Theoretical study of cruciform states in superhelical DNAs

A.V. Vologodskii and M.D. Frank-Kamenetskii

Institute of Molecular Genetics, USSR Academy of Sciences, Moscow 123182, USSR

Received 15 March 1982

1. INTRODUCTION

The most probable sites for the attack of single-strand specific endonucleases in some superhelical DNAs are the centers of the longest palindromic sequences [1,2]. These findings accord with the earlier theoretical prediction that the probability of cruciform states for large palindromes should increase with superhelix density [3]. Does this novel conformational state of DNA play any biological role? A search for the answer to this question has begun. It may be of interest to know how the probability of occurrence of cruciform states depends on the various conditions (e.g., temperature, ionic environments, superhelix density). We report here the theoretical calculations of the cruciform state probability for φX174, fd, SV40, pBR322 and pAO3 (part of ColE1) DNAs. Our results show that this probability is a weak function of temperature and ionic strength whereas its dependence on the superhelix density is most spectacular. We compare these results with the available experimental data.

2. METHODS OF CALCULATION

Theoretical analysis of structural perturbations in superhelical DNA begins with the addition to the ordinary microstate energy of a linear molecule of a term allowing for the topological constraints inherent in the closed circular form of DNA [3–6]. This additional term is known as superhelix energy G and it is assumed to depend on three variables: superhelix density σ the fraction of open basepairs ϑ and the fraction of basepairs μ participating in the formation of cruciform structures. (Generally speaking one has to consider all the other states that affect the superhelix energy, such as the Z-form. However, we are not in a position to allow for the transition to the Z-form because, in contrast with the open and cruciform states, the energy parameters of this transition are unknown. So we confine ourselves to a consideration of $G(\sigma, \vartheta, \mu)$ as a function of three variables.)

The $G(\sigma, \vartheta, \mu)$ function is unavailable directly from the experimental data. However, a reliable expression for G may be obtained on the basis of experimental data and sound theoretical considerations [6]. We considered the case of $\mu = 0$ but allowance for $\mu \neq 0$ does not present any difficulty because the formation of cruciform structures is tantamount to the elimination of a certain number of basepairs from the molecule without changing the linking number. This means that instead of σ we have to substitute the value of $\sigma + \mu$ into the equation obtained in [6], and get:

$$G(\sigma,\vartheta,\mu) = 10\,RTN\,[(1-b)(\sigma+\vartheta+\mu)^2 + b(\sigma+\mu)^2] \quad (1)$$

Earlier [3,4] we calculated the probability of open and cruciform states assuming that $b = 0$. However, our recent analysis showed that the correct values of b are 0.2 and 0.4 depending on the model chosen for the open state, i.e., depending on the first term of the overall expression for microstate energy [6]. Using eq. (1) and the simplest model of helix–coil transition (the Ising model) we can calculate the cruciform state probability in any DNA provided that its sequence is known.

The algorithm for such calculations was proposed in [4]. However, the algorithm as presented in [4] corresponded to a specific case when the change in superhelix energy due to opening of a basepair:

$$\delta G_\vartheta = \frac{1}{N}\frac{\partial G}{\partial \vartheta}$$

Published by Elsevier Biomedical Press
00145793/82/0000–0000/$2.75

was equal to the change in the superhelix energy due to the addition of a basepair to the cruciform state:

$$\delta G_\mu = \frac{1}{N}\frac{\partial G}{\partial \mu}$$

One can see that this is the case for eq. (1) only if $b = 0$. So we have used a modified procedure which allows calculations for any b. In the first step of the procedure we calculated the self-consistent value of ϑ for $\mu = 0$ using algorithm (2) from [4]. To do this we chose an interval $[a,b]$ which should contain the self-consistent value, e.g., the interval [0,1]. For the value $\vartheta_1 = (b-a)/2$ and $\mu = 0$, the values δG_ϑ and δG_μ were calculated using eq. (1). These values made it possible to calculate the corresponding statistical weights of the helical and cruciform states and the equilibrium value of ϑ. If it proved to be less than ϑ_1, then ϑ_2 would be chosen so as:

$$\vartheta_2 = \vartheta_1 - \frac{b-a}{4}$$

otherwise it would be taken as:

$$\vartheta_2 = \vartheta_1 + \frac{b-a}{4}$$

Next the corresponding values of δG_ϑ and δG_μ were calculated for the new ϑ value and $\mu = 0$ and so on. As a result we obtained the self-consistent value of ϑ at $\mu = 0$. In the second step of the procedure we obtained the self-constistent value of μ by the same method taking the ϑ value determined in the first step. Then we found the new self-consistent value of ϑ for the obtained value of μ. A few such steps proved to be sufficient to obtain the self-consistent values of ϑ and μ with an accuracy of 10^{-5}.

The calculations were performed for the following thermodynamic parameters of DNA: the melting temperature of AT pairs $T_{AT} = 70°C$; $T_{GC} = 111°C$, the melting enthalpy of AT pairs $U_{AT} = 8550$ cal/mol, cooperativity factor $\exp(-F_s/RT) = 5 \times 10^{-5}$, which correspond to normal ionic conditions 1 × SSC. The loop-weighting factor α was assumed to be zero and in eq. (1) $b = 0.4$. This pair of parameters leads practically to the same results as the pair $\alpha = 1.5$ and $b = 0.2$, which corresponds to a more realistic helix–coil model [6]. We have allowed for the possible formation of isolated non-complementary pairs AC and GT within the cruciform stems and assumed the equilibrium constants for them equal to unity.

There is an ambiguity in our calculations stemming from the energy of the helical boundaries associated with the appearance of a cruciform structure. In the model used we assumed that the boundary formation increases the energy of the microstate by $F_s/2$. A cruciform structure increases the energy by $3F_s$ because it is associated with the formation of 6 boundaries. Our quantitative results are sensitive to the choice of this energy; however, it would be impossible to refine until quantitative experimental data on the probability of cruciforms are available.

3. RESULTS AND DISCUSSION

Using the above methods we calculated the probability of cruciforms for the DNA of viruses SV40, φX174 and fd of plasmids pAO3 (a part of the

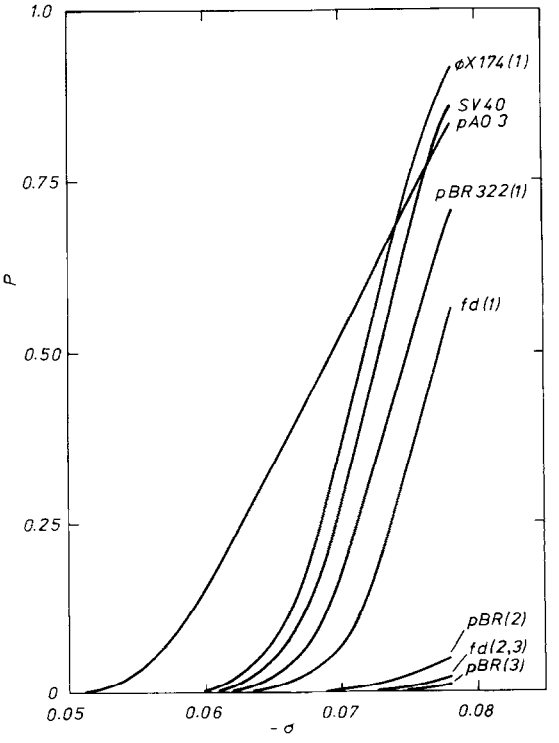

Fig.1. The calculated probability of cruciform states in palindromic regions as a function of superhelix density using eq. (1) with $b = 0.4$ as superhelix energy. The sequences and positions of the palindromes are presented in table 1. The plot embraces all cruciforms with an occurrence probability higher than 0.02 at $-\sigma = 0.08$.

ColE1 plasmid) and pBR322 for different values of the superhelix density. Fig.1 shows the results for the most probable cruciforms at 37°C and 1 × SSC. The corresponding palindromes are presented in table 1. One can see that the probability of cruciform occurrence for the palindromic regions under study becomes significant within the physiological range of superhelix densities (e.g., [7]); within the same range of superhelix densities and under the same environmental conditions the opening probability does not exceed 0.02.

The most striking feature of our results is a dramatic increase of the cruciform state probability with growing superhelix density. A change of σ by as little as 0.01 may lead to a change of two orders of magnitude in the cruciform probability! The only exception is pAO3 DNA, which is so small that the formation of the cruciform structure leads to an appreciable relaxation of the tension in the whole molecule.

Table 1

Palindromes for which the probabilities of cruciform states are shown in fig.1 and 3

AGCGGTAAAAATTTTAATTTTTGCCGCT 2316	ϕX 174	(1)
GCCTCCAAATCTTGGAGGC 3956	ϕX 174	(2)
AAAGGCTCCTTTTGGAGCCTTT 1538	fd	(1)
TAATACATCTAAATCCTCAAATGTATTA 4685	fd	(2)
AACCTCCCGCAAGTCGGGAGGTT 3320	fd	(3)
CAGAGGCCGAGGCGGCCTCGGCCTCTG 3448	SV40	
AAACCACCGCTGGTAGCGGTGGTTT 3053	pBR 322	(1)
AAAAGGATCTTCACCTAGATCCTTTT 3208	pBR 322	(2)
AAAGGATCTCAAGAAGATCCTTT 3112	pBR 322	(3)
GTCCTAGCAATCCCATTTGGATTGCTAGGAC 1499	pAO 3	

Our calculations have shown that the probability of cruciforms in a specific palindrome is quite insensitive to all factors other than superhelix density σ. Fig.2 shows the temperature dependence of the cruciform state probability for pAO3 DNA for $-\sigma = 0.078$ (this is the native superhelix density of ColE1 DNA). The probability proved to be as weakly dependent on the ionic strength.

Although formerly we used an incorrect equation for G (corresponding to eq. (1) with $b = 0$) the final results have changed only slightly. Fig.3 shows the results of calculations for $b = 0$. The results are virtually the same as in fig.1. This makes our results more reliable. Since now we allow for the formation of non-complementary AC and GT pairs, whereas formerly we considered only ideal hairpins, the main palindrome in ϕX174 DNA now proves to be centered in position 2330 instead of 3965 as in [3]. These results confirm the main conclusion in [3] about the dramatic increase of the cruciform state probability for sufficiently large palindromes with growing superhelix density.

A quantitative comparison of theory with experiment cannot be made because of the qualitative nature of the available experimental data. Nevertheless, our data agree with experiment:

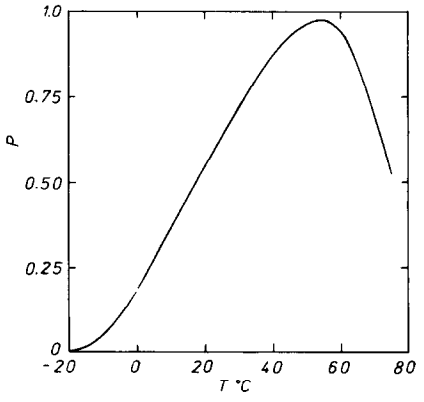

Fig.2. The calculated probability of cruciform states for pAO3 DNA as a function of temperature ($-\sigma = 0.078$).

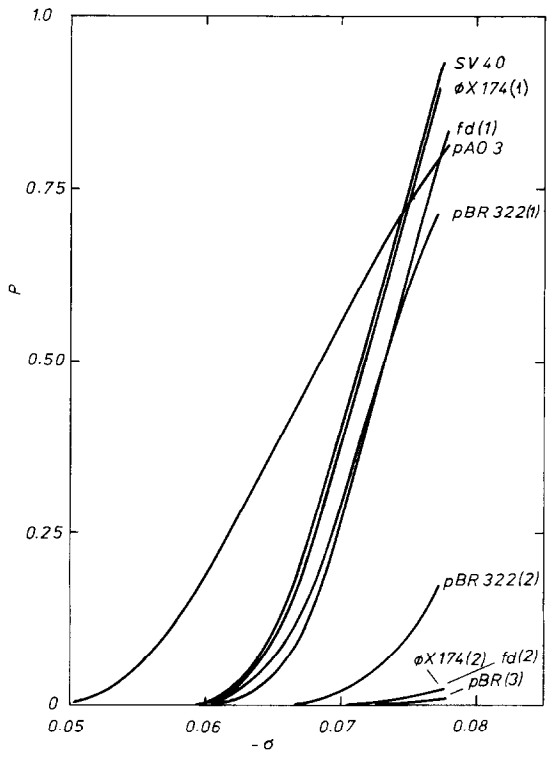

Fig.3. The same calculations as in fig.1 but for the formerly assumed superhelix energy equation [3], corresponding to eq.(1) with $b = 0$.

(i) The fact that a single-strand-specific endonuclease cuts superhelical DNAs at specific sites corresponding to the centers of the largest palindromic regions in consistent with results of our calculations.

(ii) Our data on the spectacular dependence of the probability of cruciform states on superhelix density may explain some peculiarities of the experimental data.

In [2] cruciforms may not have been observed in φX174 DNA because the preparations had a slightly lower superhelix density than those in [1] where a cruciform structure at position 2330 was observed (fig.1, table 1). Similarly, the failure to observe the cruciform structure in SV40 DNA is obviously due to the comparatively low superhelix density of this DNA ($-\sigma = 0.05$). Our results predict the appearance of a cruciform structure in this DNA centered in position 3448 at $-\sigma > 0.06$. The same is true for fd DNA. Our results agree very well with Lilley's data [1] on the distribution of probabilities of 3 cruciforms in pBR322 (fig.1, table 1). Our results concerning the weak temperature dependence of the cruciform state probability also correlate with [1].

The theory explains, at least qualitatively, all the major features of the available experimental data. We believe that our results may form a reliable basis for the further study of the influence of cruciform states on the physical, chemical and biological properties of supercoiled DNAs.

REFERENCES

[1] Lilley, D.M. (1980) Proc. Natl. Acad. Sci. USA 77, 6468–6472.
[2] Panayotatos, N. and Wells, R.D. (1981) Nature 289, 466–470.
[3] Vologodskii, A.V., Lukashin, A.V., Anshelevich, V.V. and Frank-Kamenetskii, M.D. (1979) Nucleic Acids Res. 6, 967–982.
[4] Anshelevich, V.V., Vologodskii, A.V., Lukashin, A.V. and Frank-Kamenetskii, M.D. (1979) Biopolymers 18, 2733–2744.
[5] Benham, C.J. (1980) J. Chem. Phys. 72, 3633–3639.
[6] Vologodskii, A.V. and Frank-Kamenetskii, M.D. (1981) FEBS Lett. 131, 178–180.
[7] Bauer, W.R. (1978) Annu. Rev. Biophys. Bioeng. 7, 287–313.