr{N}} \tag{30}$$

where the spatial probability density is evaluated at $r=0$, $\Gamma_r(\gamma)$ is evaluated when $r=0$ and $\cos(\gamma)=1$ (parallel ends) and $\Phi_{r,\gamma}(\tau)$ is evaluated when $r=0$, $\cos(\gamma)=1$ and $\tau=\tau^\circ$, the angle corresponding to the perfect torsional alignment of the chain.

For a worm-like chain, an exact closed form for the probability density in Eq. (30) has not been derived. Some approximate distributions have been proposed [54,55], but they are not generally applicable to systematically bent molecules. In order to overcome these limitations, some authors used Monte Carlo simulations to calculate ring closure probabilities [56–60].

In this paper, a statistical-mechanical model is used to calculate circularization probability. We have modeled the circularization process of a DNA tract constraining the ends of the two strands to a correct ring closure. Therefore, circularization probability will also depend on the required deformation energy and Eq. (26) becomes

$$\Delta G^\circ_{\mathrm{circ}} = \Delta G^* - RT\ln\big(W(0)dV\big) + \Delta E_{\mathrm{circ}} \tag{31}$$

The spatial probability density, $W(0)$, has been evaluated assuming a Gaussian distribution of end-to-end distances calculated following Kratky and Porod [51]. As a consequence, the J factor becomes

$$J = \frac{1}{\mathscr{N}}\left(\frac{3}{2\pi\langle\, r^2_{KP}\rangle}\right)^{3/2}\exp\left(-\frac{\Delta E_{\mathrm{circ}}}{RT}\right) \tag{32}$$

$\langle r^2_{KP}\rangle$ is the average value of the squared end-to-end distance: it depends mainly on the persistence length and could be sensitive to the flexibility distribution along the chain. However, changing the persistence length in the range of $\pm 10\%$ seems to have only slight effects on the pre-exponential term.

The term ΔE_{circ} includes contributions due to both the ground-state elastic energy and the thermic fluctuation differences. They have been evaluated by calculating the Fourier amplitudes necessary to fulfil the geometrical properties of the circular form.

However, there are some fluctuations that cause changes in the writhing number. It is worthy of noting that this requires a simultaneous variation of two or three amplitudes (see Table 3) and the relative deformation energy strongly depends on the intrinsic curvature [42].

Therefore, the presence of all the possible structures with different writhing number is considered as a result of a series of parallel circularization reactions and the total thermodynamic circularization constant is

$$K_{\mathrm{circ}} = \frac{\sum_W [C]_W}{[L]} = \sum_W (K_{\mathrm{circ}})_W = \sum_W \exp\left(-\frac{\Delta G_{\mathrm{circ}}}{RT} \right)_W \quad (33)$$

where the subscript W represents the writhing number of each circular structure considered in Eq. (33). Circularization free energy relative to each equilibrium is

$$(\Delta G_{\mathrm{circ}})_W = \Delta G^* - RT\ln\left(W(0)\mathrm{d}V\right) + (\Delta E_{\mathrm{circ}})_W \quad (34)$$

As a result, deformation energy, ΔE_{circ}, in Eq. (31) can be simply expressed as

$$\Delta E_{\mathrm{circ}} = -RT\ln\sum_W \exp\left(-\frac{\Delta E_{\mathrm{circ}}}{RT} \right)_W \quad (35)$$

Each term $(\Delta E_{\mathrm{circ}})_W$, contains the two contributions due to ground-state $(\Delta E^{\circ}_{\mathrm{circ}})_W$, and fluctuations $\langle \Delta E_{\mathrm{circ}} \rangle_W$. The first, can be defined with the simple harmonic approximation, as the sum of the energies required by the bending and twisting transformations. They are evaluated according to Eqs. (20) and (22), following Parseval's theorem [46].

Calculating the thermodynamic equilibrium constant between two DNA forms requires also the estimate of those thermal fluctuations which do not cause changes in the writhing number. Fixing the linking and the writhing numbers, we can define a superstructure ensemble characterized by an average energy and entropy. Therefore, by definition, the term $\langle \Delta E_{\mathrm{circ}} \rangle_W$, is

$$\langle \Delta E_{\mathrm{circ}} \rangle_W = RT^2 \frac{\partial \ln(q_{\mathrm{C}}/q_{\mathrm{L}})_W}{\partial T} \quad (36)$$

where q_{C} and q_{L} are the partition functions for the circular and linear forms, respectively. In the hypothesis that the bending and twisting contributions are independent, for fixed linking and writhing number, the partition functions can be written as $q(W, \Delta L) = q_{\mathrm{ben}}(W)\, q_{\mathrm{tws}}(W, \Delta L)$, where q_{ben} and q_{tws} refer to the bending and twisting contributions respectively. The first one can be calculated as follows, recalling Eq. (20),

$$q_{\mathrm{ben}}(W) = \sum g(W) \exp\left(-\frac{\Delta E_{\mathrm{ben}}}{\mathrm{RT}} \right) = \sum g(W) \exp\left(-\frac{b}{2NRT} \sum_j \Delta A_j^2 \right) \quad (37)$$

where $g(W)$ accounts for energy states degeneration and ΔA_j refers to amplitude fluctuations. As amplitudes vary continuously in the complex plane

$$q_{\mathrm{ben}}(W) = \prod_j \int_0^{\Delta \mathbf{A}_j^M} \exp\left(-\frac{b}{2NRT} \Delta A_j^2 \right) d\Delta \mathbf{A} = \prod_j \frac{2\pi NRT}{b} \times \left[1 - \exp\left(-\frac{b}{2NRT} \right) \Delta A_j^{M^2} \right] \quad (38)$$

where ΔA_j^M is the maximum modulus fluctuation of the amplitude with frequency j. If fluctuations are large, as in the linear form, the values of ΔA_j^M are high enough to let the corresponding bending integrals in Eq. (38) be $2\pi NRT/b$. As a result, the ratio of the bending partition functions for the circular and the linear forms at a certain writhing number, W, becomes

$$\left(\frac{q_{\mathrm{C}}}{q_{\mathrm{L}}} \right)_{\mathrm{ben},W} = \prod_j \left[1 - \exp\left(-\frac{b}{2NRT} \Delta A_j^{M^2} \right) \right]_{\mathrm{C}} \quad (39)$$

which only depends on the terms relative to the circular form. Therefore the corresponding contribution to $\langle \Delta E_{\mathrm{circ}} \rangle_W$, is

$$\langle \Delta E_{\mathrm{circ}} \rangle_{\mathrm{ben},W} = RT^2 \frac{\partial \ln}{\partial T} \left(\frac{q_{\mathrm{C}}}{q_{\mathrm{L}}} \right)_{\mathrm{ben},\,W} = \left[\sum_j \frac{-\frac{b}{2N} \Delta A_j^{M2}}{\exp\left(\frac{b}{2NRT} \Delta A_j^{M^2} \right) - 1} \right]_{\mathrm{C}} \quad (40)$$

Actually, evaluation of ΔA_j^M is easier than one could expect, as the effect of amplitude fluctuations on structure do not depend on its writhing number.

Fig. 7 shows the derivative of the end-to-end distance with respect to the amplitude perturbation, vs. the frequency: a Gaussian-like curve interpolates all the points corresponding to structures with different writhing numbers. It is worth noting that only the fluctuations of the amplitudes with low frequencies can distort the structure. As a consequence, to preserve the closed circular structure, their modulus can change only within limits. On the contrary, amplitudes with higher frequencies are unconstrained and can assume any value. As a result, the corresponding bending contributions in Eq. (40) tend to $-RT$.

Consequently we chose to represent $\Delta A_j^M = NA/G(j)$, where A is a parameter that takes into account for the fluctuations per base pair and $G(j)$ is the normalized Gaussian from Fig. 7, which contains the dependence on the frequency j. When $G(j)$ is small, for high frequencies, ΔA_j^M is large, whereas for lower frequencies, the maximum fluctuations are restrained by the value assumed by A. As a result, the summation in Eq. (40) can be conveniently substituted by the integral

$$\langle \Delta E_{\mathrm{circ}} \rangle = \frac{1}{\sigma} \int \frac{-bN(A/G(j))^2}{\exp\left(\frac{bN}{2RT}(A/G(j))^2\right) - 1} \mathrm{d}j \qquad (41)$$

where σ is the integration factor. It is worth noting that when N is small, the circular form becomes rigid enough to avoid effective fluctuations. Therefore σ should be chosen to have $\langle \Delta E_{\mathrm{circ}} \rangle_{\mathrm{ben}} = -3RT$, corresponding to fix in modulus and phase the Fourier modes corresponding to the frequencies 0, 1 and -1, necessary to obtain the circular form with $W=0$. On the contrary, when N is large, $\langle \Delta E_{\mathrm{circ}} \rangle_{\mathrm{ben}}$ tends to zero, as the fluctuations of the circular and linear forms are practically equivalent. The parameter A modulates the rate at which $\langle \Delta E_{\mathrm{circ}} \rangle_{\mathrm{ben}}$ changes as a function of N.

Analogously, taking into account for Eq. (22),

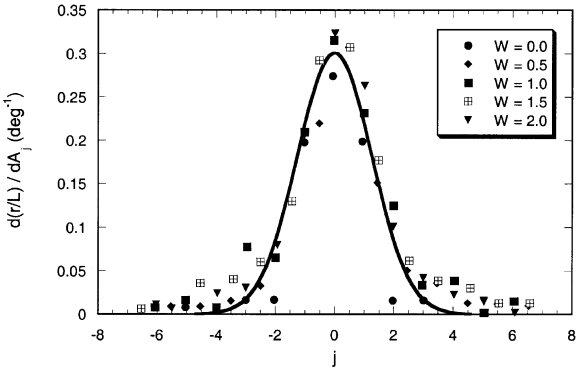

Fig. 7. Gradient components of the end-to-end distance, r, normalized with the contour length, L, for 10 degree increments of the $\mathbf{A}(j)$ amplitudes. The symbols refer to some forms with different writhing numbers; the solid line is the Gaussian fitting.

the twisting contribution is evaluated as

$$q_{\mathrm{tws}}(W,\Delta L) = \prod_j \int_0^{\Delta \mathbf{B}_j^M} \mathrm{e}^{-\frac{t}{2NRT}\Delta B_j^2} d\Delta \mathbf{B}_j$$

$$= \prod_j \left[\frac{2\pi NRT}{t} \times \left[1 - \mathrm{e}^{\left(-\frac{t}{2NRT} \Delta B_j^{M2} \right)} \right] \right]^{1/2} \qquad (42)$$

The exponent $\frac{1}{2}$ takes into account the twisting is a real function. The only constrained amplitude in the circular form is that with frequency equal to zero, because it is proportional to the twisting number, as discussed before [see Eq. (22)]. All the other amplitudes are unconstrained and the corresponding integrals tend to $(2\pi NRT/t)^{1/2}$.

However, it can be stated that in the framework of the energy equipartition, each torsional degree of freedom contributes in equal measure with respect to the bending ones and could be calculated accordingly. Namely, σ could be chosen to have $\langle \Delta E_{\mathrm{circ}} \rangle = -3.5\ RT$ when N is small (for example $N=100$).

Applying these results to Eq. (32), the J factor becomes

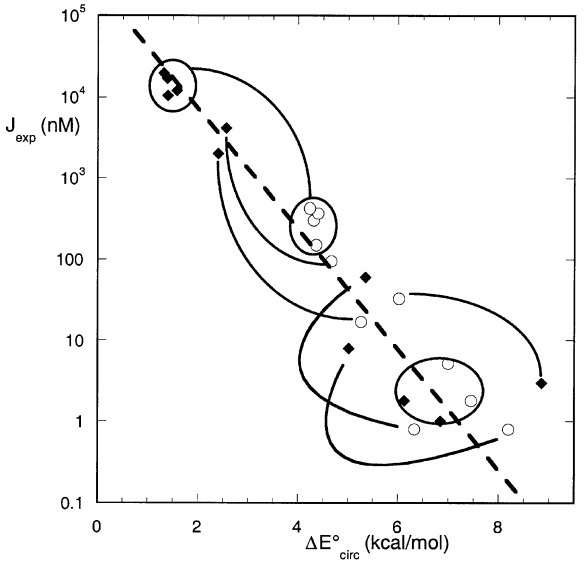

Fig. 8. Theoretical values of circularization free energy vs. *J*-factor experimental data relative to sequences characterized by the presence of the CAP binding-site [58]. Lines connect points referring to the same sequences in the presence (♦) or absence (○) of the protein.

$$J = \frac{1}{\mathcal{N}}\left(\frac{3}{2\pi}\right)^{3/2}\langle r_{KP}^2\rangle^{-3/2}\exp\left(-\frac{\langle \Delta E_{\mathrm{circ}}\rangle}{RT}\right) \times \exp\left(-\frac{\Delta E^{\circ}_{\mathrm{circ}}}{RT}\right) \quad (43)$$

where the numerical term $1/\mathcal{N}(3/2\pi)^{3/2}$ is assumed equal to 5.5×10^{11} when J is given in nM and $\langle r_{KP}^2\rangle$ in Ångstroms.

The described model for calculating the circularization probability was applied to a set of experimental data obtained by several authors [58,61–64].

Bending force constant b, has been obtained according to the relation $b=RTP/l$, where P, the persistence length, and l, the helix rise per base pair, are 450 Å and 3.4 Å, respectively. Therefore b is equal to 78.4 kcal mol^{-1} rad^{-2}. Twisting force constant t is 88.8 kcal mol^{-1} rad^{-2}. The values adopted for those force constants correspond to averaging those found in the literature.

As for the A value in Eq. (41), we assumed a tentative value equal to 1.0×10^{-3} rad.

In Fig. 8 are reported theoretical values of circularization free energy vs. *J*-factor experimental data relative to the sequences studied by Kahn and Crothers [58] characterized by the presence of the binding-site for CAP protein. Therefore, for such sequences it was necessary to calculate circularization probability for the bound and free DNA molecule.

We evaluated the *J*-factor in presence of CAP protein by imposing a curvature of approximately 90° to the binding site, according to the crystallographic structure [65].

Fig. 8, shows a good correlation between the two sets of data: the model we developed works satisfactorily in predicting circularization probability also in the presence of a protein that influences intrinsic curvature.

Finally, Fig. 9 shows the correlation between *J*-factor theoretical values and the corresponding experimental data obtained by different authors [58,61–64]. There is a good agreement especially for those sequences with an high circularization probability. It is interesting that all the values spread in a range of approximately six orders of magnitude, even if most sequences have practically the same length (approx. 150–160 bp). This is a

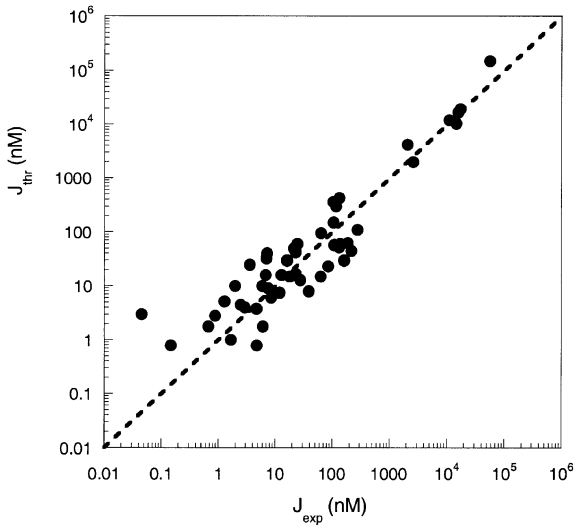

Fig. 9. Correlation between *J*-factor theoretical values and the corresponding experimental data obtained by different authors [58,61–64].

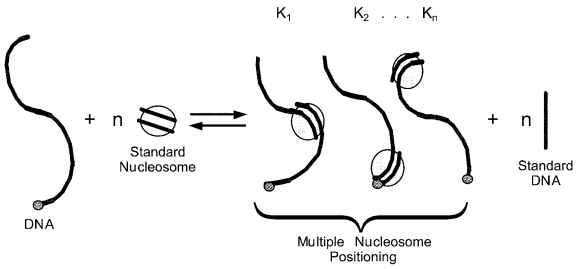

Fig. 10. Schematic drawing of competitive nucleosome reconstitution for a certain DNA, long enough to form more than one nucleosome. Each nucleosome positioning is characterized by an equilibrium constant (K_1, K_2, … K_n). An ideal straight random-sequence of a DNA of 144 bp with a uniform average flexibility is adopted as a standard. Adopting a different standard corresponds to adding a constant term to the free-energy difference; consequently, it does not affect relative nucleosome stability and positioning.

demonstration that *J*-factor values strongly depend on curvature effects.

Very recently, Roychoudhury et al. [60], published a set of circularization probabilities data relative to some sequences characterized by the nucleosome positioning site (NPS). In this case the model does not allow theoretical values to be obtained which are in agreement with the experimental values. Anyway, this could be justified by the particular mechanical properties of the binding site, according to the conclusions drawn by the authors, and/or by the non-uniform distribution of the flexibility along the DNA chain, which influences the average value of the end-to-end distance. However, this effect is at present under investigation.

4.3. A simple elastic model for predicting the nucleosome positioning

Nucleosome is the association complex of DNA (145 bp) with the histone octamer; it is the elemental unit of chromatin. The first model of nucleosome was proposed by Kornberg [66], while its structure was later determined at low resolution by Klug et al. [67] by means of electron microscopy and by Richmond et al. [68] by x-ray diffraction techniques. It is characterized by a flat solenoid-like structure where the DNA axis is wrapped on the proteic core of the histone octamer with a pseudo-dyad symmetry. Recently, a 2.8 Å resolution electron density map was obtained by Luger et al. [69] which confirms the previous structure and contains relevant details of both DNA and protein core.

The experimental evidence indicates that the reconstitution of nucleosome can be practically obtained with any DNA sequence, although a number of papers have shown the existence of preferential positioning of nucleosomes along DNA as well as the occurrence of nucleosome-free regions in chromatin. Therefore, the ability to form nucleosomes appears to be enhanced or decreased by some base sequences.

Competitive nucleosome reconstitution experiments (Fig. 10) provide a quantitative estimate of differential thermodynamic nucleosome stability, indicating that nucleosome positioning along a DNA sequence occurs with different affinity; therefore, a large set of experimental data of nucleosome reconstitution free energy has been accumulated in literature.

Attempts to find the sequence features controlling the stability of nucleosomes, were made in different laboratories with the aim to identify the genomic location of such sequences in the organization of chromatin [70–81].

The experimental results show that the affinity of the histone octamer to DNAs with different sequences, curvature and length, appears to be restricted within only a few kcal per mole of nucleosome. This could suggest that the DNA-histones recognition could be driven by simple elastic energy differences resulting from the DNA intrinsic curvature and/or flexibility.

By adopting the same approach described before, in the framework of first-order elasticity, we calculated the canonical ensemble free energy to transform the DNA intrinsic superstructure into the nucleosomal structure, and evaluated the thermodynamic equilibrium of the competitive nucleosome reconstitution experiments.

If ΔG_k represents the nucleosome reconstitution free-energy difference from a standard nucleosome of the *k*th DNA tract of $l = 144$ bp along a sequence, the free energy per mole of nucleosome, ΔG, pertinent to the whole DNA chain with *N* bp

is

$$\Delta G = -RT\ln\sum_{k=l/2}^{N-l/2}\exp\left[-\frac{\Delta G\,(k)}{RT}\right] \tag{44}$$

The relation with the pertinent canonical partition functions allows us to write the nucleosome reconstitution free energy difference as:

$$-\frac{\Delta G\,(k)}{RT} = \ln\frac{Q_{\mathrm{n}}(k)}{Q_{\mathrm{n}}^{*}} - \ln\frac{Q_{\mathrm{f}}(k)}{Q_{\mathrm{f}}^{*}} \tag{45}$$

where $Q_{\mathrm{n}}(k) = \sum\exp(-\Delta E_j(k)/RT)$ is the configurational canonical partition function of the kth nucleosomal DNA tract along the sequence; $Q_{\mathrm{f}}(k)$ is that of the corresponding free DNA; Q_{n}^{*} and Q_{f}^{*} are those pertinent to a standard intrinsically straight DNA with a random sequence. The partition functions of the flanking DNA tracts are considered unaffected by the nucleosome formation and cancel in the ratio.

The partition functions were first evaluated taking into account the elastic terms, namely the sum of the bending, $\Delta E_{\mathrm{ben}}(k)$, and twisting, $\Delta E_{\mathrm{tws}}(k)$, energies necessary to distort the kth 144 bp DNA tract in the nucleosomal form.

The elastic twisting contribution can be expressed, as in Eq. (22)

$$\Delta E_{\mathrm{tws}}(k) = \frac{2\pi^2 t}{l}\left[\Delta T(k)\right]^2 \tag{46}$$

where t is the twisting force constant and $\Delta T(k)$ represents twisting number difference between the nucleosome and the free kth DNA tract. The latter term was evaluated imposing a DNA periodicity of 10.2 bp per turn according to the experimental evidence.

Owing to Parseval's theorem [46], the only necessary condition to have the nucleosome-like curvature, suggested by the crystallographic structure [69], is to constrain the Fourier term with periodicity $\mu = -0.18$ to assume the value 10.9 rad leaving the maximum of the intrinsic curvature features invariant in order to minimize the elastic energy.

The elastic bending energy contribution can be expressed as Eq. (20)

$$\Delta E_{\mathrm{ben}}(k) = l\frac{b}{2}\left\langle\Delta c_k(n)^2\right\rangle = \frac{b}{2l}\left[\Delta A(\mu)\right]^2 \tag{47}$$

where b refers to the apparent isotropic bending force. Both b and t can be expressed as $b = b^*\langle f^{-1}\rangle$ and $t = t^*\langle f^{-1}\rangle$, where $\langle f^{-1}\rangle$ is a rigidity parameter that allows accounting for sequence-dependent flexibility and b^* and t^* are the bending and twisting force constants for a standard DNA.

Let $A_{\mathrm{n}}(\mu)$ and $A^{\circ}_{\mathrm{f}}(\mu)$ be the Fourier terms pertinent to nucleosome and free DNA, respectively.

Setting $A_{\mathrm{n}} = A_{\mathrm{n}}(\mu)$, $A^{\circ}_{\mathrm{f}} = A^{\circ}_{\mathrm{f}}(\mu)$ and $B = (b/lRT)\,\langle f^{-1}\rangle$ for the sake of clarity

$$Q_{\mathrm{n}}(k) = \exp\left[-\frac{\Delta E_{\mathrm{tws}}(k)}{RT}\right]\exp\left[-\frac{B}{2}(A_{\mathrm{n}}^2 + A^{\circ 2}_{f})\right]\int\exp(BA_{\mathrm{n}}A^{\circ}_{f}\cos\phi)\,\mathrm{d}\phi \tag{48}$$

where ϕ is the relative phase angle between $A_{\mathrm{n}}(\mu)$ and $A^{\circ}_{\mathrm{f}}(\mu)$. The last integral becomes equal to 2π times $J_0(iZ)$, the zero-order Bessel function of the imaginary argument $Z = BA_{\mathrm{n}}A^{\circ}_{\mathrm{f}}$. Therefore

$$Q_{\mathrm{n}}(k) = 2\pi\exp\left[-\frac{\Delta E^{\circ}(k)}{RT}\right]\exp(-Z)J_0(iZ) \tag{49}$$

where $\Delta E^{\circ}(k)$ contains both the ground-state bending and twisting energy contributions. For the standard nucleosome, $A^{\circ}_{\mathrm{f}} = 0$ and therefore

$$\frac{Q_{\mathrm{n}}(k)}{Q_{\mathrm{n}}^{*}} = \frac{\exp\left[-\Delta E^{\circ}(k)/RT\right]\exp(-Z)J_0(iZ)}{\exp[-\Delta E^{\circ *}/RT]} \tag{50}$$

It should be stressed that only the ground states are represented in the partition function ratio of nucleosomal DNA because the association with the histone core is considered to freeze the DNA superstructure in the minimum elastic energy state.

The ratio of the canonical partition functions of the free DNAs is equal to the ratio of the bending and twisting fluctuation terms, analogous to those in Eqs. (38) and (42) for the linear DNA. In fact, the free DNA fluctuates around the ground-state

superstructure, which corresponds to the intrinsic curvature. Therefore, as $b=b^*\langle f^{-1}\rangle$ and $t=t^*\langle f^{-1}\rangle$ and since for the standard nucleosome $\langle f^{-1}\rangle=1$, the ratio of the pertinent canonical partition functions becomes practically equal to:

$$\frac{Q_f(k)}{Q^*_f}=\langle f^{-1}\rangle^{-(3/2)l} \tag{51}$$

where $3/2\ l$ corresponds to the sum of the number of degrees of freedom of the bending and twisting modes. Consequently:

$$\frac{\Delta G_{el}(k)}{RT}=\frac{\Delta E^{\circ}(k)}{RT} -\frac{3}{2}l\ln\langle f^{-1}\rangle+Z-\ln J_0(iZ) \tag{52}$$

$\Delta E^{\circ}(k)$ corresponds to the minimum elastic energy required to distort the kth tract in the nucleosomal form. A°_f represents the effective curvature, namely the Fourier term of the free DNA curvature function, which coherently contributes to the nucleosomal structure; whereas fixing A_n is a necessary condition to ensure that DNA assumes a nucleosome-like form. The term $-3/2l\ln\langle f^{-1}\rangle$ represents the free DNA loss of degrees of freedom during nucleosome formation. It has effect on elastic free energy according to the flexibility of the considered sequences: if a DNA is more flexible than a standard DNA, it experiments a higher decrease of the accessible states during nucleosome formation. As a consequence, ΔG_{el} raises. On the contrary, elastic free-energy difference is more negative for rigid DNAs. However, this corresponds to the assumption that nucleosomal DNA is completely frozen. We considered the rigidity parameter, $\langle f^{-1}\rangle$, as corresponding to the ratio $\langle T/T^*\rangle$, where T is the dinucleotide empirical melting temperature [26] and T^* the relative mean value assumed as a standard. This ratio, averaged along the kth nucleosomal DNA, modulates the force constant producing a sequence-dependent flexibility [24,25]. Anyway, we found that dynamic fluctuations are also correlated to the ratio between stacking energies [28] [and paper in progress], and that there is a correlation between

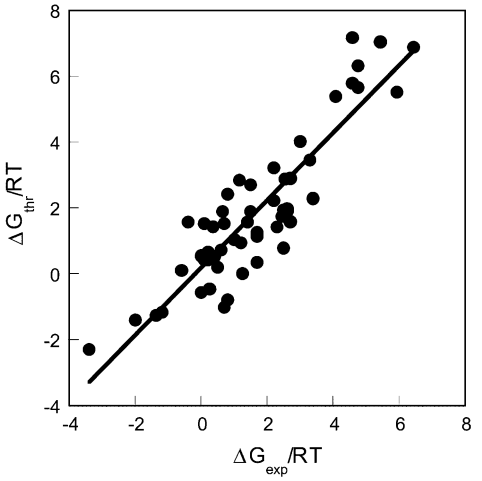

Fig. 11. Theoretical vs. experimental nucleosome reconstitution free energies of synthetic and natural DNA fragments [70–81]. The experimental data are related to the TG pentamer as a standard [70,71].

flexibility scales based on melting temperatures and on stacking energies (Table 2). This suggests the possibility of using the two scales, indifferently.

Therefore, we evaluated the elastic free energy difference, $\Delta G_{el}(k)$, of the competitive nucleosome reconstitution for a large set of DNAs, different for sequence and length investigated in several laboratories [70–81].

Surprisingly, the differences between the experimental and theoretical (elastic) nucleosome reconstitution free energies are not constant with respect to the square of the intrinsic effective curvature of the free DNA as expected (they are related to the choice of the standard DNA, Crothers TG [70,71], in the case of experimental data and an ideal straight DNA in our model), but appear to be a parabolic function of the DNA curvature [24,25]. This would suggest the existence of a free energy contribution directly related to the intrinsic curvature, namely to a property of the free DNA, which destabilizes the nucleosome formation.

The strikingly significant correlation with the average DNA effective curvature strongly supports the hypothesis that DNA curvature could play two opposite roles in nucleosome stability: one favoring the nucleosome formation by reducing the elastic free energy required to distort a given DNA

tract in the nucleosomal structure; the other, related to the curvature of the DNA free form, reducing the DNA affinity for histonic octamer. The latter curvature effect can be plausibly related to the small-groove narrowing and to the sequence-dependent DNA hydration.

This problem has been investigated by several authors in relation to its role in protein binding and specificity [82–89]. All the investigations provide evidence of differential hydration and stability of DNA sequences related to the minor groove width in agreement with the Drew and Dickerson model of the 'spine of water' [90].

Therefore it is reasonable to consider a further energy cost to the nucleosome formation. This corresponds to the free energy that DNA spends to release a part of the 'spine of water' (and counterions) displaced by the histone interactions.

This additive contribution is characterized by an inverse quadratic dependence on minor groove contraction from the straight DNA, to represent the water dipole energy difference in the electrostatic field of phosphates. Consequently it is possible to evaluate it as a function of the average DNA effective curvature [25].

Fig. 11 shows the comparison between experimental and theoretical nucleosome reconstitution free energies of synthetic and natural DNA fragments of different length, intrinsic curvature and flexibility.

The accord is straightforward proving the general validity of the model.

On the basis of the reported results, we can advance the conclusion that DNA intrinsic curvature is the main factor which controls nucleosome stability, and as a consequence nucleosome positioning. DNA curvature plays a dual role: by decreasing the distortion free energy of the DNA tract, when it assumes the nucleosomal shape, and by increasing the energy cost which the corresponding DNA free form spends to release a part of the 'spine of water' (and counterions) which is displaced by histone interactions. Other factors concerning chemical recognition arising from specific interactions between base pairs and aminoacidic residues, play probably a minor role, which requires further investigations.

Also the flexibility (bendability and twistability) appears to play a dual role: by decreasing the distortion energy necessary for nucleosome formation, but in contrast by increasing the entropy difference between the flexible free form and the final rigid nucleosomal structure.

Furthermore, it is important to note that the DNAs investigated have different length, which obviously increases the number of virtual positions and therefore the histone affinity in competitive nucleosome reconstitution.

Finally, it is interesting that a general feature of curved DNAs is the recurrence of repeated AA·TT sequences, in phase with the double helix periodicity, faced in the bending direction of the free form, as well as toward the histone core in the nucleosome. The presence of A·T base pair adds further stability to the 'spine of water' as first shown by x-ray crystallographic studies [90] and further validated by thermodynamic and NMR data from aqueous solutions [82–89]. A significant part of such water molecules (and counterions) is displaced by the histone minor groove binding, thus increasing the energy cost of nucleosome formation. Therefore AA·TT stretches appear to control the equilibrium, stabilizing and limiting the nucleosome stability.

5. Mapping sequence-dependent DNA curvature and flexibility from scanning force microscopy images

Electron microscopy (EM) and scanning force microscopy (SFM) are unique techniques, which provide information on both the contour of individual molecules and a population of such contours [91,92].

The data coming from these techniques were explained by using either static or dynamic models of the behavior of DNA in solution. The former models are focused on time-averaged conformations, i.e. on static intrinsic curvature, the latter on dynamic contributions, i.e. on the flexibility that characterizes the deformability about those averaged structures [14,15,93–95].

Attempts to characterize and separate the effects of static curvatures from those we can ascribe to the flexibility, were so far made only on peculiar

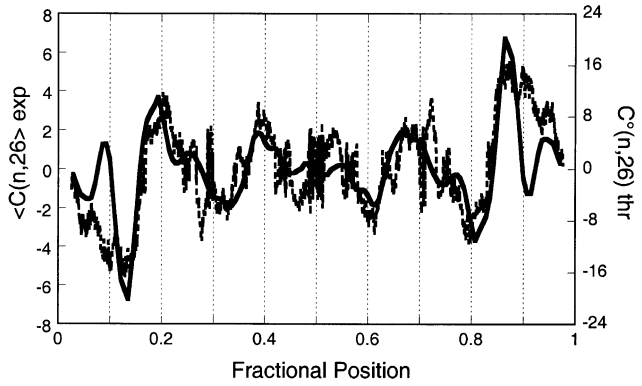

Fig. 12. Plot of the averaged experimental curvature (solid line) superimposed to the theoretical curvature (dashed line). The base pair resolution theoretical curvature has been smoothed to facilitate the comparison with the experimental data. All data refer to $m=26$ bp.

DNA constructs with anomalous flexibility [93,96–101]. Therefore the problem is still open for a 'natural' randomly sequenced DNA.

The possibility to estimate both intrinsic curvature and differential flexibility along the chain, as discussed before, can be considered a sound basis to approach this problem.

Recalling Eq. (1), the time average value of the curvature at the nth dinucleotide step is

$$\left\langle \mathbf{c}(n,t)\right\rangle = \left\langle \mathbf{c}^{\circ}(n)+\chi(n,t)\right\rangle = \mathbf{c}^{\circ}(n)+\left\langle \chi(n,t)\right\rangle \tag{53}$$

where the superscript $^{\circ}$ is used here to indicate the intrinsic curvature. If the fluctuations under thermal perturbations are considered to follow first-order elasticity due to the relatively high rigidity of DNA, their average value vanishes for all the sequence positions. As a consequence the observed average curvature profile must be equal to the intrinsic curvature. This conclusion is obviously true for both three- and two-dimensional structures.

However, SFM makes it possible to image the conformations of a DNA molecule deposited from a buffer solution on a mica substrate [102–104]. These single molecule observations can give a deeper insight on the way the base sequence determines the local intrinsic curvatures and drives the elastic responses to thermal energy, provided that the base-sequence of the imaged DNA can be assigned at any position along the chain in all images.

Therefore, it is possible to introduce a vector $\mathbf{c}(n,m)$, which represents the modulus and direction of the DNA curvature at the nth position, when it is sampled at intervals of m bp, where m depends on the resolution of the SFM images. However, it is possible to introduce the vector $\chi(n,m)$ to represent the DNA fluctuation at the nth position. Assuming isotropicity of the elastic tensor, we can write:

$$\left\langle \chi(n,m)^2\right\rangle = \left\langle \left|\mathbf{c}(n,m)-\left\langle \mathbf{c}(n,m)\right\rangle\right|^2\right\rangle = m\frac{RT}{2b^{*}\left\langle f^{-1}\right\rangle} \tag{54}$$

The results are analogous to that of Eq. (11), expressing the apparent bending force constant as in Eq. (13) and taking into account the sampling of the curvature every m base pair and for the two-dimensionality of the system. Eq. (54) represents the curvature dispersion at the nth sequence position, whereas its square root is the corresponding standard deviation of the distribution of the curvature around its mean (intrinsic) value in the plane.

In general, when measuring the curvature of a DNA molecule in a micrograph, we never know

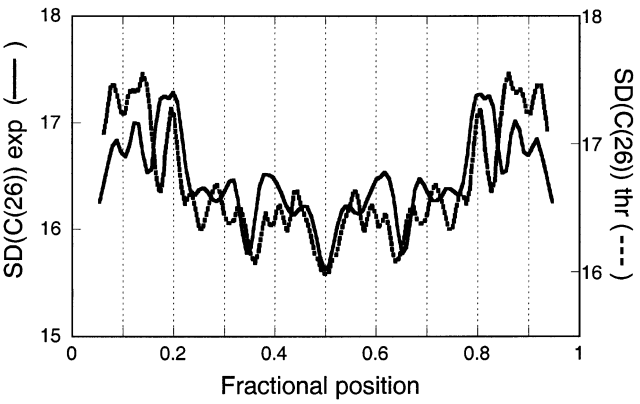

Fig. 13. Comparison between the experimental (solid line) and theoretical (dashed line) curvature standard deviation (S.D.). The theoretical S.D. was calculated modulating the average bending force constant, b^{*}, by multiplying it by the local $\langle f^{-1}\rangle$ value [see Eq. (54)], evaluated from the ratio T/T^{*} (Table 2) and averaged over $m=26$ bp.

which end we start with. In order to remove this uncertainty, Zuccheri et al. [27] studied a palindromic dimer of the DNA plasmid pBR322. In fact, in palindromic molecules, the sequence is the same reading from either end to the other and no uncertainty on the sequence orientation can exist. Furthermore, the curvature profile of a population of a palindromic DNA molecule should reflect the binary symmetry of the sequence, if the number of analyzed molecules is big enough, and the equilibration conditions are good enough.

Therefore, both the theoretical DNA average curvature and its standard deviation were compared with the corresponding experimental values calculated from SFM images of the pBR322 dimer [27].

Fig. 12 reports the theoretical intrinsic vs. the experimental average curvature. It is evident that not only the dyad symmetry, but also the location of the most significant curvature peaks is similar between the two profiles.

Fig. 13 shows the theoretical and experimental differential flexibility along the sequence, represented by the standard deviation of the curvature distribution. The correspondence between the profiles of this plot is an additional evidence of the validity of the flexibility scale we adopted (Table 2).

6. Conclusions

The implication of DNA curvature and flexibility 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.

In this sense, a large number of different approaches have been proposed in the literature to model DNA behavior.

It could be stated that each approach has its own advantages and disadvantages which justify the development of the method and constitute its limitations.

However, the increasing amount of the experimental data in the literature continuously requires the development of the old models and/or proposal of new points of view.

Given the overwhelming complexity of the systems, it is interesting to compare the results from different researches. In fact, the development of proper interpreting models is necessary to understand the main features of DNA behavior, avoiding the false conclusions that could driven through the 'common sense'.

The models we developed are all based on a sound evaluation of the sequence-dependent DNA curvature. We have shown that curvature can be related to thermodynamic functions such as Gibbs free energy, internal energy and entropy, if it is possible to express the variations of the DNA superstructures in Fourier space.

This method has demonstrated great flexibility and capacity of predicting equilibrium constants of many biological processes in which DNA is involved.

Starting from this sound basis, we have proposed a new point of view in the interpretation of the different experimental data concerning sequence-dependent DNA flexibility. We have advanced a set of dinucleotide parameters that modulated the ground-state DNA distortion energy as well as fluctuations along the DNA chain. Also in this case, the same approach seems to be able to justify the experimental data from different techniques, available in the literature.

It casts light that DNA curvature and, in same cases, flexibility are the main determinant of DNA macroscopic behavior, whereas all the other effects, such as interactions with the solution and the counterions, seem to play a secondary role.

This advanced the hypothesis that in molecular systems with such a large degree of complexity, the average molecular properties dominate over the local features, as in a statistical ensemble.

Acknowledgments

This work was supported by 'Progetto 60% Facoltà' of University 'La Sapienza' of Rome, by 'Cofinanziamento 40%' of MURST and by Istituto Pasteur, Fondazione Cenci-Bolognetti.Computer programs derived by some of the proposed models has been made available on line at the address http://www.chem.uniroma1.it/~desantis.

References

[1] L.D. Landau, E.M. Lifshitz, Theory of Elasticity, Pergamon Press, Oxford and New York, 1970.

[2] P. De Santis, A. Palleschi, S. Morosetti, M. Savino, in: E. Clementi, S. Chin (Eds.), Structures and dynamics of nucleic acids, proteins and membranes, structure and superstructures in periodical polynucleotides, Plenum Press, New York, 1986, 31.

[3] P. De Santis, A. Palleschi, M. Savino, A. Scipioni, A theoretical model of DNA curvature, Biophys. Chem. 32 (1988) 305–317.

[4] J.C. Marini, S.D. Levene, D.M. Crothers, P.T. Englund, Bent helical structure in kinetoplast DNA, Proc. Natl. Acad. Sci. USA 79 (1982) 7664–7668.

[5] H.-M. Wu, D.H. Crothers, The locus of sequence-directed and protein-induced DNA bending, Nature 308 (1984) 509–513.

[6] H.S. Koo, H.M. Wu, D.M. Crothers, DNA bending at adenine-thymine tracts, Nature 320 (1986) 501–506.

[7] P.J. Hagerman, Sequence dependence of the curvature of DNA: a test of the phasing hypothesis, Biochemistry 24 (1985) 7033–7036.

[8] P.J. Hagerman, Sequence-directed curvature of DNA, Nature 321 (1986) 449–450.

[9] H.S. Koo, D.M. Crothers, Calibration of DNA curvature and a unified description of sequence-directed bending, Proc. Natl. Acad. Sci. USA 85 (1988) 1763–1767.

[10] E.N. Trifonov, Sequence-dependent deformational anisotropy of chromatin DNA, Nucleic Acids Res. 8 (1980) 4041–4053.

[11] A. Bolshoy, P. McNamara, R.E. Harrington, E.N. Trifonov, Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles, Proc. Natl. Acad. Sci. USA 88 (1991) 2312–2316.

[12] P.G. De Gennes, Reptation of a polymer chain in the presence of fixed obstacles, J. Chem. Phys. 55 (1971) 572–579.

[13] O.J. Lumpkin, P. Dejardin, B.H. Zimm, Theory of gel electrophoresis of DNA, Biopolymers 24 (1985) 1573–1593.

[14] E.N. Trifonov, R.K.-Z. Tan, S.C. Harvey, in: W.K. Olson, M.H. Sarma, M. Sundaralingam (Eds.), Structure & Expression, DNA Bending and Curvature, vol. 3, Adenine Press, 1987, p. 243.

[15] J.A. Schellman, S.C. Harvey, Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature, Biophys. Chem. 55 (1995) 95–114.

[16] S.D. Levene, B.H. Zimm, Understanding the anomalous electrophoresis of bent DNA molecules: a reptation model, Science 245 (1989) 396–399.

[17] W.K. Olson, V.B. Zhurkin, Biological Structure and Dynamics, Proceedings of the Ninth Conversation, Twenty Years of DNA Bending, Adenine Press, 1996, p. 341.

[18] S. Diekmann, DNA methylation can enhance or induce DNA curvature, EMBO J. 6 (1987) 4213–4217.

[19] D.L. Milton, M.L. Casper, R.F. Gesteland, Saturation mutagenesis of a DNA region of bend. Base steps other than ApA influence the bend, J. Mol. Biol. 213 (1990) 135–140.

[20] P.D. Chastain, E.E. Eichler, S. Kang, D.L. Nelson, S.D. Levene, R.R. Sinden, Anomalous rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes, Biochemistry 34 (1995) 16125–16131.

[21] P.D. Chastain, R.R. Sinden, CTG repeats associated with human genetic disease are inherently flexible, J. Mol. Biol. 275 (1998) 405–411.

[22] W.K. Olson, A.A. Gorin, X.-J. Lu, L.M. Hock, V.B. Zhurkin, DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, Proc. Natl. Acad. Sci. USA 95 (1998) 11163–11168.

[23] W.K. Olson, N.L. Marky, R.L. Jernigan, V.B. Zhurkin, Influence of fluctuations on DNA curvature. A comparison of flexible and static wedge models of intrinsically bent DNA, J. Mol. Biol. 232 (1993) 530–554.

[24] C. Anselmi, G. Bocchinfuso, P. De Santis, M. Savino, A. Scipioni, Dual role of DNA intrinsic curvature and flexibility in determining nucleosome stability, J. Mol. Biol. 286 (1999) 1293–1301.

[25] C. Anselmi, G. Bocchinfuso, P. De Santis, M. Savino, A. Scipioni, A theoretical model for the prediction of sequence-dependent nucleosome thermodynamic stability, Biophys. J. 79 (2000) 601–613.

[26] O. Gotoh, Y. Tagashira, Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles, Biopolymers 20 (1981) 1033–1042.

[27] G. Zuccheri, A Scipioni, V. Cavaliere, G. Gargiulo, P. De Santis, B. Samorì, Mapping the intrinsic curvature and the flexibility along the DNA chain, Proc. Natl. Acad. Sci. USA 98 (2001) 3074–3079.

[28] R.L. Ornstein, R. Rein, D.L. Breen, R.D. Macelroy, An optimized potential function for the calculation of nucleic acid interaction energies. I. Base stacking, Biopolymers 17 (1978) 2341–2360.

[29] P.J. Hagerman, Flexibility of DNA, Annu. Rev. Biophys. Biophys. Chem. 17 (1988) 265–286.

[30] C.J. Benham, Elastic model of supercoiling, Proc. Natl. Acad. Sci. USA 74 (1977) 2397–2401.

[31] M. Le Bret, Catastrophic variation of twist and writhing of circular DNAs with constraint, Biopolymers 18 (1979) 1709–1725.

[32] M. Le Bret, Twist and writhing in short circular DNAs according to first-order elasticity, Biopolymers 23 (1984) 1835–1867.

[33] J. Shimada, H. Yamakawa, Statistical mechanics of DNA topoisomers. The helical worm-like chain, J. Mol. Biol. 184 (1985) 319–329.

[34] S.D. Levene, D.M. Crothers, Topological distributions and the torsional rigidity of DNA. A Monte Carlo study of DNA circles, J. Mol. Biol. 189 (1986) 73–83.

[35] W.K. Olson, Simulating DNA at low resolution, Curr. Opinion Struct. Biol. 6 (1996) 242–256.
[36] M.-H. Hao, W.K. Olson, Global equilibrium configurations of supercoiled DNA, Macromolecules 22 (1989) 3292–3303.
[37] T. Schlick, W.K. Olson, T. Westcott, J.P. Greenberg, On higher buckling transitions in supercoiled DNA, Biopolymers 34 (1994) 565–597.
[38] T.P Westcott, I. Tobias, W.K Olson, Elasticity theory and numerical analysis of DNA supercoiling: an application to DNA looping, J. Phys. Chem. 99 (1995) 17926–17935.
[39] W.R. Bauer, R.A. Lund, J.H. White, Twist and writhe of a DNA loop containing intrinsic bends, Proc. Natl. Acad. Sci. USA 90 (1993) 833–837.
[40] J.H. White, R.A. Lund, W.R. Bauer, Twist, writhe, and geometry of a DNA loop containing equally spaced coplanar bends, Biopolymers 38 (1996) 235–250.
[41] J.A. Gebe, J.M. Schurr, Thermodynamics of the first transition in writhe of a small circular DNA by Monte Carlo simulation, Biopolymers 38 (1996) 495–503.
[42] C. Anselmi, G. Bocchinfuso, P. De Santis, M. Fuà, A. Scipioni, M. Savino, Statistical thermodynamic approach for evaluating the writhe transformations in circular DNAs, J. Phys. Chem. B 102 (1998) 5704–5714.
[43] F.B. Fuller, The writhing number of a space curve, Proc. Natl. Acad. Sci. USA 68 (1971) 815–819.
[44] R. Bracewell, The Fourier Transform and its Applications, McGraw-Hill, New York, 1965.
[45] G. Calagareanu, L'intégral de Gauss et l'analyse des noeuds tridimensionnels, Rev. Math. Pur. Appl. 4 (1959) 5–20.
[46] M.R. Spiegel, Fourier Analysis, McGraw-Hill, New York, 1974.
[47] H. Jacobson, W.H. Stockmayer, Intramolecular reaction in polycondensations. I. The theory of linear systems, J. Chem. Phys. 18 (1950) 1600–1606.
[48] J.C. Wang, N. Davidson, Thermodynamic and kinetic studies on the interconversion between the linear and circular forms of phage λ DNA, J. Mol. Biol. 15 (1966) 111–123.
[49] J.C. Wang, N. Davidson, On the probability of ring closure of λ DNA, J. Mol. Biol. 19 (1966) 469–482.
[50] V. Bloomfield, D.H. Crothers, I. Tinoco, The Physical Chemistry of Nucleic Acids, Harper and Row, New York, 1974.
[51] O. Kratky, G. Porod, X-ray investigation of dissolved chain molecules, Recl. Trav. Chim. 68 (1949) 1106–1122.
[52] H. Yamakawa, W.H. Stockmayer, Statistical mechanics of wormlike chains. II. Excluded volume effects, J. Chem. Phys. 57 (1972) 2843–2854.
[53] P.J. Flory, U.W. Suter, M. Mutter, Macrocyclization equilibriums. 1. Theory, J. Am. Chem. Soc. 98 (1976) 5733–5739.
[54] J.E. Hearst, W.H. Stockmayer, Sedimentation constants of broken chains and wormlike coils, J. Chem. Phys. 37 (1962) 1425–1433.
[55] J. Shimada, H. Yamakawa, Ring closure probability for twisted wormlike chains. Application to DNA, Macromolecules 17 (1984) 689–698.
[56] S.D. Levene, D.M. Crothers, Ring closure probabilities for DNA fragments by Monte Carlo simulation, J. Mol. Biol. 189 (1986) 61–72.
[57] D. Sprous, R.K.-Z. Tan, S.C. Harvey, Molecular modeling of closed circular DNA thermodynamic ensembles, Biopolymers 39 (1996) 243–258.
[58] J.D. Kahn, D.M. Crothers, Measurement of the DNA bend angle induced by the catabolite activator protein using Monte Carlo simulation of cyclization kinetics, J. Mol. Biol. 276 (1998) 287–309.
[59] S.C. Hockings, J.D. Kahn, D.M. Crothers, Characterization of the ATF/CREB site and its complex with GCN_4, Proc. Natl. Acad. Sci. USA 95 (1998) 1410–1415.
[60] M. Roychoudhury, A. Sitlani, J. Lapham, D.M. Crothers, Global structure and mechanical properties of a 10-bp nucleosome positioning motif, Proc. Natl. Acad. Sci. USA 97 (2000) 13608–13613.
[61] D. Shore, J. Langowski, R.L. Baldwin, DNA flexibility studied by covalent closure of short fragments into circles, Proc. Natl. Acad. Sci. USA 78 (1981) 4833–4837.
[62] D. Shore, R.L. Baldwin, Energetics of DNA twisting. I. Relation between twist and cyclization probability, J. Mol. Biol. 170 (1983) 957–981.
[63] J.D. Kahn, D.M. Crothers, Protein-induced bending and DNA cyclization, Proc. Natl. Acad. Sci. USA 89 (1992) 6343–6347.
[64] A. Sitlani, D.M. Crothers, Fos and Jun do not bend the AP-1 recognition site, Proc. Natl. Acad. Sci. USA 93 (1996) 3248–3252.
[65] S.C. Schultz, G.C. Shield, T.A. Steitz, Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees, Science 253 (1991) 1001–1007.
[66] R.D. Kornberg, Structure of chromatin, Annu. Rev. Biochem. 46 (1977) 931–954.
[67] A. Klug, D. Rhodes, J. Smith, T.J Finch, J.O. Thomas, A low resolution structure for the histone core of the nucleosome, Nature 287 (1980) 509–516.
[68] T.J Richmond, J.T. Finch, B. Rushton, D. Rhodes, A. Klug, Structure of the nucleosome core particle at 7 Å resolution, Nature 311 (1984) 532–537.
[69] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution, Nature 389 (1997) 251–260.
[70] T.E. Shrader, D.M. Crothers, Artificial nucleosome positioning sequences, Proc. Natl. Acad. Sci. USA 86 (1989) 7418–7422.
[71] T.E. Shrader, D.M Crothers, Effects of DNA sequences and histone-histone interactions on nucleosome placement, J. Mol. Biol. 216 (1990) 69–84.

[72] J.S. Godde, A.P Wolffe, Nucleosome assembly on CTG triplet repeats, J. Biol. Chem. 271 (1996) 15222–15229.
[73] J.S. Godde, S.U. Kass, M.C. Hirst, A.P Wolffe, Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter, J. Biol. Chem. 271 (1996) 24325–24328.
[74] Y.R. Wang, R. Gellibolian, M. Shimizu, R.D. Wells, J. Griffith, Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes, J. Mol. Biol. 263 (1996) 511–516.
[75] Y.R. Wang, J. Griffith, The $[(G/C)_3NN]_n$ motif: a common DNA repeat that excludes nucleosomes, Proc. Natl. Acad. Sci. USA 93 (1996) 8863–8867.
[76] H.R. Widlund, H. Cao, S. Simonsson, , et al., Identification and characterization of genomic nucleosome-positioning sequences, J. Mol. Biol. 267 (1997) 807–817.
[77] S Cacchione, M.A. Cerone, M. Savino, In vitro low propensity to form nucleosomes of four telomeric sequences, FEBS Lett. 400 (1997) 37–41.
[78] P.T. Lowary, J. Widom, New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning, J. Mol. Biol. 276 (1998) 19–42.
[79] L. Rossetti, S. Cacchione, M. Fuà, M. Savino, Nucleosome assembly on telomeric sequences, Biochemistry 37 (1998) 6727–6737.
[80] M. Del Cornò, P. De Santis, B. Sampaolese, M. Savino, DNA superstructural features and nucleosomal organization on the two centromers of *Kluveromyces lactis* chromosome 1 and *Saccharomyces cerevisiae* chromosome 6, FEBS Lett. 431 (1998) 66–70.
[81] H. Cao, H.R. Widlund, T. Simonsson, M. Kubista, TGGA repeats impair nucleosome formation, J. Mol. Biol. 281 (1998) 253–260.
[82] F. Fack, V. Sarantoglou, Curved DNA fragments display retarded elution upon anion exchange HPLC, Nucleic Acid Res. 19 (1991) 4181–4188.
[83] E. Liepnish, G. Otting, K. Wuthrich, NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution, Nucleic Acid Res. 20 (1992) 4549–4553.
[84] H.M. Berman, Hydration of DNA: take 2, Curr. Opin. Struct. Biol. 4 (1994) 345–350.
[85] X. Shui, L. McFail-Isom, D. VanDerveer, L.D. William, The B-DNA dodecamer at high resolution reveals a spine of water on sodium, Biochemistry 37 (1998) 8341–8355.
[86] X. Shui, C.C. Sines, L. McFail-Isom, D. VanDerveer, L.D. William, Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations, Biochemistry 37 (1998) 16877–16887.
[87] J. Woda, B. Schneider, K. Patel, K. Mistry, H.M. Berman, An analysis of the relationship between hydration and protein-DNA interactions, Biophys. J. 75 (1998) 2170–2177.
[88] L. McFail-Isom, C.C. Sines, L.D. William, DNA structure: cations in charge?, Curr. Opin. Struct Biol. 9 (1999) 298–308.
[89] N.V. Hud, V. Sklenar, J. Feigon, Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending, J. Mol. Biol. 286 (1999) 651–660.
[90] H.R. Drew, R.E. Dickerson, Structure of a B-DNA dodecamer. III. Geometry of hydration, J. Mol. Biol. 151 (1981) 535–556.
[91] G. Muzard, B. Theveny, B. Revet, Electron microscopy mapping of pBR322 DNA curvature. Comparison with theoretical models, EMBO J. 9 (1990) 1289–1298.
[92] J.A. Cognet, C. Pakleza, D. Cherny, E. Delain, E.L. Cam, DNA bending induced by the archaebacterial histone-like protein MC1, J. Mol. Biol. 285 (1999) 997–1009.
[93] J. Bednar, P. Furrer, V. Katritch, A.Z. Stasiak, J. Dubochet, A. Stasiak, Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA, J. Mol. Biol. 254 (1995) 579–594.
[94] D. Porschke, E.R. Schmidt, T. Hankeln, G. Nolte, J. Antosiewicz, Structure and dynamics of curved DNA fragments in solution: evidence for slow modes of configurational transitions, Biophys. Chem. 47 (1993) 179–191.
[95] C.R. Calladine, H.R. Drew, A useful role for 'static' models in elucidating the behaviour of DNA in solution, J. Mol. Biol. 257 (1996) 479–485.
[96] C. Rivetti, C. Walker, C. Bustamante, Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility, J. Mol. Biol. 280 (1998) 41–59.
[97] A. Grove, A. Galeone, L. Mayol, E.P. Geiduschek, Localized DNA flexibility contributes to target site selection by DNA-bending proteins, J. Mol. Biol. 260 (1996) 120–125.
[98] J.D. Kahn, E. Yun, D.M. Crothers, Detection of localized DNA flexibility, Nature 368 (1994) 163–166.
[99] J.K. Strauss, L.J. Maher, DNA bending by asymmetric phosphate neutralization, Science 266 (1994) 1829–1834.
[100] E. Le Cam, F. Fack, J. Menissier-de Murcia, , et al., Conformational analysis of a 139 base-pair DNA fragment containing a single-stranded break and its interaction with human poly(ADP-ribose) polymerase, J. Mol. Biol. 235 (1994) 1062–1071.
[101] T. Akiyama, M.E. Hogan, Structural analysis of DNA bending induced by tethered triple helix forming oligonucleotides, Biochemistry 36 (1997) 2307–2315.
[102] C. Rivetti, M. Guthold, C. Bustamante, Scanning force

microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis, J. Mol. Biol. 264 (1996) 919–932.
[103] C. Bustamante, C. Rivetti, Visualizing protein-nucleic acid interactions on a large scale with the scanning force microscope, Annu. Rev. Biophys. Biomol. Struct. 25 (1996) 395–429.
[104] C. Bustamante, D.A. Erie, D. Keller, Biochemical and structural applications of scanning force microscopy, Curr. Opin. Struct. Biol. 4 (1994) 750–760.