FEBS Letters 417 (1997) 352–354 FEBS 19465

A crystallographic approach to DNA bending: prediction of nucleosome formation by DNA triple repeats and other repetitive sequences

Juan A. Subirana*

Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, Diagonal 647, E-08028 Barcelona, Spain

Received 25 September 1997

Abstract DNA bending is due to two main factors: the inherent curvature of the sequence and its flexibility. Most methods of analysis do not allow a differentiation between these two factors. In this paper I show that the flexibility of DNA sequences can be estimated from the standard deviation of roll values determined by X-ray crystallography for each base step. As an application of this approach, the nucleosome formation ability of triple repeat sequences has been determined and shown to be in agreement with the experimental results. Local variations in twist do not appear to have any influence on nucleosome formation.

Key words: DNA bending; DNA curvature; DNA triple repeat; Repetitive DNA; X-ray crystallography; Nucleosome

1. Introduction

Recently it has been shown [1] that one class of triple repeats, $(CCG)_n$, prevents the formation of nucleosomes. On the other hand triple repeats such as $(CTG)_n$ form hyperstable nucleosomes [2]. Both repeats are involved in pathological situations. In this paper I want to show that this behavior may be derived from our current knowledge of the variability of DNA structure as established from crystallographic data of oligonucleotides.

Any DNA with a strictly repetitive sequence cannot show in solution any tendency to bend in a definite direction. If we consider for example a sequence such as $(XYZ)_n$, all (..Z)XYZ(X..) trinucleotides will show an average local bend which will depend on the roll/tilt/slide values of the three individual base steps. If we consider that the conformation of the whole sequence is frozen with each trinucleotide in its average conformation, we will obtain a DNA molecule with a helical axis that is not straight, but itself follows a helix or supercoil. In other words, the local intrinsic curvature due to the features of individual XYZ triplets will be canceled since consecutive trinucleotides will regularly change their orientations: no overall spatial bend will be favored. Of course Brownian motion will distort DNA molecules from the frozen conformation that we have considered, but without imposing any favored orientation in the random bends arising.

Since repetitive sequences do not have a tendency to bend, what will happen when they encounter a histone core? This will depend on the stiffness of the sequence. If the individual base steps involved in a repetitive sequence are stiff, nucleosomes will form only with difficulty. If they bend easily, nucleosomes will form. Since bending in B-DNA is mainly due to changes in roll [3,4], the stiffness of a sequence towards bending can be estimated from the variability of the roll parameter.

2. Materials and methods

The statistical analysis of crystallographic data carried out by Gorin et al. [5] shows that the variability of roll for individual base steps in B-DNA is striking. A similar (but not identical) variability of roll has been reported [6], using a data set which includes A and B-DNA structures. For the purpose of this paper the values derived by Gorin et al. [5] appear to be more adequate, since nucleosomes have their DNA mainly in the B form. The values obtained by the latter authors show that the GC, AC, TA, CA and AG base steps have a standard deviation of roll in the range of 6.1–6.9° and can therefore be considered variable. On the other hand the AT, CG, GA, AA and GG base steps have a much smaller standard deviation (2.9–3.5°) and should be considered rigid. The tendency to form nucleosomes by repetitive sequences will only depend on the flexibility of the base steps involved, which can be determined from the variability of roll calculated by Gorin et al. [5].

3. Results

When the variability of roll values (standard deviation) described above is applied to repetitive sequences, we find a striking agreement between the observed and predicted tendency to form nucleosomes, as shown in Table 1. We conclude therefore that the local variability of roll detected in crystalline structures of DNA oligonucleotides has a predictive value for the behavior of this molecule in biological systems.

Changes in twist of the individual base steps do not appear to have any influence on nucleosome formation. The $(AT)_n$ sequence is known to show a strong alternation, with high twist in the AT step and low twist in the TA step [7]. Nevertheless this sequence can easily form nucleosomes, as shown in Table 1. In fact most repetitive sequences will tend to show alternations in twist. However, it is interesting to note that all repetitive sequences given in Table 1 will have between 10 and 10.3 base pairs per helical turn, as calculated from the average values of twist found in oligonucleotide crystals [8]. Thus the average helical structure of all the triple repeats coincides with the average structure of mixed sequence DNA.

It should be stressed that the nucleosome formation tendencies given in Table 1 only refer to strictly repetitive sequences. In a non-repetitive sequence the nucleosome forming ability will depend both on the flexibility of the sequence and on the intrinsic bending of the individual sequences. Thus the trinucleotide bending propensities determined from DNase cutting rates [9] are not directly related to the patterns presented in Table 1, since by the latter method it is not possible to deter-

*Fax: (34) (3) 401.71.50.
E-mail: Subirana@EQ.UPC.ES

PII S0014-5793(97)01317-3

Table 1
Nucleosome formation by DNA repeats

DNA repeat sequence	Roll variability (σ)	Nucleosome formation	
		Predicted	Found
A/T	3.2	No	No [16,17]
G/C	2.9	No	No [16,17]
CG	3.2/6.1	Yes	Yes [16,17]
AT	3.5/6.4	Yes	Yes [16,17]
GA/TC	3.0/6.9	Yes	–
GT/AC	6.7/6.3	Yes (easy)	–
AAG/CTT	3.2/6.9/3.0	No	–
GGA/TCC	2.9/3.0/6.9	No	–
AGC/GCT	6.9/6.1/6.3	Yes (easy)	Yes (hyperstable) [2]
AGT/ACT	6.9/6.7/6.4	Yes (easy)	–
GAC/GTC	3.0/6.7/3.2	No	–
GAT/ATC	3.0/3.5/6.3	No	–
GGC/GCC	2.9/6.1/3.2	No	No [1]
GGT/ACC	2.9/6.7/6.3	Yes	–
AAC/GTT	3.2/6.7/6.3	Yes	–
AAT/ATT	3.2/3.5/6.4	No	–

The roll variability is measured by the standard deviation of the values found in oligonucleotide crystalline structures [5]. It is assumed that nucleosome formation will be favored when at least 50% of the base steps in the sequence have a variable roll ($\sigma > 6°$). In order to prevent nucleosome formation a sufficient length of the repeated sequence should be present.

mine whether the base sequences analyzed are flexible or inherently bent.

4. Discussion

In order to validate the predictions based on Table 1 it would be of interest to determine experimentally the nucleosome forming ability of the other trinucleotide sequences and of longer repetitive sequences. Exceptions to the rules presented in Table 1 might be found, since nearest neighbors may have an effect on the central base step, in particular in pyrimidine/purine steps. For example, the CA step in the sequence CCAA/TTGG appears to be more variable than in GCAA/TTGC [8]. Unfortunately there are not enough crystallographic data to introduce this effect in Table 1. Also some of the roll values determined by X-ray diffraction may be biased by packing effects in the crystal. Furthermore, if some triplet repeats have unusual helical structures [10], this would certainly have an influence on their nucleosome forming ability.

The approach presented in this paper is in general agreement with the results obtained using other methods.

1. The GAT(G) repeats have recently been found to be reluctant to form nucleosomes [11], as predicted by the values given in Table 1.
2. The AGC repeat has been confirmed to be inherently flexible [12].
3. In the study of Satchwell et al. [13], the AAT, AAA and GGC triplets showed a strong positional preference on nucleosomes, which indicates that the two steps in these triplets should be rather rigid, in agreement with their low variability of roll. The only exception is the GC step, which has a variable roll. However, these results are not strictly comparable, since Satchwell et al. [13] studied sequences that are not repetitive.
4. In a related study it was found [14] that the AAA, TGG and CGG trinucleotides prefer to lie outside of the nucleosome cores, which may also indicate that the two base steps present in each triplet are not flexible. This result is also in agreement with the roll values determined by Gorin et al. [5], except for the TG/CA step which is predicted to be variable. In this case the influence of neighboring sequences may determine the behavior of the TGG triplet.

Finally it is interesting to note that an extended repeat of the dodecamer sequence CCCCGCCCCGCG is involved in progressive myoclonus epilepsy [15]. This sequence contains nine rigid base pairs and only three flexible ones, so that it should be expected to be rigid and not able to form nucleosomes. However, the relationship between DNA structure and pathological situations is not clear, as discussed by Lalioti et al. [15]. In fact, both AGC and GGC expanded repeats are known to be involved in pathological situations, whereas the results available show that these two sequences respectively do not form nucleosomes [1] or form very stable ones [2].

References

[1] Wang, Y.-H., Gellibolian, R., Shimizu, M., Wells, R.D. and Griffith, J. (1996) J. Mol. Biol. 263, 511–516.
[2] Wang, Y.-H., Amirhaeri, S., Kang, S., Wells, R.D. and Griffith, J.D. (1994) Science 265, 669–671.
[3] Young, M.A., Ravishanker, G., Beveridge, D.L. and Berman, H.M. (1995) Biophys. J. 68, 2454–2468.
[4] Dickerson, R.E., Goodsell, D. and Kopka, M.L. (1996) J. Mol. Biol. 256, 108–125.
[5] Gorin, A.A., Zhurkin, V.B. and Olson, W.K. (1995) J. Mol. Biol. 247, 34–48.
[6] El Hassan, M.A. and Calladine, C.R. (1997) Phil. Trans. R. Soc. Lond. A 355, 43–100.
[7] Yuan, H., Quintana, J. and Dickerson, R.E. (1992) Biochemistry 31, 8009–8021.
[8] Subirana, J.A. and Faria, T. (1997) Biophys. J. 73, 333–338.
[9] Brukner, I., Sánchez, R., Suck, D. and Pongor, S. (1995) EMBO J. 14, 1812–1818.
[10] Chastain II, P.D., Eichler, E.E., Kang, S., Nelson, D.L., Levene, S.D. and Sinden, R.R. (1995) Biochemistry 34, 16125–16131.

[11] Cao, H., Widlund, H.R. and Kubista, M. (1997) J. Biomol. Struct. Dyn. 14, 920.
[12] Sinden, R.R., Chatain II, P.D., Eichler, E.E., Levene, S.D., Nelson, D.L. and Pearson, C.E. (1997) J. Biomol. Struct. Dyn. 14, 881.
[13] Satchwell, S.C., Drew, H.R. and Travers, A.A. (1986) J. Mol. Biol. 191, 659–675.
[14] Drew, H.R. and Calladine, C.R. (1987) J. Mol. Biol. 195, 143–173.
[15] Lalioti, M.D., Scott, H.S., Buresi, C., Rossier, C., Bottani, A., Morris, M.A., Malafosse, A. and Antonarakis, S.E. (1997) Nature 386, 847–851.
[16] Rhodes, D. (1979) Nucleic Acids Res. 6, 1805–1816.
[17] Simpson, R.T. and Künzler, P. (1979) Nucleic Acids Res. 6, 1387–1415.