ELSEVIER

Biophysical Chemistry 112 (2004) 77–88

Biophysical Chemistry

www.elsevier.com/locate/bpc

# Sequence-dependent DNA torsional rigidity: a tetranucleotide code

Francesco Pedone[a,b,*], Filomena Mazzei[c], Daniele Santoni[d,e]

[a]*Istituto Nazionale di Fisica della Materia, Università La Sapienza, Rome, Italy*
[b]*Dipartimento di Genetica e Biologia Molecolare, Università La Sapienza, Rome, Italy*
[c]*Dipartimento di Ambiente e connessa prevenzione primaria, Istituto Superiore di Sanità, Rome, Italy*
[d]*Consorzio per le Applicazioni di Supercalcolo per Università e Ricerca, Rome, Italy*
[e]*Istituto Universitario di Scienze Motorie, Rome, Italy*

Received 20 April 2004; received in revised form 6 July 2004; accepted 6 July 2004
Available online 28 August 2004

**Abstract**

Using fluorescence polarization anisotropy (FPA), we measured the torsional constant of various DNA oligomers in different sequences and calculated the value for each of the 136 unique tetranucleotides. From these values, we obtained a «rigidity profile» for every double-stranded DNA sequence. We tested the code in the analysis of DNA sequences able to form nucleosomes. More than 50% of the sequences studied showed a common 20 and/or 30 bp modulation of the torsional constant. Many other profiles of rigidity were observed in the remaining sequences and this variety in torsional constant modulation may be related to functional differences between nucleosomes.


*Keywords:* Fluorescence polarization anisotropy; Tetranucleotides; Nucleosomes; DNA flexibility

## 1. Introduction

The dependence of DNA flexibility on base composition was first investigated by triplet anisotropy decay measurements of methylene blue intercalated in synthetic polydeoxynucleotides by Hogan et al. [1]. These authors observed that the rigidity of a base step may vary in different sequence contexts and reported that poly(dG) · poly(dC) has a torsional constant 40 times larger than poly(dA–dC) poly(dT–dG) and 20 times larger than poly(dA) · poly(dT).

Further studies on this topic, carried out with fluorescence depolarization measurements of ethidium bromide complexed with DNA in various forms [2–4], did not confirm the previous results and small sequence effects on DNA flexibility were reported. Thereafter, long DNA fragments (>1 kbase) [5] and short oligonucleotides were studied, with respect to sequence effects [6–8], but with no systematic approach. We recently studied several short synthetic oligonucleotides using fluorescence polarization anisotropy (FPA) [9] and found that DNA rigidity is sequence dependent.

The simplest model for interpreting FPA experimental data obtained from DNA is based on the properties of the dinucleotide steps which constitute the smallest significant unit as far as torsional motion is concerned. Differences in torsional constant values at the level of dinucleotide base steps can also be calculated by the variance of twist values derived from the X-ray crystal structure database [10]. However, Yanagi et al. [11] observed that the twist, like other DNA structural parameters, varies for the same base step depending both on crystal packing forces and flanking sequences, and postulated that an accurate analysis of DNA geometry must be based not only on 10 possible dinucleotide steps but on 136 tetranucleotide ones. At that time, only 33 of the 136 unique four-base steps were represented in the X-ray solved structures and no attempt was made to develop a tetranucleotide code.

Recently, Packer et al. [12] obtained 136 theoretical tetranucleotide conformational maps useful to predict

* Corresponding author. Dipartimento di Genetica e Biologia Molecolare, Università La Sapienza, Piazzale A. Moro 3, Rome 00161, Italy. Tel.: +39 649912235; fax: +39 6449970.
*E-mail address:* francesco.pedone@uniroma1.it (F. Pedone).

0301-4622/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2004.07.005

structures not yet crystallographically characterized. This data was extracted from 60 crystal structures of DNA oligomers [13] containing 66 four-base steps of the 136 unique ones.

We extended one of our previous studies [9] to derive the value of the torsional constant for each of the 136 possible tetranucleotides. 53 DNA oligomers of length ranging from 24 to 30 bp were analyzed; they contained 1142 four-base steps, in which all the 136 unique tetranucleotides were represented.

In order to find a possible correlation between the rigidity of the DNA tract calculated using our tetranucleotide code and some general DNA feature, we analyzed sequences known to form nucleosomes. The complexation of 146 bp of DNA around the histone octamer is a well-known and characterized phenomena [14–16] and, in principle, is not sequence dependent in that the interaction is between the negative DNA phosphate groups and the positive charges of the histones' basic amino acids. There are 12 interaction sites on the histone octamer surface [16], spaced one-helix-turn apart and the precise wrapping of the DNA must be linked to some elastic properties of this polymer.

We observed that DNA fragments lacking sequence homology but able to form nucleosomes yield similar rigidity profiles. These profiles may be divided into different classes, represented more or less in the analyzed sequences, with a characteristic modulation of rigidity around a mean value. This finding suggests that our tetranucleotide code—although a few values are uncertain—is correct and may be useful for further studies of DNA properties.

## 2. Materials and methods

### 2.1. Oligonucleotide preparation

Oligonucleotides were purchased from Roche and checked for length homogeneity by denaturing polyacrylamide gel electrophoresis (PAGE) [17]. The fraction of short contaminating fragments was usually less than 3% so that the samples were generally used without further purification. In order to obtain duplexes, equimolar amounts of the complementary single-strand oligonucleotides were heated in 10 mM Tris–HCl pH 7.5, 50 mM NaCl, 0.1 mM $Na_2EDTA$ buffer, to 90 °C for 10 min and then slowly cooled to 4 °C over a 16- to 18-h period. The annealing was checked by nondenaturing PAGE [17]. Calf thymus DNA was purchased from Sigma and was used without further purification.

Table 1 shows the top strand sequence of the duplexes.

Table 1
Name, sequence and torsional constant of the samples analyzed

| No. | DNA samples | $\alpha 10^{-19} \pm$ 0.2 (J) |
|---|---|---|
| — | Calf thymus DNA | 4.6 |
| 1 | GGCCGTCCGTCATCGTCCGTCCGG | 1.8 |
| 2 | GGCAAAAAAAAAAAAAAAAAGG | 2.3 |
| 3 | GGCAATTAATTATAATTAATTCGG | 2.3 |
| 4 | GGCTCAATCAAATTCAATCAACGG | 2.7 |
| 5 | GGCGACCGACCATGACCGACCCGG | 2.8 |
| 6 | GGCACAGACAGATACAGACAGCGG | 3.1 |
| 7 | GGCTCACTCACATTCACTCACCGG | 3.2 |
| 8 | GGCCACCCACCATCACCCACCCGG | 3.5 |
| 9 | GGCAACCAACCATAACCAACCCGG | 3.5 |
| 10 | GGCCATGCATGATCATGCATGCGG | 3.5 |
| 11 | GGCCCACCCACATCCACCCACCGG | 3.5 |
| 12 | GGCAATGAATGATAATGAATGCGG | 3.5 |
| 13 | GGCGCAAGCAAATGCAAGCAACGG | 3.6 |
| 14 | GGCGATCGATCATGATCGATCCGG | 3.6 |
| 15 | GGCATATATATATATATATATACGG | 3.6 |
| 16 | GGCCGCGCGCGCGCGCGCGCGCCGG | 3.6 |
| 17 | GGCTCAGTCAGATTCAGTCAGCGG | 3.9 |
| 18 | GGCCCATCCATATCCATCCATCGG | 4.0 |
| 19 | GGCGATAGATAATGATAGATACGG | 4.0 |
| 20 | GAGTTAGATGACGTCACCTCCAGG | 4.0 |
| 21 | GAAGAATAAGAATAACAAGCAGAAGAG | 4.0 |
| 22 | GGCTCGATCGAATTCGATCGACGG | 4.0 |
| 23 | GGCACGAACGAATACGAACGACGG | 4.0 |
| 24 | GGCTAGTTAGTATTAGTTAGTCGG | 4.1 |
| 25 | ACGTGCATGCACATACATGCGTACATG | 4.1 |
| 26 | GGCCGGACGGAATGGACGGACGG | 4.1 |
| 27 | GGCAGGCAGGCATAGGCAGGCCGG | 4.1 |
| 28 | GGCGACAGACAATGACAGACACGG | 4.2 |
| 29 | ACGTGCGTGCACATACATGCGTACATG | 4.2 |
| 30 | GGCCACTCACTATCACTCACTCGG | 4.3 |
| 31 | GGCTGGATGGAATTGGATGGACGG | 4.3 |
| 32 | TTCCTTTTACGTCATCCGGGGGCA | 4.3 |
| 33 | GTCCCTGTTCGGGCGCCACTGCTA | 4.4 |
| 34 | GGCTGCATGCAATTGCATGCACGG | 4.4 |
| 35 | GGCAGGAAGGAATAGGAAGGACGG | 4.4 |
| 36 | GGCTGGTTGGTATTGGTTGGTCGG | 4.4 |
| 37 | GGCGTACGTACATGTACGTACCGG | 4.5 |
| 38 | GGCACAACAACATACAACAACCGG | 4.6 |
| 39 | GGCAACTAACTATAACTAACTCGG | 4.6 |
| 40 | AGGCTCGGATCCAAATCTAGACTCAGA | 4.6 |
| 41 | GGCAGCAAGCAATAGCAAGCACGG | 4.6 |
| 42 | GGCCTAGCTAGATCTAGCTAGCGG | 4.7 |
| 43 | GGCACGTACGTATACGTACGTCGG | 4.8 |
| 44 | GGCAGGTAGGTATAGGTAGGTCGG | 4.8 |
| 45 | GGCAGCTAGCTAGCTAGCTAGCGG | 5.0 |
| 46 | AGGCTCAGATCCAGATCTAGACTCAGA | 5.0 |
| 47 | GATCCTCTAGAGTCGACCTGCAGGCATGCA | 5.0 |
| 48 | GGCAGTAAGTAAGTAAGTAAGCGG | 5.2 |
| 49 | GAAGAAGAGAAGGAAGGGAGAGGAAGA | 5.3 |
| 50 | AAAAGAAAAGGGGGGACTCGAAGGGCTAAT | 5.4 |
| 51 | GGCACACACACATACACACACCGG | 7.0 |
| 52 | CGAACAAATCTCTAGCAGTGGCGC | 7.5 |
| 53 | GGCAGAGAGATAGAGAGAGCGG | 7.9 |

### 2.2. Dynamic fluorescence measurements

Lifetime and fluorescence polarization anisotropy measurements were performed by a K2-ISS phase fluorometer (ISS, Urbana IL, USA), monitoring the fluorescence emission of the ethidium bromide (EB) dye bound to the DNA duplexes [18]. Fluorophore excitation was allowed using the 514-nm output of a Coherent Innova 90 °C Argon laser. The modulation ratio of the excitation light was always in the 60–70% range, and the detection cross-correlation frequency was 80 Hz.

For lifetime measurements, 10 frequencies logarithmically spaced in the 2–40 MHz interval were acquired. The excitation polarizer was set at 36°, with respect to vertical polarization of the laser, and polystyrene latex was used as a reference. The fluorescence emission was analysed through a long pass filter (550 nm). The data was analysed with K2-ISS fluorometer software, using the lifetime value of free ethidium equal to 1.7 ns. Lifetimes were measured as a function of temperature in the 15–30 °C range, and data was fitted by linear regression.

Anisotropy data was acquired at 20 frequencies in the 2- to 40-MHz interval, with the excitation polarizer kept at 0° (i.e., vertical) and the emission polarizer automatically rotated to 0° and 90° for each acquisition.

In the frequency domain, the quantities usually measured to characterise the fluorescence polarization anisotropy decay are the difference between the phase angles $\Delta\Phi$ of the parallel and perpendicular components with respect to the excitation beam and the demodulation ($M$) ratio. The error for phase and modulation ratio was 0.2° and 0.004, respectively.

$$\Delta\Phi = \Phi_{\perp} - \Phi_{||}$$

$$\Delta\Lambda = M_{||}/M_{\perp}$$

These quantities are related to the time decay of the fluorescence intensity ($I$), as measured when the relative orientation of the excitation polarizer and the emission analyser are parallel ($I_{||}$) or orthogonal ($I_{\perp}$), by means of Laplace transforms [8,19].

The model adopted to analyze FPA data using DNA was first proposed by Allison and Schurr [19]. The DNA helix is represented as a bendable and twistable set of $N$+1 stacked cylinders which represent the bp. The decay of fluorescence anisotropy is well described by the product of correlation functions, which refer to the internal motion of the nucleic acids themselves:

$$r(t) = \frac{I_{||} - I_{\perp}}{I_{I_{||}} + 2I_{\perp}} = r_0 \sum_{n=-2}^{2} A_n(t) C_n(t) F_n(t)$$

where $r_0$ is the EB limiting anisotropy value, usually equal to 0.36. $A_n$ are the internal correlation functions fully described by trigonometric functions of the angle $\theta$ formed by the EB dipole and the helical axis, assumed to be equal to 70.5°. $C_n(t)$ and $F_n(t)$ are the torsional and the bending correlation functions, respectively. All the correlation functions are fully described elsewhere [18,19]. Due to their different lengths, two different expressions of the correlation function $C_n(t)$ were used to analyze the data derived from calf thymus DNA and from the synthetic oligonucleotides. The two complete expressions for $C_n(t)$ can be found in Refs. [6,19], respectively.

Phase differences and demodulation ratios were measured as a function of the frequency at different temperatures. Data derived from the FPA measurements were fitted with a Vax station in order to obtain the hydrodynamic parameters and the torsional constant. During the fitting procedure, the torsional constant $\alpha$, the hydrodynamic radius $R_h$ and the rise $b$ were the free parameters. The temperature dependence of the lifetimes was also accounted for. The fluctuation of the twist angle ($\Delta\xi$) was evaluated according to Eq. (1).

$$\Delta\xi = \left(K_B T/\alpha\right)^{1/2} \quad (1)$$

where $K_B$ is the Boltzmann constant.

### 2.3. Mathematical treatment

The torsional rigidity data were analyzed with a genetic algorithm implemented in perl language.

The algorithm manipulates the bit strings in which the problem is encoded in order to find the best values of $\alpha$ for the 136 unique tetranucleotides, so that the rigidity estimated for each sequence using these values ($\alpha_{cal}$) is very close to the experimental one ($\alpha_{exp}$).

We started with a population of 10 chromosomes (arrays of 136 components) and for each computational step generated a new generation of 10 chromosomes by means of crossing over and single mutations. We then evaluated the fitness of the chromosomes and chose the best ones. The initial value of each tetranucleotide was assumed equal to the corresponding central dinucleotide. We allowed tetranucleotide values such that $0<\alpha<40$.

The fitness function $F$ is the sum of the square difference between the calculated rigidity and the experimental one for each $j$-th sequence as reported in Eq. (2)

$$F = \sum_{j=1}^{53} \left[ \left( \alpha_{exp}^{j} - \alpha_{cal}^{j} \right)^2 \right] \quad (2)$$

We computed the $j$-th rigidity of a sequence from the tetranucleotide occurrences using Eq. (3):

$$\frac{1}{\alpha_{cal}^{j}} = \left[ \left( \sum_{i=1}^{136} n_i^j \frac{1}{\alpha_i} \right) + \frac{1}{\alpha_{init}^{j}} + \frac{1}{\alpha_{end}^{j}} \right] \frac{1}{L^j - 1} \quad (3)$$

where $i$=1–136 is the index of tetranucleotides, $j$=1–53 is the index of sample sequences, $L^j$ is the length of the $j$-th sequence, $\alpha_i$ is the value of the $i$-th tetranucleotide, $n^j{}_i$ is the occurrence of the $i$-th tetranucleotide in the $j$-th sequence, $\alpha_{init}^{j}$ is the value of the dinucleotide at 5′ end and $\alpha_{end}^{j}$ is the value of the dinucleotide at 3′ end.

The algorithm stops when, for at least one chromosome, the fitness function $F$ is less than the parameter set at the beginning as the target. In order to analyze a sample

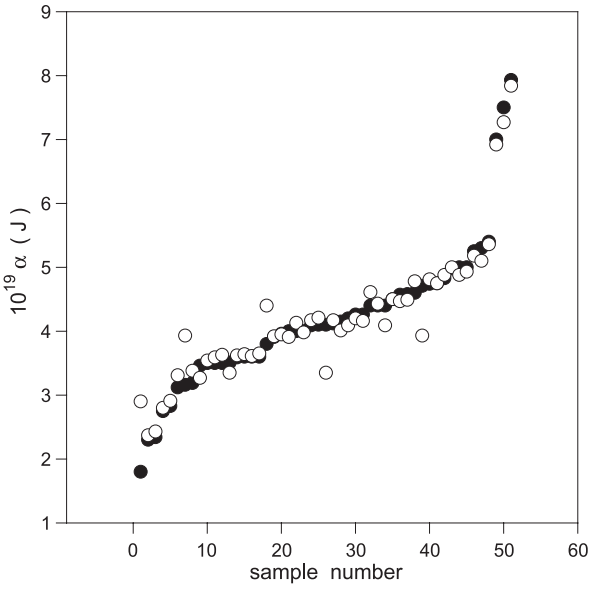


Fig. 1. Torsional constants of the samples at 20 °C. (●) Experimental values; (○) fitted values. Sample numbers are those of Table 1.

sequence, the values of the terminal dinucleotides must be assigned, and these values were arbitrarily assumed equal to AA/TT=2, CA/TG=2, AG/CT=2, TA=2, GA/TC=3, GG/CC=3, GC=3, CG=3, AC/GT=4, AT=4 and not varied in the fitting procedure. These values were lowered with respect to those measured in a previous study [9], considering that greater exposure to solvent, at the extremities of the DNA helix, would lower the rigidity.

The values of the 10 dinucleotide steps at the center of the 136 possible tetranucleotides were the free parameters in the fitting procedure, and their initial values were assumed to be AA/TT=2.3, CA/TG=2.1, AG/CT=3.3, TA=3.2, GA/TC=4.9, GG/CC=7.0, GC=6.0, CG=4.4, AC/GT=13.8, AT=9.7 according to the values reported in Ref. [9] calculated from the data of Suzuki and Yagi [20].

Fourier transform analysis of the data was conducted by the Matlab 4.2 FFT routine.

## 3. Results

### 3.1. Dynamic fluorescence measurements

The torsional constant α of DNA from calf thymus and from the 53 oligomers was evaluated from FPA data at four temperatures in the 10–30 °C range and fitted by exponential regression analysis [9]. The values at 20 °C are reported in ascending order in Table 1.

Samples 2,15, 16, 40, 46 and 49 had been studied previously [9] and were reexamined in the experimental conditions used in this paper. Samples 20, 21, 29, 32, 38, 47, 50, 52 and 53 were synthesized to represent random sequences; the remaining samples were constructed with:

(i) a common central AT step, in order to contain a preferential binding site for ethidium bromide [21]
(ii) the same 5′ and 3′ three bases GGC and CGG in order to ensure favourable annealing and reduced fraying ends effects
(iii) at least four repeats of the same tetranucleotide.

In Fig. 1, the α values at 20 °C of the 53 oligomers are reported, numbered as in Table 1. For 50 out the 53 samples analyzed, the measured α was in the range between 1.8 and 5.4 $10^{-19}$ J, whilst for only three samples was it in the range between 7 and 7.9.

The number of oligomers synthesized and analyzed was not very large so we cannot put forward any hypothesis on the distribution of rigidities shown in Fig. 1. We can only conclude that, to date, we have been able to sort out more sequences with moderate and low rigidity, and that very rigid sequences are poorly represented in our samples. It was mentioned previously that higher values of DNA

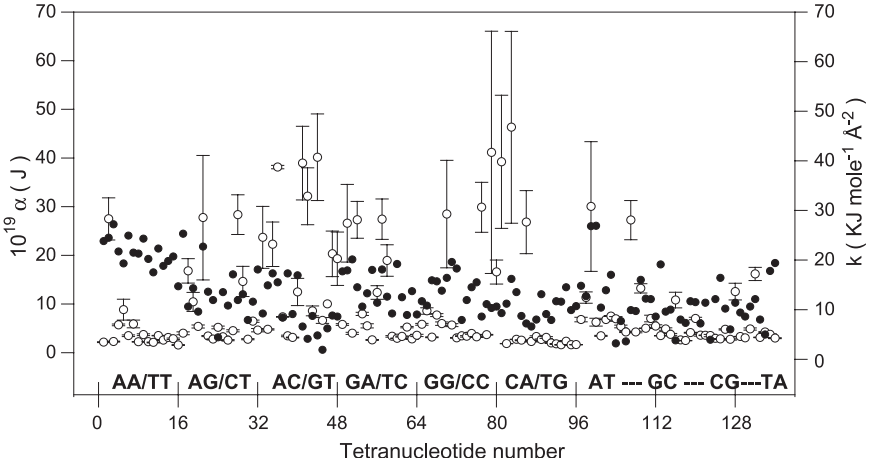


Fig. 2. Torsional constant (α) of the 136 unique tetranucleotides. (○) calculated values; (●) elastic constant (K) values calculated by Packer et al. [12]. The tetranucleotides are grouped according to the common central dinucleotide steps reported in the figure.

rigidity are found in sequences containing poly(dG) poly(dC) [1]. Our efforts to reproduce this result in short synthetic oligonucleotide (18 consecutive GC bp) failed due to the difficulty in annealing these samples properly. The presence of contaminating bands of apparently higher molecular weight in polyacrylamide gels indicated the formation of secondary structures, i.e., tetraplex forms [22].

Because the mean of the $\alpha$ values of the 53 samples at 20 °C (Table 1 and Fig. 1) was $4.2\pm0.2\times10^{-19}$ J, not far from the mean value $4.6\pm0.2\times10^{-19}$ J found for calf thymus DNA, we calculated the torsional constants of the 136 unique tetranucleotides from them.

Fig. 1 shows the values of rigidity calculated for each sample using Eq. (3). It is evident that the fit obtained is satisfactory, with the exception of samples No. 1 and 26. We cannot explain the behaviour of these samples, which is probably related to their particular sequence. However, we observed that these samples contain a triple repetition of the same TCCG/CGGA tetranucleotide with the central dinucleotide CC/GG. Hunter [23] reported evidence of bistability in DNA bp stacked in the DNA helix, mainly in the pyrimidine–purine steps CG and TA but also in CC/GG and AG/CT. The existence of bistable dinucleotide steps complicates the study of DNA dynamical geometry because of the increased number of conformations that a tetranucleotide can assume. This may occur when the central, bistable dinucleotide step is influenced by neighbouring nucleotides or when two or more consecutive bistable steps are adjacent. Further investigations with properly designed samples are needed to verify this hypothesis.

The best fit for the rigidity of the 53 oligomers shown in Fig. 1 was obtained, as described in Material and Methods section, by using 136 $\alpha$ values calculated for the 136 unique tetranucleotides. These values are reported in Fig. 2 and Table 2, where the occurrence ($n$) of each tetranucleotide in our database and the torsional constant values calculated for the 136 unique tetranucleotides by Packer et al. [24] are also reported. The standard deviation of the data reported in Fig. 2 is obtained by three best fits, and shows a good convergence for most of the data. In processing this data, we assumed a common $\alpha$ value equal to 2 or 5 or 8 for the 136 tetranucleotides and the results were always similar to those reported in Fig. 3. When we increased the number of data samples by measuring the rigidity of 10 successive oligomers and adding it to the database, the genetic algorithm become unstable and the convergence was worse.

It is worth noting, in Fig. 2, the influence of flanking bases on the rigidity of central dinucleotide steps which is more pronounced for groups bearing AC/GT, GA/TC and AG/CT as central steps. In the CA/TG group, there is also a marked increase in rigidity for 5 out of 16 tetranucleotides, but the remaining ones have very similar values; the other groups exhibit an intermediate behaviour.

In Fig. 2, we plotted the values of flexibility force constants for the 136 unique tetranucleotides calculated by

Table 2
Torsional constant values of the 136 unique tetranucleotide sequences

| No. | Tetranucleotide | $n$ | $\alpha10^{-19}$(J) | rmsd | [24] |
|---|---|---|---|---|---|
| 1 | AAAA/TTTT | 21 | 2.2 | – | 23.8 |
| 2 | AAAG/CTTT | 1 | 27.5 | 4.3 | 24.5 |
| 3 | AAAC/GTTT | 1 | 2.3 | – | 27.2 |
| 4 | AAAT/ATTT | 4 | 5.7 | 0.2 | 21.7 |
| 5 | GAAA/TTTC | 1 | 8.8 | 2.1 | 19.3 |
| 6 | GAAG/CTTC | 10 | 3.5 | 0.2 | 24.9 |
| 7 | GAAC/GTTC | 4 | 5.9 | 0.8 | 21.5 |
| 8 | GAAT/ATTC | 14 | 2.2 | – | 21.3 |
| 9 | CAAA/TTTG | 5 | 3.7 | 0.2 | 24.3 |
| 10 | CAAG/CTTG | 5 | 2.2 | – | 20.2 |
| 11 | CAAC/GTTG | 12 | 2.1 | 0.1 | 17.5 |
| 12 | CAAT/ATTG | 10 | 3.5 | 0.1 | 22.3 |
| 13 | TAAA/TTTA | 1 | 2.6 | 0.1 | 18.8 |
| 14 | TAAG/CTTA | 5 | 3.1 | – | 19.8 |
| 15 | TAAC/GTTA | 8 | 2.9 | 0.1 | 20.7 |
| 16 | TAAT/ATTA | 10 | 1.6 | – | 14.7 |
| 17 | AAGA/TCTT | 7 | 4.0 | 0.3 | 25.3 |
| 18 | AAGG/CCTT | 7 | 16.8 | 2.5 | 10.6 |
| 19 | AAGC/GCTT | 6 | 10.4 | 2.0 | 14.3 |
| 20 | AAGT/ACTT | 3 | 5.4 | 0.3 | 9.6 |
| 21 | GAGA/TCTC | 8 | 27.7 | 12.8 | 22.7 |
| 22 | GAGG/CCTC | 3 | 3.4 | 0.1 | 13.6 |
| 23 | GAGC/GCTC | 7 | 2.8 | – | 11.9 |
| 24 | GAGT/ACTC | 8 | 5.2 | 0.3 | 4.4 |
| 25 | CAGA/TCTG | 12 | 3.3 | 0.1 | 13.5 |
| 26 | CAGG/CCTG | 9 | 2.6 | 0.1 | 10.8 |
| 27 | CAGC/GCTG | 7 | 4.5 | 0.2 | 17.1 |
| 28 | CAGT/ACGT | 5 | 28.3 | 4.1 | 11.9 |
| 29 | TAGA/TCTA | 13 | 14.6 | 3.1 | 13.1 |
| 30 | TAGG/CCTA | 6 | 2.8 | 0.1 | 7.9 |
| 31 | TAGC/GCTA | 18 | 6.5 | 0.2 | 11.6 |
| 32 | TAGT/ACTA | 8 | 4.6 | 0.3 | 18.1 |
| 33 | AACA/TGTT | 6 | 23.7 | 6.4 | 9.2 |
| 34 | AACG/CGTT | 5 | 4.8 | 0.1 | 14.9 |
| 35 | AACC/GGTT | 7 | 22.3 | 4.6 | 17.3 |
| 36 | AACT/AGTT | 7 | 38.2 | – | 15.5 |
| 37 | GACA/TGTC | 6 | 7.3 | – | 8.4 |
| 38 | GACG/CGTC | 15 | 3.4 | 0.2 | 17.3 |
| 39 | GACC/GGTC | 7 | 3.1 | 0.1 | 9.1 |
| 40 | GACT/AGTC | 7 | 12.5 | 2.8 | 16.9 |
| 41 | CACA/TGTG | 12 | 39.0 | 7.6 | 6.6 |
| 42 | CACG/CGTG | 7 | 32.1 | 5.8 | 4.1 |
| 43 | CACC/GGTG | 10 | 8.6 | 1.0 | 8.7 |
| 44 | CACT/AGTG | 8 | 40.2 | 8.9 | 4.8 |
| 45 | TACA/TGTA | 9 | 6.6 | 0.6 | 1.9 |
| 46 | TACG/CGTA | 17 | 10.0 | – | 6.2 |
| 47 | TACC/GGTA | 5 | 20.3 | 4.7 | 8.8 |
| 48 | TACT/AGTA | 5 | 19.3 | 5.5 | 8.6 |
| 49 | AGAA/TTCT | 6 | 5.8 | 0.2 | 17.7 |
| 50 | AGAG/CTCT | 12 | 26.6 | 8.0 | 17.9 |
| 51 | AGAC/GTCT | 6 | 4.0 | 0.2 | 20.1 |
| 52 | AGAT/ATCT | 13 | 27.3 | 3.8 | 14.5 |
| 53 | GGAA/TTCC | 7 | 8.0 | 0.3 | 10.6 |
| 54 | GGAG/CTCC | 2 | 5.5 | 0.4 | 13.3 |
| 55 | GGAC/GTCC | 10 | 2.6 | – | 18.0 |
| 56 | GGAT/ATCC | 12 | 12.4 | 1.4 | 11.4 |
| 57 | CGAA/TTCG | 9 | 27.4 | 4.2 | 18.1 |
| 58 | CGAG/CTCG | 5 | 19.0 | 3.2 | 12.6 |
| 59 | CGAC/GTCC | 12 | 3.3 | 0.1 | 9.2 |
| 60 | CGAT/ATCG | 13 | 3.0 | 0.1 | 19.2 |
| 61 | TGAA/TTCA | 5 | 3.3 | 0.1 | 12.5 |
| 62 | TGAG/CTCA | 10 | 5.2 | 0.1 | 8.9 |

(continued on next page)

Table 2 (*continued*)

| No. | Tetranucleotide | *n* | $\alpha 10^{-19}$(J) | rmsd | [24] |
|---|---|---|---|---|---|
| 63 | TGAC/GTCA | 8 | 2.8 | – | 13.7 |
| 64 | TGAT/ATCA | 10 | 3.6 | 0.2 | 9.0 |
| 65 | AGGA/TCCT | 8 | 5.8 | 0.2 | 11.7 |
| 66 | AGGG/CCCT | 4 | 8.6 | 0.6 | 10.8 |
| 67 | AGGC/GCCT | 8 | 3.2 | 0.1 | 15.9 |
| 68 | AGGT/ACCT | 6 | 7.7 | 0.3 | 15.7 |
| 69 | GGGA/TCCC | 3 | 6.0 | 0.4 | 13.8 |
| 70 | GGGG/CCCC | 5 | 28.5 | 11.0 | 16.4 |
| 71 | GGGC/GCCC | 5 | 5.7 | 0.3 | 19.6 |
| 72 | GGGT/ACCC | 7 | 3.0 | 0.1 | 18.3 |
| 73 | CGGA/TCCG | 10 | 3.4 | 0.1 | 7.9 |
| 74 | CGGG/CCCG | 5 | 3.3 | 0.1 | 11.9 |
| 75 | CGGC/GCCG | 5 | 3.9 | 0.1 | 14.5 |
| 76 | CGGT/ACCG | 7 | 3.2 | – | 15.5 |
| 77 | TGGA/TCCA | 11 | 29.9 | 5.1 | 8.6 |
| 78 | TGGG/CCCA | 6 | 3.7 | 0.1 | 11.1 |
| 79 | TGGC/GCCA | 5 | 41.1 | 24.9 | 10.3 |
| 80 | TGGT/ACCA | 9 | 16.6 | 2.5 | 10.6 |
| 81 | ACAA/TTGT | 7 | 39.2 | 13.7 | 9.3 |
| 82 | ACAG/CTGT | 7 | 1.9 | 0.1 | 11.2 |
| 83 | ACAC/GTGT | 7 | 46.3 | 19.7 | 16.2 |
| 84 | ACAT/ATGT | 12 | 2.7 | – | 13.5 |
| 85 | GCAA/TTGC | 14 | 2.5 | 0.1 | 8.7 |
| 86 | GCAG/CTGC | 15 | 26.8 | 6.5 | 7.2 |
| 87 | GCAC/GTGC | 12 | 2.3 | 0.1 | 6.6 |
| 88 | GCAT/ATGC | 19 | 3.3 | 0.2 | 8.0 |
| 89 | CCAA/TTGG | 7 | 2.7 | 0.1 | 13.1 |
| 90 | CCAG/CTGG | 4 | 3.1 | 0.2 | 9.1 |
| 91 | CCAC/GTGG | 10 | 2.0 | 0.1 | 7.9 |
| 92 | CCAT/ATGG | 10 | 1.8 | – | 11.7 |
| 93 | TCAA/TTGA | 4 | 1.6 | – | 11.6 |
| 94 | TCAG/CTGA | 7 | 2.3 | 0.1 | 14.5 |
| 95 | TCAC/GTGA | 9 | 1.6 | – | 9.9 |
| 96 | TCAT/ATGA | 13 | 1.6 | – | 10.7 |
| 97 | AATA/TATT | 7 | 6.8 | 0.5 | 14.8 |
| 98 | AATG/CATT | 8 | 11.3 | 1.2 | 12.7 |
| 99 | AATC/GATT | 6 | 30.0 | 13.3 | 26.8 |
| 100 | AATT/AATT | 8 | 6.2 | 0.8 | 26.9 |
| 101 | GATA/TATC | 8 | 3.5 | 0.1 | 10.4 |
| 102 | GATG/CATC | 11 | 6.8 | – | 13.9 |
| 103 | GATC/GATC | 12 | 7.5 | – | 17.0 |
| 104 | CATA/TATG | 8 | 7.0 | 0.5 | 3.2 |
| 105 | CATG/CATG | 15 | 5.4 | 0.3 | 7.7 |
| 106 | TATA/TATA | 13 | 4.2 | 0.1 | 3.6 |
| 107 | AGCA/TGCT | 9 | 27.2 | 4.0 | 9.9 |
| 108 | AGCG/CGCT | 6 | 4.2 | 0.2 | 9.7 |
| 109 | AGCC/GGCT | 11 | 13.2 | 1.0 | 16.0 |
| 110 | AGCT/AGCT | 6 | 5.0 | – | 12.2 |
| 111 | GGCA/TGCC | 23 | 7.0 | 0.7 | 12.1 |
| 112 | GGCG/CGCC | 9 | 5.5 | 0.3 | 8.6 |
| 113 | GGCC/GGCC | 10 | 3.5 | – | 19.1 |
| 114 | CGCA/TGCG | 6 | 4.9 | 0.3 | 9.6 |
| 115 | CGCG/CGCG | 9 | 3.8 | 0.2 | 10.0 |
| 116 | TGCA/TGCA | 12 | 10.8 | 1.6 | 3.8 |
| 117 | ACGA/TCGT | 5 | 2.6 | 0.1 | 8.0 |
| 118 | ACGG/CCGT | 18 | 2.5 | – | 7.3 |
| 119 | ACGC/GCGT | 4 | 4.1 | 0.1 | 11.7 |
| 120 | ACGT/ACGT | 10 | 7.0 | 0.3 | 11.5 |
| 121 | GCGA/TCGC | 4 | 3.6 | 0.2 | 7.2 |
| 122 | GCGG/CCGC | 8 | 3.6 | 0.1 | 11.4 |
| 123 | GCGC/GCGC | 12 | 3.5 | 0.2 | 3.9 |
| 124 | CCGA/TCGG | 13 | 2.9 | 0.2 | 12.1 |
| 125 | CCGG/CCGG | 14 | 2.8 | – | 16.4 |
| 126 | TCGA/TCGA | 8 | 4.8 | 0.1 | 10.2 |

Table 2 (*continued*)

| No. | Tetranucleotide | *n* | $\alpha 10^{-19}$(J) | rmsd | [24] |
|---|---|---|---|---|---|
| 127 | ATAA/TTAT | 7 | 2.7 | 0.1 | 6.0 |
| 128 | ATAG/CTAT | 9 | 12.5 | 1.7 | 11.4 |
| 129 | ATAC/GTAT | 13 | 3.3 | 0.1 | 9.4 |
| 130 | ATAT/ATAT | 10 | 3.0 | – | 8.6 |
| 131 | GTAA/TTAC | 5 | 4.9 | 0.2 | 10.6 |
| 132 | GTAG/CTAC | 2 | 16.2 | 1.4 | 12.1 |
| 133 | GTAC/GTAC | 8 | 3.1 | 0.1 | 8.0 |
| 134 | CTAA/TTAG | 7 | 4.2 | 0.1 | 5.0 |
| 135 | CTAG/CTAG | 13 | 3.7 | 0.1 | 17.8 |
| 136 | TTAA/TTAA | 2 | 3.0 | 0.1 | 19.4 |

Packer et al. [24] in a study on stacking energy. The flexibility was calculated from the curvature in the tetranucleotide potential energy surface in shift–slide contour maps. The force constants obtained are expressed in kJ $mol^{-1}$ $Å^{-2}$ and although direct comparison with our data is not possible, we reported them in order to find correlations.

As can be seen from the figure, there is no correlation between the two data sets. The force constants are distributed over a wider range and have higher values with respect to the torsional constants. This trend is more evident for the tetranucleotides bearing AA/TT as the central step. We recently observed [9] an overestimation of AA/TT step rigidity in data derived from X-ray analysis of DNA crystals and analyses derived therefrom, like that of Packer et al. [24], who probably keep this overestimation.

### *3.2. Application of the tetranucleotide code: nucleosomal DNA analysis*

The $\alpha$ values of the 136 tetranucleotides obtained in this study were used as a torsional rigidity code to obtain rigidity profiles for all DNA sequences. Starting from the identification of the first tetranucleotide of a sequence, its $\alpha$ value is taken from the code. After a shift along the sequence of one base, the second tetranucleotide is identified, its $\alpha$ value added to the previous and this process continued until the last tetranucleotide of the sequence is reached. It is easy to calculate that for a sequence containing $n$ bases, $n$-3 consecutive $\alpha$ values are present in the rigidity profile.

Instead of substituting the values of $\alpha$ of the code at each four base steps of the DNA sequence under study, we used the $\Delta\xi$ values calculated according to Eq. (1) to obtain profiles related to the twist angle variations that occur along the sequence. After this transformation, the calf thymus mean $\Delta\xi$ became 5.3°.

Widlund et al. [25] isolated, from the mouse genome, sequences able to form very stable mononucleosomes ranging in length from 117 to 151 bp. Cao et al. [26] synthesized and studied sequences unable to form mononucleosomes and ranging in length from 81 to 126 bp. Both sets of sequences are available at the address http://www.molbiotech.chalmers.se/research/mk/nuc/nuc.htm; they are composed of 50 sequences there named from phcn1 to phcn50 that indicate very stable nucleosomes and 37

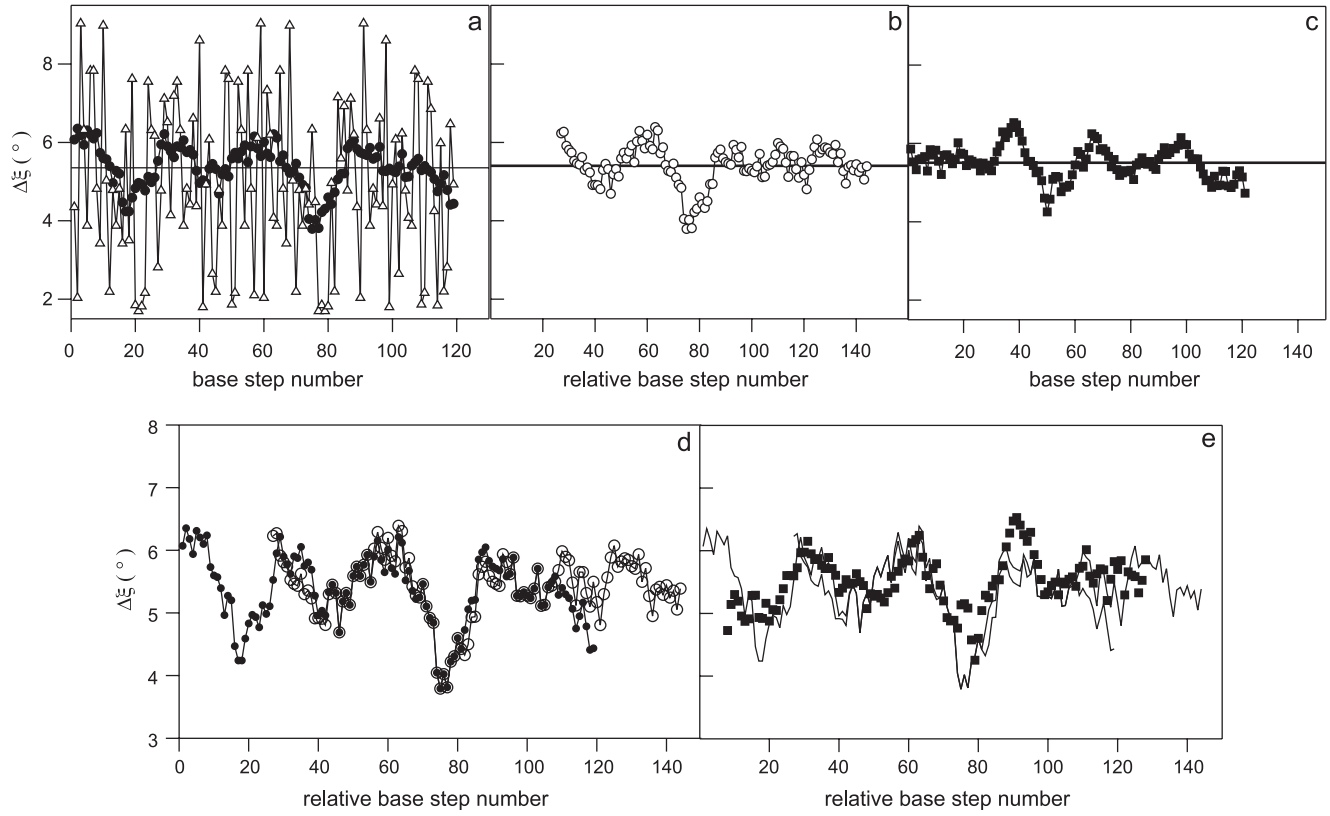


Fig. 3. Torsional rigidity profiles of nucleosomal sequences. (a) CN3 sample profiles without (△) and with (●) a smoothing window of 11 data points; (b) CN4 sample (○); (c) 5*S* rRNA from *Xenopus borealis* (■); (d) overlapping of the CN3 and CN4 rigidity profiles; (e) overlapping of the CN3, CN4 (continuous lines) and the 5*S* rRNA (■) rigidity profiles. Sequences are reported in Table 3.

sequences there named from phwn1 to phwn37 that indicate less-stable nucleosomes. The 40 sequences not able to form nucleosomes are indicated with pasHC#, where # refers to various numbers.

We obtained rigidity profiles from these sequences in order to correlate the ability of DNA to form nucleosomes with elastic properties, and referred to the original classification of the samples in a simplified manner using the notation CN1 to CN50, WN1 to WN37 and HC#.

Fig. 3a shows the rigidity profile obtained for sample CN3 without smoothing of the data. As can be seen, smoothing with a window of 11 points (about one DNA helix turn) substantially decreases the noise present in the initial data and facilitates comparison of the various rigidity profiles. It is evident from Fig. 3a that the twist angle fluctuation of CN3 is modulated around a mean value of 5.3°, coincident with the mean value of calf thymus DNA. Both profiles, with and without smoothing, show a 30-bp phase of $\Delta\xi$ that is confirmed by Fourier transform analysis of the data. In Fig. 3b, the profile of CN4 shares a similar tract with CN3 as highlighted in Fig. 3d where the two profiles are overlapped. Fig. 3c shows the profile obtained from the gene for 5*S* rRNA of *Xenopus borealis* [27], which is a well-characterized sequence able to form stable nucleosomes. Fig. 3e shows the overlapping of the CN3 and CN4 profiles and the 5*S* rRNA profile. It can be seen that, in order to obtain the best overlapping, the profile of 5*S* rRNA in Fig. 3c has to be inverted. The mean $\Delta\xi$ values for CN4 and 5*S* rRNA are 5.4° and 5.5°, and the 30-bp phasing is evident. Table 3 reports the original sequences of CN3 and CN4 from which the rigidity profiles were derived. The reported 5*S* rRNA sequence is complementary to the one taken from the EMBL database with accession number V01426. It is evident that high homology is present in the two overlapping parts of CN3 and CN4 and this may justify the similarity in their rigidity. The 5*S* rRNA has no sequence homology in the overlapping zone and only through application of the code was it possible to identify the observed similarities in rigidity.

It is known [14] that the first step of nucleosomal DNA complexation with histone octamer is the binding of the central zone of the 146 bp involved and that this zone has the strongest binding. Both CN3 and CN4 are reported [25] to be very stable nucleosome-forming sequences and we hypothesized that the common zone of their profiles contains the center of the nucleosome.

The X-ray structure of the nucleosome at 2.8-Å resolution was obtained from Luger et al. [28]. They used a 73-bp sequence of the human α satellite DNA made palindromic (NCP146; see Table 3) in order to achieve crystallization of these 146 bp and the histone octamer. The resolution was increased to 1.9 Å by Richmond et al. [29]

Table 3
Nucleosomal DNA sequence

| | |
|---|---|
| CN3 | ACGTGAAAAATGAGAAATGCACACTGAA**GGACCTGAAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATACACACTTTAG-GACGTGAAATATGGCGAGGAAAA**CTGACAAAGCGG |
| CN4 | TA**GGACCTGGAGTATGGCGAGAAAACTGAAAATCACAGAAAATGAGAAATACACACTTTAGGATGTGAAATATGGCGAGGAAAA**CT-GAAAAAGGTGGAAAATTCAGAAATGTCCACTGTAGG |
| 5SrRN | **AAAGCCTACGACACCTGGTATTCCCAGGCGGTCTCCCATCCAAGTACTACCAGCCCCGACCCTGCTTGGCTTCCGAGATCAGACGA-TATCGGGCACTTTCAGGGTGGTATGGCGTAGGC** |
| CN17 | CCGGAATATGGCGAGAAAACTGAAAATCACGTAAAATGAGAAATACACACTTTAGGACGTGAGATATCGCGAGGAAACCTGAAAAAGGTG-GAAAATTTAGAAATGTCACAGTAGGACGTG |
| CN48 | ATGTAGGACAGTGTATATCAACGAGTTACAATGAGAAACATGGAAAATGATAAAAACCACACTGTAGAACAGATTAGATGAGTGAGTTACACT-GAGAAACACATTCGTTGGAAACGGC |
| WN2 | GAGAGTAACATAGGCACAGGTGTGGAGAGTAACACAGGCACAGGTGTGGAGAGTACACACAGGCACAGGCGTGGAGAGTACACACAGGCA-CAGGTGTGGAGAGCACACACAGGCACAGGTGTGGAGAG |
| WN32 | TGCCAGATGACACGTGCTATGCCCAGGTGACACAAGCTATGCCCAGATGATACGTGCTATGCCCAGGTGACACATGCTATGCCCAGGTGACA-CATGCTATGCCCAAGTGACACATGCCATGCCAGGTGACACAAGCTA |
| HSP26 | ACGACAACAACTACATTTAAAATTTCTCGAAACTCATGGCATTTATTGGGAAAGGTTAGTTAGTTTTATTTTTTGTTTTTAGAGCAGCATT-CAATTTAGACTTTTATAAAAGAAATTTCTAATTTGATCCCTCGTTTATCAAACGATAC |
| TG5 | **TCGGTGTTAGAGCCTGTAACTCGGTGTTAGAGCCTGTAACTCGGTGTTAGAGCCTGTAACTCGGTGTTAGAGCCTGTAACTCGGTGT-TAGAGCCTGTAACTCGGTGTTAGAGCCTGTAAC** |
| MMTV A and B | ATCAATAGCCTTTATTGGCCCAACCTTGCGGTTCCCAGGGCT**TAAGTAAGTTTTTGGTTACAAACTGTTCTTAAAACGAGGATGTGAGA-CAAGTGGTTTCCTGACTTGGTTTGGTATCAAAGGTTCTGATCTGACCRCTGAGTGTTCTATTTTCCTATGTTCTTTTGGATTTTATC-CAAATCTTATG**TAAATGCTTATGTAAACCAAGATATAAAAGAGTGCTGATTTTTTGAGTAAA**CTTGCAACAGTCCTAACATT-CACCTCTTGTGTGTTTGTGTCTGTTCGCCATCCCGTTCCCGCTCGTCACTTATCCTTCACTTTCCAGAGGGTCCCCCCGCA-GACCCCGGCGACCCTCAGGTCGGCCGACTGCGGCACAGTTTTTT**GCTCCTTTTTCTAGATGTAATTTTTTAAAGCTTATTTTTTAACTTTCA-CATGGTGAT |
| NCP146 | ATCAATATCCACCTGCAGATTCTACCAAAAGTGTATTTGGAAACTGCTCCATCAAAAGGCATGTTCAGCT**$G_{71}$**AATTCAGCTGAACATGCCTTTT-GATGGAGCAGTTTCCAAATACACTTTTGGTAGAATCTGCAGGTGGATATTGAT |
| NCP147 | ATCAATATCCACCTGCAGATACTACCAAAAGTGTA**TTTGGAAACTGCTCCATCAAAAGGCATGTTCAGCTGGA(A)TCCAGCTGAA-CATGCCTTTTGATGGAGCAGTTTC**CAAATACACTTTTGGTAGTATCTGCAGGTGGATATTGAT |

Overlapping tracts, mentioned in the text, are marked in bold character.

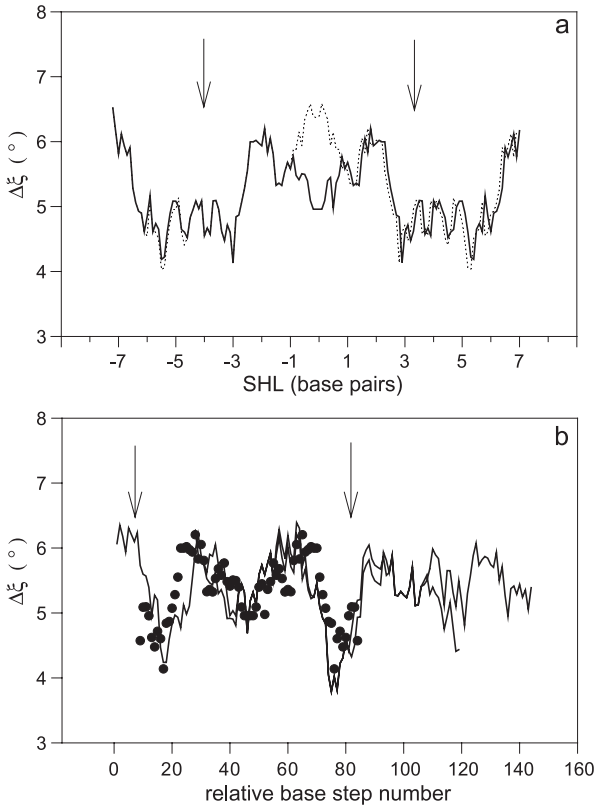


Fig. 4. Torsional rigidity profiles of DNA in crystallized nucleosomes. (a) NCP 147 (continuous line); NCP 146 (dotted line); (b) overlapping of the CN3 and CN4 profiles reported in Fig. 3e (continuous lines) and the NCP 147 profile (●). The overlapping tract is marked by arrows in panels a and b graphs. Sequences are reported in Table 3.

with a 147-bp sequence obtained through two variations of the previous sequence, as can be seen in Table 3: a substitution of the T21 with an A base and the insertion of a G after the G71. The new 147-bp sequence was made palindromic by achieving complementarity from A1 to A73 excluding the central A74. The crystals obtained with this sequence yielded an unprecedented accuracy in comparison to the previously solved nucleosome structures. We derived the rigidity profiles of the two sequences, named NCP146 and NCP147, and compared their rigidity, as shown in Fig. 4.

For the NCP147 sample, a marked reduction in rigidity is evident in the center of the sequence at the superhelix location (SHL) 0 which refers to the diad pseudosymmetry axis of the nucleosome and corresponds to position 74 of the

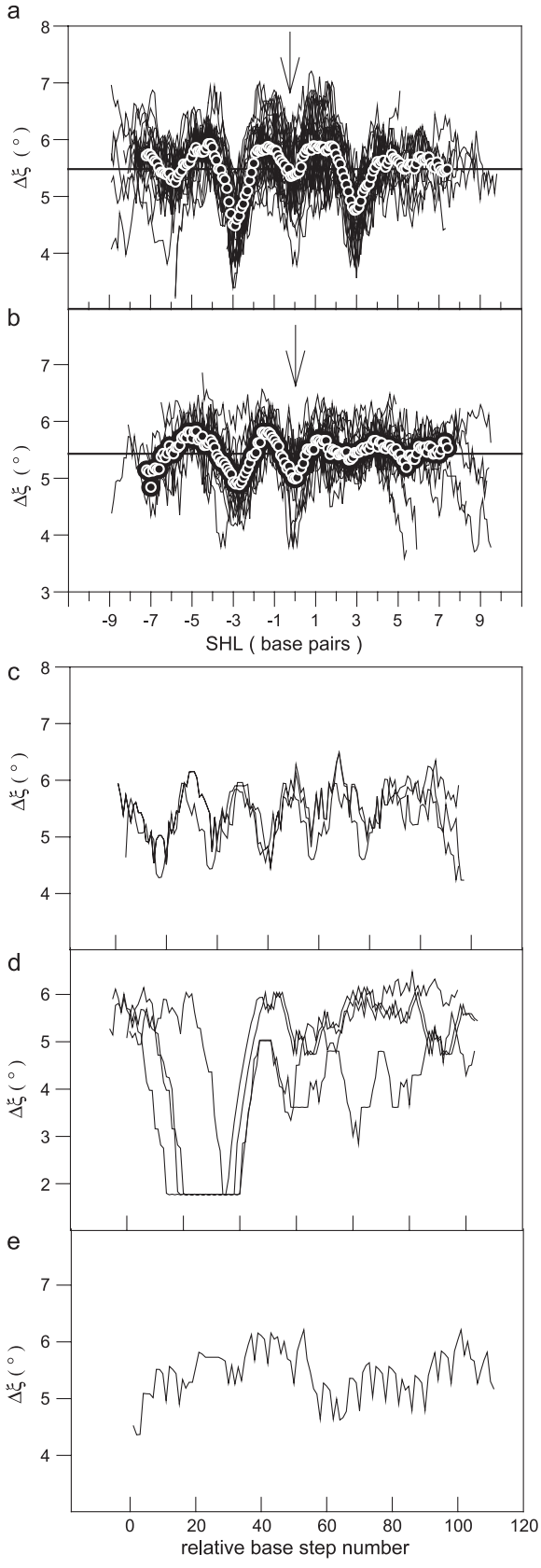


Fig. 5. Torsional rigidity profiles of nucleosomal sequences. (a) CN 1–4, 6–12, 14, 16, 18, 20, 21, 23, 26, 27, 29, 30, 37–40, 42, 43, 46–48, 50 and WN 5, 19, 26, 27, 29, 33, 35, 36 profiles (continuous lines) are overlapped; weighted mean of the profiles (white edged circles); (b) CN 15, 17, 19, 22, 24, 28, 35, 44 and WN 1, 8, 11, 12, 16, 22, 23, 28, 30, 31, 37 (continuous lines) are overlapped; weighted mean of the profiles (white edged circles); (c) WN 3, 24, 32 (continuous lines) are overlapped; (d) WN 10,14, 18, 25 (continuous lines) are overlapped; (e) HC 49 (continuous lines). Sequences are available from the address http://www.molbiotech.chalmers.se/research/mk/nuc/nuc.htm.

sequence (see Table 3). Fig. 4a shows that, whilst the effect on rigidity of the substitution around SHL-5 is very low, the insertion of the G near the center of the palindrome at SHL 0 yields a marked reduction in rigidity. These changes are mainly due to the insertion of the G base which creates the new tetranucleotide TGGA, with a $\Delta\xi$ of 2.1° instead of the existing TGAA with a $\Delta\xi$ of 6.3°. We suggest that the increased rigidity of the two DNA turns of the central sequence accounts for the higher resolution observed for NCP147 by X-ray analysis. The NCP147 profile reported in Fig. 4a contains a tract, extending from SHL −3 to SHL +3, similar to the profile of Fig. 3d between the 20 and 80 base step positions. The overlapping of the NCP147 profiles indicated by the arrows in Fig. 4a and the profiles of Fig. 3d are reported in Fig. 4b and supports our hypothesis that the rigidity profiles of the nucleosomal DNA shares a characteristic central zone.

To verify this hypothesis, we compared all the sequences named CN and WN to find as many similarities as possible in their rigidity profiles. Fig. 5 reports the result of our approach. We overlapped many different rigidity profiles by shifting and sometimes inverting the CN and WN profiles. Fig. 5a reports the 39 profiles comprising the CN3 and CN4 sequences mentioned previously. The other samples making up this figure are listed in the captions and comprise both CN- and WN-type sequences. There is a high degree of symmetry in these profiles with respect to SHL 0 (central arrow). The mean $\Delta\xi$ value of the profiles is 5.4°, almost equal to that of calf thymus DNA. The mean of all superimposed profiles is marked by white circles and the alternation of low and high rigidity tracts along the nucleosomal sequences is evident. A certain degree of specular symmetry with respect to the central zone at SHL 0 is present in the mean profile, with the rigidity modulation more marked to the left than to the right of this position. This trend is more evident in Fig. 5b. The specular symmetry disappears and the rigidity modulation is lower to the right of SHL 0. As in the previous profile, the mean $\Delta\xi$ value is 5.4°.

In the graphs of Fig. 5c, the modulation of the rigidity for three nucleosomal profiles is phased every 20 bp. In Fig. 5d, a very rigid zone is evident in four nucleosomal sequences, comprising from 20 up to 30 bp; in the adjacent zone, the rigidity reaches the usual values of $\Delta\xi$ around 5.4° without specific modulation.

Fig. 5a–d just mentioned represent 65 profiles out of the 87 examined, the others (data not shown) being a complex mixture of the ones reported—they always show rigidity modulation but with many differences between them and no sharing of coincident characteristic tracts.

Fig. 5e reports only one of the 40 profiles of the HC# sequences not able to form nucleosomes [26]. The absence of clear phasing or symmetry is evident in this and the other profiles not shown.

Fig. 6 gives three examples of the possible homologies that can be found between the rigidity of the nucleosomal profiles previously analyzed and those obtainable from the literature on nucleosomal DNA.

Widlund et al. [25] compared the strength of interaction of their samples with a 20-mer repeated sequence of DNA named TG5 [30] reported in the literature as a good nucleosome forming sequence. In Fig. 6a, the profile of

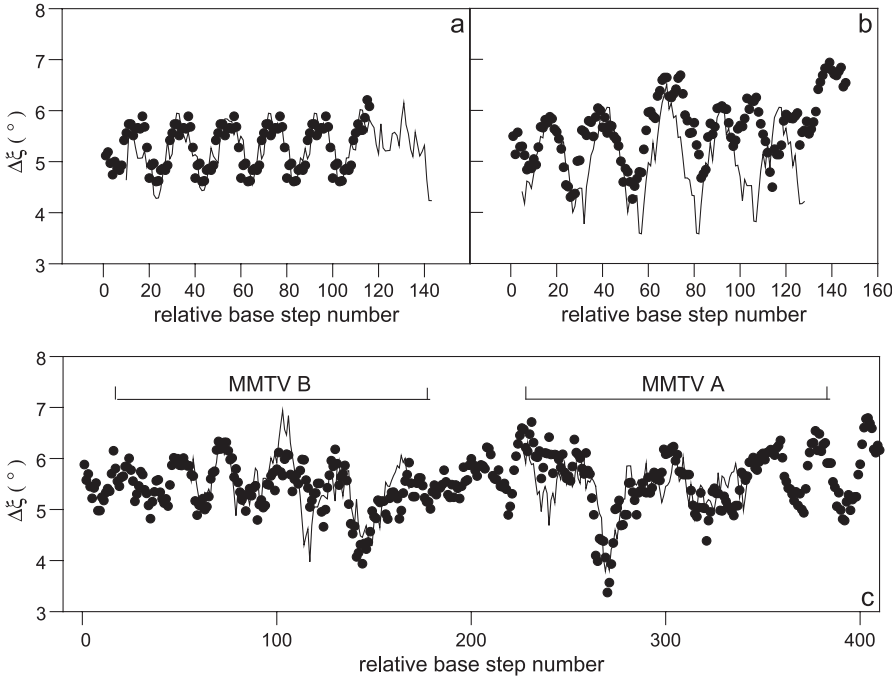


Fig. 6. Torsional rigidity profiles of nucleosomal sequences. (a) WN32 (continuous line), TG5 (●); (b) WN2 (continuous line), HSP26 (●); (c) dinucleosomes MMTV A and B profiles (●) overlapped to CN48 and CN17 (continuous lines). Sequences are reported in Table 3.

WN32 contained in Fig. 5c and the profile obtained from TG5 are overlapped and show a striking similarity.

An almost complete profile overlapping is shown in Fig. 6b and c, where the profile of WN2 is superimposed on that obtained from a nucleosome named HSP26 mapped on the *Drosophila* genome [31], and the CN48 and CN17 profiles are tentatively mapped on the dinucleosome of the mouse mammalian tumor virus (*MMTV*) [32] genes, respectively.

Fourier transform analysis of CN, WN and HC profiles were performed and the results are reported in Fig. 7. For each profile, the amplitude of the Fourier power was found for the major peak present at each phase and the sum of the Fourier amplitudes divided by the number of samples analyzed, i.e., 50, 37 and 40 for CN, WN and HC samples, respectively. This analysis was performed to find the preferential distribution of the phases in samples able to form more and less-stable nucleosomes. It is evident from the figure that CN samples, able to form the most stable nucleosomes, share a preferential phase centered at 30 bp. Less-intense peaks are present at 20 and 40 bp. In the WN samples, the preferential phases are distributed over a wider range, with two maximums at 24 and 18 bp. The HC samples show no preferential phasing evidence.

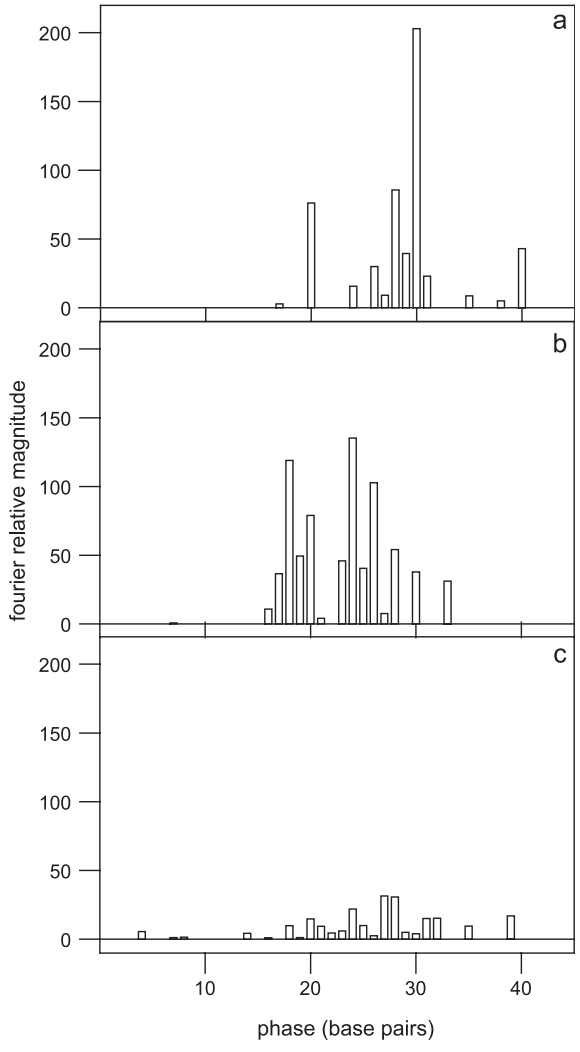


Fig. 7. Fourier transform analysis of nucleosomal DNA sequences. The Fourier relative magnitudes are reported as a function of the preferential phase for (a) CN, (b) WN and (c) HC samples. Sequences are available from the address: http://www.molbiotech.chalmers.se/research/mk/nuc/nuc.htm.

## 4. Conclusions

Early X-ray analysis of nucleosomes was obtained at 7-Å resolution for crystals obtained from native-purified nucleosomes [33] and reported the presence of sharp bends on the nucleosomal surface positioned at SHL ±1 and ±4. The DNA length between the two central bends is 20 bp and the DNA length between the left and right bends is 30 bp. A subsequent and more refined analysis of nucleosome crystals, obtained with homogeneous palindromic DNA sequences [29] with 1.9 Å resolution, located the position of the sharp bends at SHL ±1.5 and ±4–5. The distance between the bends located at the center of the nucleosome was therefore increased from 20 to 30 bp, but the left and right distances could again be averaged as 30 bp.

DNAse I fingerprinting on nucleosomal DNA [14] shows the presence of tracts resistant to digestion at positions 30, 60, 80 and 110, confirming that DNA tracts in nucleosomes spaced 20 and 30 bp assume particular conformations.

We found a correlation between the location of sharp bends on the nucleosomal surface and the length separating tracts of lower rigidity in our profiles. We assumed that less-rigid DNA tracts preferentially adapt to the histone octamer surface where sharp bends are present.

In the mean profile of rigidity shown in Fig. 5a, we can identify three tracts with lower rigidity at positions around SHL ±1–2 and −4. These sequences therefore present three preferential points of interaction spaced at approximately 30 bp, and this may be the situation in which the most stable binding of DNA takes place. In the mean profile reported in Fig. 5b, the tracts with lower rigidity are around SHL −1.5 and in the range from SHL −4 to −5. The part of this profile to the right of SHL 0 has a less-marked modulation and we can imagine that only the DNA tract to the left of SHL 0 is able to initiate and maintain the complex with DNA linked at two preferential points of interaction. These two points of interaction may also occur with the characteristic 20-bp phased distribution shown by the samples in Fig. 5c, considering that three 20-bp-spaced tracts could interact with the histones like the two 30-bp-spaced ones do. In the other profiles, such as the WN samples (Fig. 6b), there is a unique preferential point of interaction with the sharp bends located on the histone surface, and the nucleosome is less stable.

In order to verify the coherence of the tetranucleotide code, we calculated the mean $\Delta\xi$ values of some DNA sequences of the *Bos taurus* genome available in the EMBL

database. The calculation was carried out on almost 200 kbases and yielded a mean $\Delta\xi$ of 5.4°, very similar to the experimental value of 5.3° measured for calf Thymus DNA, supporting the reliability of our tetranucleotide code.

Application of the tetranucleotide code to DNA sequences able to form nucleosomes has shown that the modulation of DNA rigidity phased at 20 and/or 30 bases is responsible for nucleosome formation and stabilization. The variety of nucleosomal profiles that we observed in this study may be related to functional differences between nucleosomes.

Finally, using the tetranucleotide code reported here, mapping of nucleosomes along DNA sequences might be attempted in order to study chromatin organization. Further studies may also be carried out on DNA looping in DNA–protein interaction, the effects of mutations on DNA rigidity, promoters and consensus sequence characterization. Use of the code also allows for the design of specific sequences according to rigidity requirements for nanotechnology developments.

## Acknowledgements

We wish to thank Mikael Kubista and Hans Widlund for the availability of nucleosomal sequences.

## References

[1] M. Hogan, J. LeGrange, B. Austin, Dependence of DNA helix flexibility on base composition, Nature 304 (1983) 752–754.
[2] D.P. Millar, R.J. Robbins, A.H. Zwail, Torsion and bending of nucleic acids studied by subnanosecond time-resolved fluorescence depolarization of intercalated dyes, J. Chem. Phys. 76 (1982) 2080–2093.
[3] J.M. Schurr, Rotational diffusion of deformable macromolecules with mean local cylindrical symmetry, Chem. Phys. 84 (1984) 71–76.
[4] B.S. Fujimoto, J.M. Schurr, Dependence of the torsional rigidity of DNA on base composition, Nature 344 (1990) 175–178.
[5] J.J. Delrow, P.J. Heath, B.S. Fujimoto, J.M. Shurr, Effect of temperature on DNA secondary structure in the absence and presence of 0.5 M tetramethylammonium chloride, Biopolymers 45 (1998) 503–515.
[6] M. Collini, G. Chirico, G. Baldini, M.E. Bianchi, Conformation of short DNA fragments by modulated fluorescence polarization anisotropy, Biopolymers 36 (1995) 211–225.
[7] G. Chirico, M. Collini, K. Toth, N. Brun, J. Langowski, Rotational dynamics of curved DNA fragments studied by fluorescence polarization anisotropy, Eur. Biophys. J. 29 (2001) 597–606.
[8] F. Barone, M. Matzeu, F. Mazzei, F. Pedone, Structural and dynamical properties of two DNA oligomers with the same base composition and different sequences, Biophys. Chem. 78 (1999) 259–269.
[9] F. Pedone, F. Mazzei, M. Matzeu, F. Barone, Torsional constant of 27-mer DNA oligomers of different sequences, Biophys. Chem. 94 (2001) 175–184.
[10] M.D. Barkley, B.H. Zimm, Theory of twisting and bending of chain macromolecules: analysis of the fluorescence depolarization of DNA, J. Chem. Phys. 70 (1979) 2291–3007.
[11] K. Yanagi, G.G. Privé, R. Richardson, Analysis of local geometry in three B-DNA decamers and eight dodecamers, J. Mol. Biol. 217 (1991) 201–214.
[12] M.J. Packer, M.P. Dauncey, C. Hunter, Sequence dependent DNA structure: tetranucleotide conformational maps, J. Mol. Biol. 295 (2000) 85–103.
[13] M.A. El Hassan, G.R. Calladine, Conformational characteristics of DNA: empirical classification and a hypothesis for the conformational behaviour of dinucleotide steps, Philos. Trans. Ser. A 355 (1997) 43–100.
[14] van Holde, Chromatin, Springer Verlag, New York, 1989.
[15] R.G. Kornberg, Y. Lorch, Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome, Cell 98 (1999) 285–294.
[16] T.J. Richmond, C.A. Davey, The structure of DNA in the nucleosome core, Nature 423 (2003) 145–150.
[17] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[18] F. Barone, G. Chirico, M. Matzeu, F. Mazzei, F. Pedone, Triple helix oligomer melting measured by fluorescence polarization anisotropy, Eur. Biophys. J. 27 (1998) 137–146.
[19] S.A. Allison, J.M. Schurr, Torsion dynamics and depolarization of fluorescence of linear molecules: I. Theory and application to DNA, Chem. Phys. 41 (1979) 35–39.
[20] M. Suzuki, N. Yagi, Stereochemical basis of DNA bending by transcription factors, Nucleic Acids Res. 23 (1995) 2083–2091.
[21] E.F. Gale, E. Cundliffe, P.E. Reynolds, M.H. Richmond, M.J. Waring, The Molecular Basis of Antibiotic Action, Wiley, London, 1981.
[22] S. Dipankaran, G. Walter, Guanine quartet structures, Methods Enzymol. 211 (1992) 191–199.
[23] C.A. Hunter, Sequence dependent DNA structure. The role of base stacking interactions, J. Mol. Biol. 230 (1993) 1025–1054.
[24] M.J. Packer, M.P. Dauncey, C.H. Hunter, Sequence dependence DNA structure: tetranucleotide conformational maps, J. Mol. Biol. 295 (2000) 85–103.
[25] H.R. Widlund, H. Cao, S. Simonsson, E. Magnusson, T. Simonsson, P.E. Nielsen, J.D. Kahn, D.M. Crothers, M. Kubista, Identification and characterization of genomic nucleosome positioning sequences, J. Mol. Biol. 267 (1997) 807–817.
[26] H. Cao, H.R. Wildlund, T. Simonsson, M. Kubista, TGGA repeats impair nucleosome formation, J. Mol. Biol. 281 (1998) 253–260.
[27] R.C. Peterson, J.L. Doering, D.D. Brown, Characterization of two *Xenopus* somatic 5S DNA and one minor oocyte-specific 5S DNA, Cell 20 (1) (1980) 131–141.
[28] 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.
[29] T.J. Richmond, C.A. Davey, The structure of DNA in the nucleosome core, Nature 423 (2003) 145–150.
[30] T.E. Shrader, D.M. Crothers, Artificial nucleosome positioning sequences, Proc. Natl. Acad. Sci. U. S. A. 86 (1989) 7418–7422.
[31] G.H. Thomas, S.C.H. Elgin, Protein/DNA architecture of the DNAse I hypersensitive region of the *Drosophila* hsp26 promoter, EMBO J. 7 (1988) 2191–2201.
[32] A. Flaus, T.J. Richmond, Positioning and stability of nucleosomes on MMTV 3′ LTR sequences, J. Mol. Biol. 275 (1998) 427–441.
[33] 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.