TERTIARY STRUCTURE FOR PALINDROMIC REGIONS OF DNA

V. I. LIM and A. L. MAZANOV

Institute of Protein Research, USSR Academy of Sciences, 142292 Poustchino, Moscow Region, USSR

Received 21 October 1977
Revised version received 28 December 1977

1. Introduction

Studies of the primary structure of chromosomal DNA have revealed an interesting structural feature. Palindromic regions (sections of double-helical DNA with an axis of 2-fold rotational symmetry) were found in DNA from eukaryotes [1–6], prokaryotes and viruses [7–13]. The length of palindromic regions varies from a few to several thousand base pairs.

In the present paper the three-dimensional structure (four-stranded helix) for palindromic regions is proposed. This structure differs from the cruciform structure suggested [14]. The possible role of the four-stranded helical structure in the formation of the tertiary structure and functioning of the DNA molecule is discussed. For a preliminary description of our structure see [15].

2. Methods and results

If we pair the Watson-Crick base pairs A–T with T–A and C–G and G–C as shown in fig.1, the base tetrads will be formed, the bases of which lie in one plane. These base tetrads are sufficiently close to each other in the distribution of glycosidic bonds to allow the construction of regular four-stranded helix when the base tetrads are stacked on top of each other. Such a four-stranded helix in which two helices of the Watson-Crick type are related by a dyad axis parallel to the longitudinal axis was built [16].

Four-stranded helical structures in the DNA molecule can be formed from homologous regions of the double-helix. Homologous regions may be represented either by true repeats or by symmetrical sequences of palindromes. The building of four-stranded helices from repeats is impossible without knots. For this reason it is interesting to consider only the symmetrical sequences of palindromes which are able to form four-stranded helices without knots. Figure 2 shows the four-stranded helices which can be built from symmetrical sequences of a palindrome.

Model building with CPK atomic models showed that it is quite possible to build a four-stranded helix from symmetrical sequences even if there is only one canonic base pair between them (n=1, fig.2; see fig.3). This helix is built according to Scheme I (see fig.2). Formation of a four-stranded helix according to Scheme II needs the presence of three or even more base pairs between the symmetrical sequences. The backbone conformation in the four-stranded helix shown in fig.3 is similar to the B-form conformation of DNA. The base pairs are packed along the axis of the helix with a small tilt. The single base pair (in this case–c) at the apex of the four-stranded helix is packed in parallel to the other pairs, the guanine and cytosine of this pair forming a Donohue type system of hydrogen bonds [18]: guanine N1–cytosine 02; guanine N2–cytosine N3. In the case of the A–T pair the following scheme of hydrogen bonds will take place: adenine N6–thymine 02; adenine N1–thymine N3 (for standard atomic numeration see [18] and fig.1). The bent section of the backbone has a sterically allowed combination of rotational isomers which differs from a combination of the B-form rotational isomers only in a Ψ_{-1} angle (see fig.3c). In our case this angle corresponds to a *trans*-isomer

Elsevier/North-Holland Biomedical Press

Fig.1a

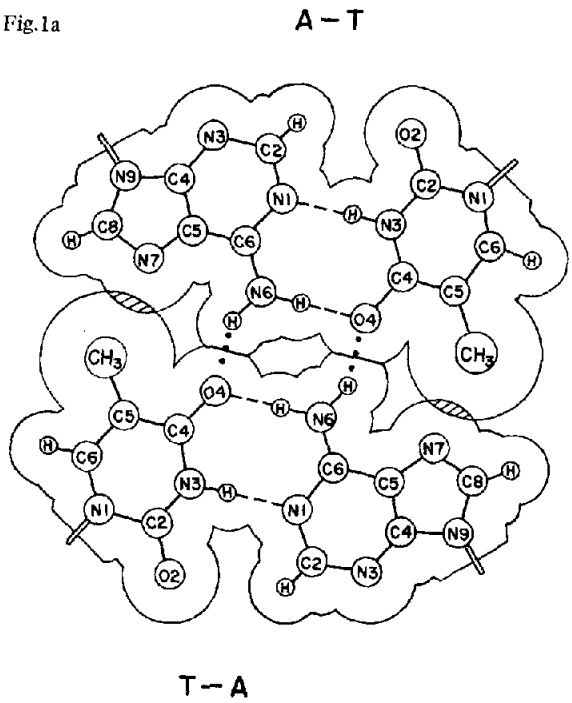

Fig.1b

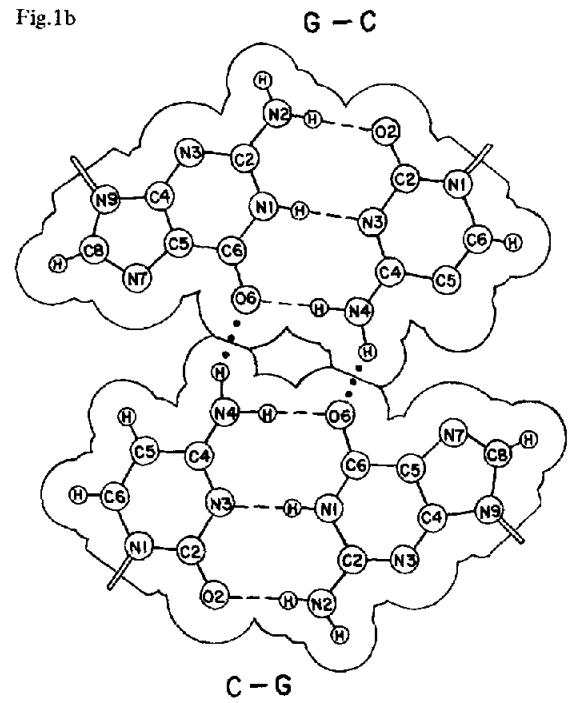

Fig.1. The base tetrads AT–TA and GC–CG which are built by a specific pairing of the Watson-Crick (canonic) base pairs. A–T is paired with T–A (which is an inverted A–T) and G–C is paired with C–G (which is an inverted G–C). In the base tetrads AT–TA there are slight sterical hindrances caused by the methyl groups (see two hatched regions). Hydrogen bonding between canonic pairs is designated as (· · ·); canonic hydrogen bonds in A–T and G–C pairs as (– – –) and glucosidic bonds as (═══).

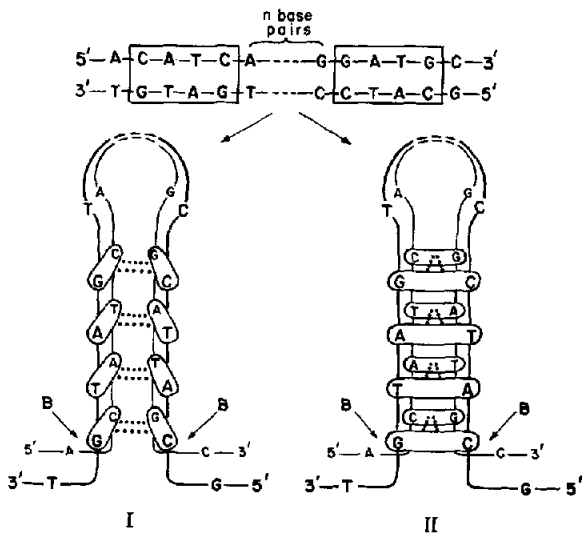

Fig.2. The four-stranded helices formed from the symmetrical sequences of the palindrome. Schemes I and II are different variants of building of a four-stranded helix from symmetrical sequences. $P_u\,P_y$ are the canonic base pairs which form the base tetrads of the four-stranded helix. In Scheme I the bases which form the canonic base pair $P_u\,P_y$ belong to different strands of the DNA molecule. In Scheme II both bases which form the canonic base pair $P_u\,P_y$ belong to one of the two strands of the DNA molecule. Hydrogen bonding between canonic pairs in the base tetrads is shown as (: : :). The number of base pairs between the symmetrical sequences can vary in a wide range ($n \geq 1$). B is the bend of the double helix. In Scheme I the Crick-Klug kinky helix can be used [17].

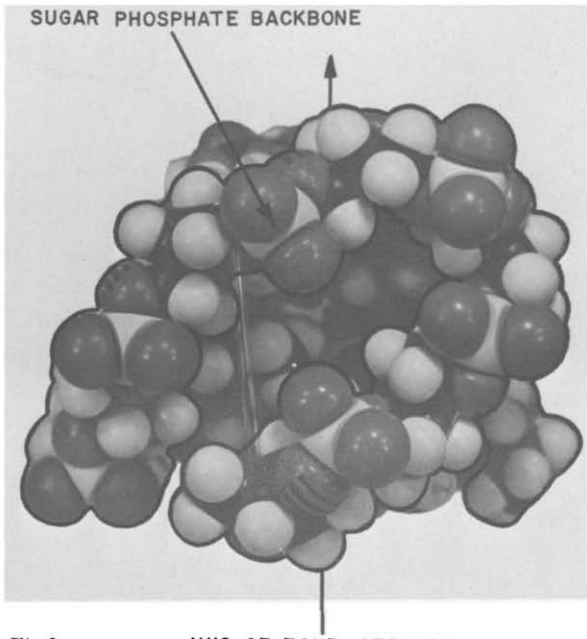

Fig.3a

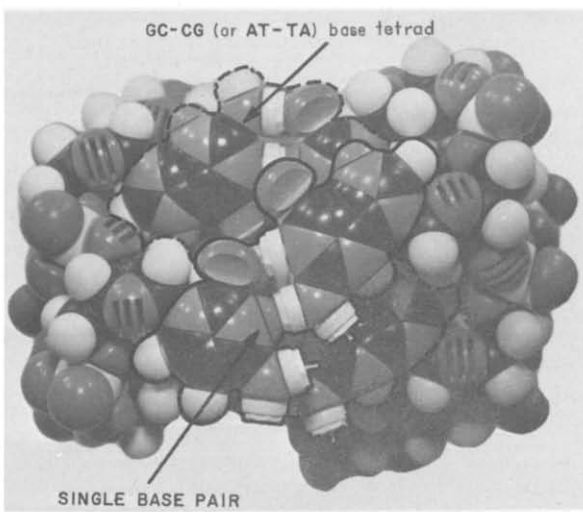

Fig.3b

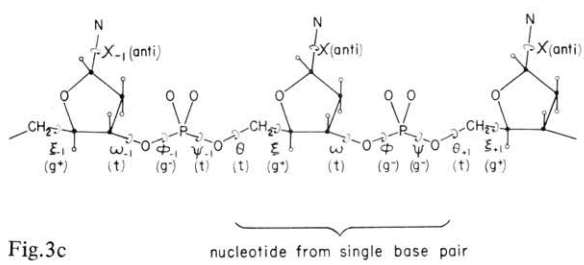

Fig.3c

while in the B-form it corresponds to a *gauche*$^-$-isomer. The *trans*-isomer is observed in monomers [18].

If there are several base pairs between the symmetrical sequences some other possibilities of DNA bending exist which may be relatively easily analyzed with atomic models. The choice of the type of a four-stranded helix (see Schemes I and II in fig.2) depends on the length and the primary structure of the spacer region which is located between symmetrical sequences of the palindrome. This choice may depend also on the proteins interacting with the palindromic region under *in vivo* conditions. All these questions are outside the scope of a short paper. Below we consider the structural properties of the four-stranded helices which are common for both types of these helices. Therefore, we shall take into account both the Schemes I and II on an equal basis and the term 'four-stranded helix' will imply both the Schemes I and II without discriminating between them.

3. Discussion

The proposed structure is in good agreement with the available experimental evidence. The hydrogen bonds shown in fig.1,3 were observed in crystals of bases and nucleosides [16,18,19]. A similar structure has been recently suggested on the basis of X-ray diffraction studies of polyinosic acid molecules which form four-stranded helices [20]. Hairpin folds and thick filaments are often observed in electron micrographs of double-stranded DNA molecules (e.g. [21]). And, finally, a branched structure has been observed for double-stranded DNA in which the bending points apparently occurred in the palindromic regions [22,23].

It is seen in fig.3 that the suggested model of palindromic regions is a very compact structure. There are several axes of 2-fold rotational symmetry. The conformation of the region located between the symmetrical sequences strongly depends on its

Fig.3. The palindromic region three-dimensional structure assembled from CPK models: (a) view from the side; (b) view from the apex, (c) rotational isomers of the bent section of a polynucleotide backbone chain. t, *trans*; g$^-$, *gaucha*$^-$; g$^+$, *gauche*$^+$.

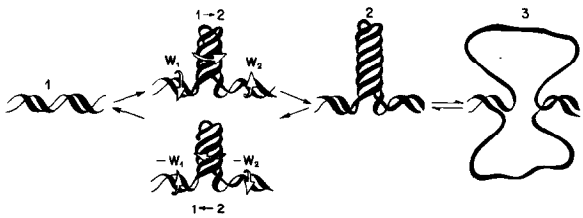

Fig.4. Conformational transitions of the palindromic region: double-helix (1) $\rightleftharpoons$ four-stranded helix (2) $\rightleftharpoons$ denatured (fully unwound) state (3). (1 $\longrightarrow$ 2) and (1 $\longleftarrow$ 2) are intermediate states at transitions (1) $\longrightarrow$ (2) and (2) $\longrightarrow$ (1), respectively, which are accompanied by W_1 and W_2 angle rotations of the palindromic region termini relative to the double helix axis. In both cases $|W_1| + |W_2| \simeq 720 \cdot N$ degrees, where N is the number of four-stranded helix turns. The transitions (2) $\rightleftharpoons$ (3) are not accompanied by a rotation of the palindromic region termini.

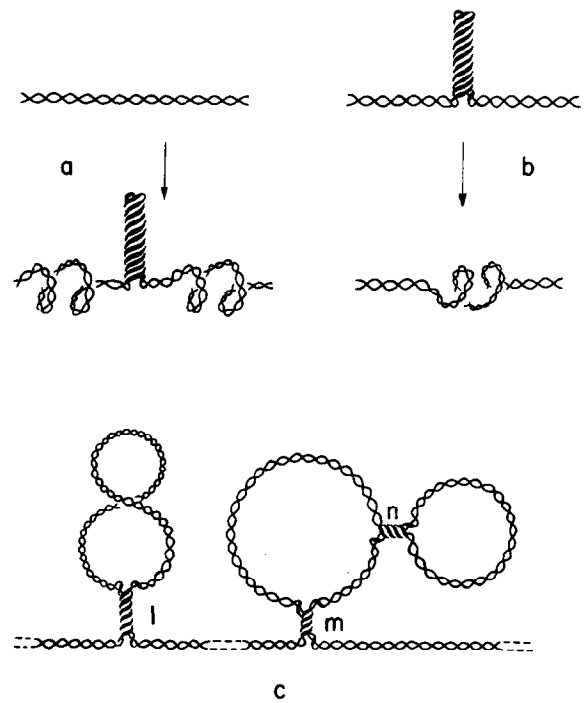

Fig.5. Elements of the DNA molecule tertiary structure which can be formed as a result of the transitions (double-helix) $\rightleftharpoons$ (four-stranded helix). (a) Transition (double-helix) $\longrightarrow$ (four-stranded helix). In this case right-handed superhelical structures will be generated to the left and to the right of the four-stranded helix. (b) Transition (four-stranded helix) $\longrightarrow$ (double-helix). In this case left-handed superhelical structures will be formed. (c) Loop structures are formed with the four-stranded helices l, m, n.

primary structure. Due to these properties the proposed structure seems to be more favourable for interaction with proteins (especially with oligomeric ones) as compared with the usual DNA double-helix.

It follows from the stereochemical analysis that at conformational transitions (double-helix) $\rightleftharpoons$ (four-stranded helix) the palindrome termini will rotate relative to the double-helix axis (fig.4). In the case of covalently closed circular or high molecular linear double-stranded DNA such rotations may promote the formation of superhelical structures from double-helix (fig.5a,b), which could be additionally stabilized by proteins. Consequently, elements of the tertiary structure of DNA (different types of superhelical structures) may be formed by building or breaking of the palindromic four-stranded helices. Loop structures may be formed also if the four-stranded helix is built from the symmetrical sequences which are far from each other along the DNA thread (see fig.5c). In particular these symmetrical sequences may be 'loop recognition sequences' in the process of the formation of loops of eukaryotic chromatin [24].

Thus the following conclusion can be made within the framework of the suggested model: the distribution of palindromes along the DNA thread is the 'primary structure' which codes the three-dimensional structure of a DNA molecule.

Specific interaction of proteins with palindromic regions may lead to large three-dimensional changes in the DNA structure. For example, it may happen that condensation or decondensation of the whole chromosome or its parts in the process of cell fission and transcription is carried out mainly by the formation or breaking of four-stranded helices, in particular, of those enriched with A–T pairs. Such four-stranded helices are easily broken because the base tetrad AT–TA is less stable than the base tetrad GC–CG due to sterical hindrance caused by the methyl group of thymine (see fig.1). Thus the stability of four-stranded helices will vary depending on the content of A–T pairs.

The transition (1) $\rightleftharpoons$ (2) $\rightleftharpoons$ (3) (see fig.4) can play an essential role in the replication and transcription of the DNA molecule. For example, the

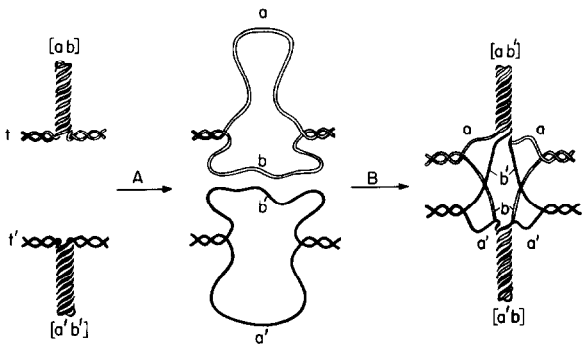

Fig.6. Pairing of two homologous DNA molecules by formation of the intermolecular four-stranded helices. t and t' are homologous palindromic regions (in the four-stranded helical state) belonging to different DNA molecules. Homologous strands belonging to t and t' are designated by the same letter (a is homologous to a' and b to b'). [ab], [a'b'], [ab'], [a'b] are the four-strand helices. A, transition of four-stranded helices [ab] and [a'b'] into the denatured state; B, transition of the denatured palindromic regions into the intermolecular four-stranded helices [ab'] and [a'b]. The intermolecular four-stranded helices [ab'] and [a'b] can be formed according to Schemes I and II (see fig.2).

replication fork can be formed with the transition (2) ⟶ (3). This transition is not accompanied by a rotation of the palindrome termini. Therefore the replication fork can be formed from a separate palindromic region without a change in the secondary and tertiary structure of the rest of the DNA molecule.

It is possible to suggest a simple mechanism of pairing homologous DNA molecules in the process of genetic recombination based on our model. According to this mechanism a pairing of DNA molecules occurs by a formation of the intermolecular four-stranded helices from homologous palindromic regions.

Let us consider two such homologous palindromic regions belonging to different DNA molecules. Suppose that before pairing of DNA molecules these palindromes exist in a four-stranded helical state (see t and t' in fig.6). One of these four-stranded helices is built from the strands a and b while the other from the strands a' and b'. The a strand and the a' strand are identical. The b strand and the b' strand are also identical. Therefore, besides the four-stranded helices [ab] and [a'b'], the intermolecular four-stranded helices [ab'] and [a'b] can be formed from the strands a, b, a', b'. The formation of the intermolecular four-stranded helices [ab'] and [a'b] may proceed in 2 steps (see the transitions A and B in fig.6). The transitions A and B occur in the same way as the transitions (2) ⟸⟹ (3) which are not accompanied by a rotation of the palindrome termini. Therefore DNA molecules will pair with each other without a change in the three-dimensional structure of the interpalindromic regions which constitute the greater proportion of their sequences.

The study of the mechanisms of homologous chromosome pairing in meiosis has shown that only a very small part of the chromosomal DNA, Z-DNA, participates in interchromosomal recognition (e.g., about 0.3% of all the *Lilium* cell DNA) [25,26]. Moreover, it has been shown that the process of DNA molecule pairing occurs in the presence of unwinding proteins [25,26]. Within the framework of our mechanism these proteins can initiate the transition A (see fig.6).

Acknowledgements

The authors thank Professor A. S. Spirin for initiating the work and constant attention, Professor D. G. Knorre and Dr M. A. Gratchev for valuable suggestions.

References

[1] Wilson, D. A. and Thomas, C. A., jr. (1974) J. Mol. Biol. 84, 115–138.
[2] Schmid, C. W., Manning, J. E. and Davidson, N. (1975) Cell 5, 159–172.
[3] Perlman, S., Philips, C. and Bishop, J. O. (1976) Cell 8, 33–42.
[4] Szala, S., Michalska, J., Paterak, H., Bieniek, B. and Chorazy, M. (1977) FEBS Lett. 77, 94–98.
[5] Cavalier-Smith, T. (1976) Nature 262, 255–256.
[6] Engberg, J. and Klenow, H. (1977) Trends Biochem. Sci. 2, 183–185.
[7] Gilbert, W. and Maxam, A. (1973) Proc. Natl. Acad. Sci. USA 70, 3581–3584.
[8] Humayun, Z., Jeffrey, A. and Ptashne, M. (1977) J. Mol. Biol. 112, 265–277.
[9] Subramanian, K. N., Dhar, R. and Weissman, S. M. (1977) J. Biol. Chem. 252, 355–367.
[10] Shen, C.-K. J., Hearst, J. E. (1977) Proc. Natl. Acad. Sci. USA 74, 1363–1364.

[11] Hsu, M.-T. and Jelinek, W. R. (1977) Proc. Natl. Acad. Sci. USA 74, 1631–1634.
[12] Bogdanov, A. A. and Ledneva, R. K. (1975) Molekulyarnaya Biologiya, vol. 5, VINITI, Moscow.
[13] Sobell, H. M. (1976) Ann. Rev. Bioph. Bioeng. 5, 307–335.
[14] Gierer, A. (1966) Nature 212, 1480–1481.
[15] Lim, V. I. and Mazanov, A. L. (1976) Dokl. Akad. Nauk SSSR 231, 492–494.
[16] McGavin, S. (1971) J. Mol. Biol. 55, 293–298.
[17] Crick, F. H. C. and Klug, A. (1975) Nature 255, 530–533.
[18] Arnott, S. and Hukins, D. W. L. (1973) J. Mol. Biol. 81, 93–105.
[19] Voet, D. and Rich, A. (1970) Progr. Nucl. Acids Res.; Mol. Biol. 10, 183–265.
[20] Arnott, S., Chandrasekaran, R. and Marttila, C. M. (1974) Biochem. J. 141, 537–543.
[21] Vollenweider, H. J., Koller, Th., Parello, J. and Sogo, J. M. (1976) Proc. Natl. Acad. Sci. USA 73, 4125–4129.
[22] Campbell, A. M. and Eason, R. (1975) FEBS Lett. 55, 212–215.
[23] Campbell, A. M. (1976) Biochem. J. 159, 615–620.
[24] Comings, D. E. and Okada, T. A. (1974) Cold Spring Harbor Symp. Quant. Biol. 38, 145–153.
[25] Radding, C. M. (1973) Ann. Rev. Genet. 7, 87–111.
[26] Stern, H. and Hotta, Y. (1973) Ann. Rev. Genet. 7, 37–66.