The relaxation time for a cruciform structure in superhelical DNA

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

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

Received 27 June 1983

We have calculated the relaxation time of a cruciform structure in superhelical DNA as a function of the superhelix density for palindromic regions of different lengths. The relaxation time has a sharp maximum at the superhelix density which corresponds to the equilibrium transition point between the cruciform structure and the regular double helix. This maximal value is shown to depend dramatically on the length of the palindromic region.

Cruciform structure *Inverted repeat* *Palindrome* *Superhelical DNA* *Kinetics* *Relaxation time* *Superhelix density*

1. INTRODUCTION

DNA in a cell usually has the form of a negative superhelix. Palindromic sequences (inverted repeats) of such DNA may exist as cruciform structures. Such structures were predicted theoretically [1,2] and first observed for artificial palindromes in [3]. Then the evidence of cruciform states in natural DNA was obtained with the aid of the single-strand specific endonuclease [4,5]. The final proof of the existence of the cruciform structures in natural DNA was obtained with the help of two-dimensional gel electrophoresis [6]. One of the striking features of a cruciform structure is its large relaxation time, as first observed in [7] for a giant artificial palindrome (about 1 min). The relaxation time of a 31 bp natural palindrome was shown to be as large as many hours [6].

What is the reason for such an unexpectedly large relaxation time? How does it depend on the superhelix density and the length of the palindromic region? To answer these questions we have performed theoretical calculations of the relaxation time for the transition between the regular double helix and a cruciform structure in superhelical DNA. The relaxation time is shown to depend dramatically on the superhelix density, reaching a maximum of many hours or more.

2. THEORY

We consider the process of the first order:

$$B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \qquad (1)$$

where by B we denote the regular double helix and by C the cruciform structure. The relaxation time τ is:

$$\tau = (k_1 + k_2)^{-1} \qquad (2)$$

We will calculate the back rate constant k_2 and the equilibrium constant K and then find the forward rate constant $k_1 = Kk_2$.

There are two different routes of decay of a cruciform structure, both leading to the formation of the regular double helix. The first route involves the sequential disruption of base pairs forming the stems of the two hairpins of the cruciform structure. As a result a large open region is formed as

Published by Elsevier Science Publishers B.V.
00145793/83/$ 3.00

an intermediate state that eventually collapses into the regular double helix. The elementary step of the second route consists in the concerted disruption of two basepairs, one from each stem, immediately followed by the formation of two additional basepairs in the principal double helix. This route also involves the transient formation of an open region, though the open region is much smaller than in the first route. It is not formed until the hairpins become so small that their stems are unstable.

Which of the two routes is more rapid depends on the temperature and the value of superhelix density. In any situation that may be of interest the first route is always less favourable than the second one. Indeed, the formation of an open basepair requires the free energy [8]:

$$\Delta F = (T_m - T)\frac{\Delta H}{T_m} + 20(1-b)RT\sigma \qquad (3)$$

The first term in the right part of this equation is the energy increase due to the disruption of the basepair (T_m is the DNA melting temperature, ΔH is the enthalpy of melting) and the second term is the energy decrease due to the relaxation of stress in the principal double helix as a result of this disruption (σ is the superhelix density, see [8]). Substituting the typical values of $\Delta H = 8.5$ kcal.mol^{-1}, $T_m = 352$ K, $T = 293$ K and $b = 0.4$ (see [8]) gives a ΔF value which is positive up to the $-\sigma = 0.1$. Note that we have chosen as T_m the melting temperature of AT pairs. So even the melting of AT pairs is unfavourable under real conditions. We conclude that for each elementary step the second route is preferable to the first one. In a qualitative discussion of the cruciform kinetics, a similar conclusion was reached [9].

It is easy to find the back rate constant k_2 for the second route because the elementary rate constant for the transition of two basepairs from one bar of the cross to the other has been thoroughly studied [10]. These authors have studied artificial complexes formed by 4 single strands. Such a complex has the form of a cross with 4 double-stranded arms, their ends free. The crossing point can freely migrate experiencing random walk up to complete dissociation that happens when one of the bars disappears and the cross turns into two identical double helices. The elementary rate constant k_+ for the transition of two basepairs from one bar of the cross to the other have been determined [10]. That is the value that we need to calculate the k_2 value.

In contrast with the free-ended cross in [10], which corresponded to undirected random walks of the crossing point, in our case of a cruciform structure in closed circular DNA the shift of the crossing point entails a change of the superhelix energy. Since the negative superhelicity is inherent in natural DNA the subtraction of one basepair from each of the two hairpins should lead to an increase of the superhelix energy:

$$\delta G = -40RT\sigma \qquad (4)$$

Eq.(4) immediately follows from the general equation of the superhelix energy (see [11]) if one recalls that the elementary step of the crossing point is a simultaneous transition of two basepairs from the two hairpins to the principal double helix.

The desired k_2 value is the reciprocal mean time required before the stems of the hairpins reach their critical length, whereupon they become unstable and form an open region. Such a region rapidly collapses into the regular double helix.

The calculation of this mean time corresponds to the well-known 'gambler's ruin' problem of the probability theory (see [12,13]) and the result is:

$$k_2 = k_+ \frac{(\alpha-1)^2}{\alpha}[\alpha^{n-m-1} - 1 - (n-m+1)(\alpha+1)]^{-1} \qquad (5)$$

where:

m = the critical length of the hairpin stems;
n = the maximum number of basepairs in the cruciform structure under consideration;
α = $\exp(-40\sigma)$

In most cases one may use a simplified equation:

$$k_2 = k_+ \alpha^{-(n-m)} \qquad (6)$$

The equilibrium constant K of the process determined by eq.(1) is equal to [2,11]:

$$K = \hat{\sigma}^3 \alpha^{n+\lambda/2} \prod_{i=1}^{\lambda} s_i^{-1} \qquad (7)$$

Here $\hat{\sigma}$ is the cooperativity factor for the helix–coil

transition, λ is the number of bases comprising the loop of the cruciform structure, s_i is the equilibrium constant for the opening of the i-th base pair to form the loop.

Thus for the rate constant k_1 of the cruciform formation one obtains:

$$k_1 = k_+ \hat{\sigma}^3 \alpha^{m+\lambda/2} \prod_{i=1}^{\lambda} s_i^{-1} \quad (8)$$

The only quantity in eq.(6),(8) that remains to be determined is m. The probability of the decay of a small cruciform structure containing m basepairs in its helical stem becomes about unity when the time of its decay is of the same order of magnitude as the time of the elementary step of the crossing point migration $1/k_+$. The mean time of decay of the helical region has been estimated in [13,14]. Using their result one can obtain an equation to define m:

$$\frac{1}{k_+} = \frac{1}{2k_f} s^{m-\nu}/(m-\nu) \quad (9)$$

Here s is the mean stability constant of a basepair in stem, k_f is the growth rate constant of a helical region and ν is the minimal number of basepairs in a helical region in an isolated hairpin (the size of nucleating region). The values k_f and ν were determined [14] as $k_f = 10^6\ s^{-1}$ and $\nu = 2$.

Let us now turn to the quantitative estimates. At room temperature and a sufficiently high ionic strength $k_+ = 2 \times 10^3\ s^{-1}$ [10]. Assuming that DNA with a GC-content of 50% melts at 80°C (this corresponds to the conditions in [6]) and taking $\Delta H = 8.5$ kcal.mol^{-1} one obtains $s = 11.76$. Substituting the values of k_+, k_f, ν and s into eq.(9) we obtain $m = 5$. Note that virtually the same value of m is obtained if one considers m as the length of a hairpin stem melting at about 20°C.

3. RESULTS AND DISCUSSION

We can now calculate the dependence of the relaxation time τ of the transition between the cruciform structure and the regular double helix on the superhelix density σ. Fig.1 shows the results for different numbers of basepairs in a palindromic region and for $m = 5$. The value of $n = 13$ corresponds to the palindromic region studied in [6].

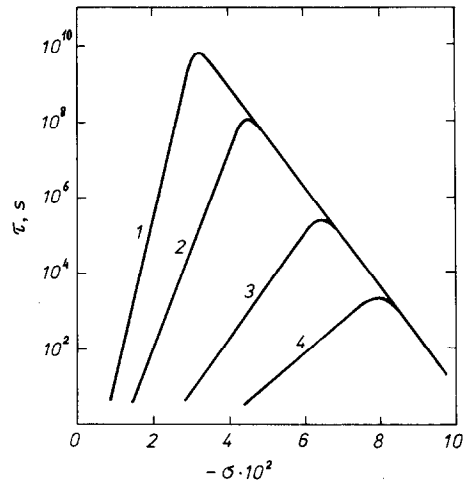

Fig.1. Dependence of the relaxation time τ on the superhelix density σ for a cruciform structure containing n basepairs in its stems: (1) $n = 30$; (2) $n = 20$; (3) $n = 13$; (4) $n = 10$.

The τ value has a sharp maximum at the superhelix density corresponding to the equilibrium transition point between the cruciform structure and the regular double helix. Most interestingly, our calculations predict that at the transition point the relaxation time for any cruciform structure observed under real conditions at room temperature should be very large. The sharp decrease of the rate constant of the cruciform decay is due to the fact that at a high superhelix density any decrease of the cruciform stem is highly unfavourable because it leads to a considerable increase of the superhelix energy. Therefore it is only after an enormous number of unsuccessful attempts that the cruciform structure undergoes the transition into the regular double helix.

The reduced relaxation time at superhelix densities above the equilibrium transition point is the result of a sharp increase of the rate of the cruciform formation. Since the process depends on the formation of the minimal open region, it is highly favoured by negative superhelicity [2]. As a result the rate constant of the cruciform formation is a function of the superhelix density but does not depend on the size of the palindromic region (see eq.(8) and fig.1).

Our data lead to the general conclusion that the formation of a cruciform structure in a palin-

dromic region of any size may be observed, at room temperature, only at a sufficiently large superhelix density, whose absolute value should be higher than 0.05. It explains why the formation of the cruciform structure was observed in giant artificial palindromes only after a considerable supercoiling of their DNA preparations with the help of DNA gyrase [7]. Authors in [7] were the first to point out that the formation of a cruciform structure is a slow process. The first observation of a sharp increase of the relaxation time at the equilibrium transition point between the cruciform structure and the regular double helix was in [6]. Our theoretical results agree very well with their data. Indeed, Lyamichev et al. [6] have shown that the relaxation time at the transition point for the cruciform structure containing 13 bp in its stem is definitely longer than 20 h. Our results suggest 50 h (see curve 3 in fig.1).

Our quantitative estimates correspond to room temperature. If one increases the temperature to, say, 37°C, the dependence of the relaxation time on the superhelix density would be basically unchanged, though the absolute values of the relaxation time would decrease. The quantitative estimates of the temperature dependence could not be quite reliable since we do not have sufficient knowledge of the temperature dependence of some parameters, such as k_f, k_+, ν. A rough estimate predicts that the temperature rise from 20–37°C should decrease the relaxation time by a factor of one hundred.

ACKNOWLEDGEMENTS

We thank Professor Y.S. Lazurkin for discussions and useful critical comments.

REFERENCES

[1] Hsieh, T.S. and Wang, J.C. (1975) Biochemistry 14, 527–535.
[2] Vologodskii, A.V., Lukashin, A.V., Anshelevich, V.V. and Frank-Kamenetskii, M.D. (1979) Nucleic Acids Res. 6, 967–982.
[3] Gellert, M., Mizuuchi, K., O'Dea, M.H., Ohmori, H. and Tomizawa, J. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 35–40.
[4] Lilley, D.M. (1980) Proc. Natl. Acad. Sci. USA 77, 6468–6472.
[5] Panayotatos, N. and Wells, R.D. (1981) Nature 289, 466–470.
[6] Lyamichev, V.I., Panyutin, I.G. and Frank-Kamenetskii, M.D. (1983) FEBS Lett. 153, 298–302.
[7] Mizuuchi, K., Mizuuchi, M. and Gellert, M. (1982) J. Mol. Biol. 156, 229–243.
[8] Vologodskii, A.V. and Frank-Kamenetskii, M.D. (1981) FEBS Lett. 131, 178–180.
[9] Lilley, D.M.J. and Markham, A.F. (1983) EMBO J. 2, 527–533.
[10] Thompson, B.J., Camien, M.N. and Warner, R.C. (1976) Proc. Natl. Acad. Sci. USA 73, 2299–2303.
[11] Vologodskii, A.V. and Frank-Kamenetskii, M.D. (1982) FEBS Lett. 143, 257.
[12] Feller, W. (1957) An Introduction to Probability Theory and Its Applications, Wiley, New York.
[13] Anshelevich, V.V., Vologodskii, A.V., Lukashin, A.V. and Frank-Kamenetskii, M.D. (1983) Biopolymers, in press.
[14] Craig, M.E., Crothers, D.M. and Doty, P. (1971) J. Mol. Biol. 62, 383–401.