M. D. Frank-Kamenetskiĭ. *Topology and structural transitions in DNA.* Perhaps the most important fact which has been established about DNA since the work of Watson and Crick is that the biologically functioning form is a ring form, rather than a linear form. We thus see that topology plays an important role.

A distinction is made between two levels of the topology of the DNA molecule.[1] First, the DNA ring may form no nodes, or it may form a node of some type or other. Second, each of the complementary chains is closed on itself; as a result, these chains form, in the mathematical sense, a high-order linkage. A quantitative measure of the extent of linkage is the linking number Lk, which is defined as the number of times a single chain pierces a surface covering another chain.

The linking number and the nature of the node are topological invariants of DNA since they cannot vary, no matter what deformations or structural changes occur in the molecule without involving a rupture of bonds. The existence of these topological invariants has several important consequences. In the first place, since the molecules at nodes cannot replicate, a cell must be able to "untie" DNA molecules at nodes.[2] As it turns out, this is done by some special enzymes called "topoisomerases." These enzymes cut the DNA double helix, pull the part of the molecule which is remote along the chain through the resulting breach, and then repair the breach. The topoisomerases also change the DNA linking number Lk.

The existence of the topological invariant Lk leads to the important concept of superspiralling of DNA. The num-

ber of superturns is found as the difference $\Delta Lk = Lk - (N/\gamma)$, where N is the number of pairs of nucleotides in the DNA, and γ is the number of pairs of nucleotides per turn of the double helix under the given external conditions. The specific magnitude of the superspiralling—i.e., the density of superturns—is found as $\sigma = 10\Delta Lk/N$. The DNA superhelix is in a stressed state; the superspiralling energy $G = 11R\,TN\sigma^2$ is stored in both the bends and the axial twisting of the double helix.[3]

There is the fundamental fact that DNA functioning in a cell is at all times not only closed but also in a negatively superspiralled state. The values of σ for all types of DNA are approixmately the same, -0.05. This fact means that a special mechanism operates in a cell to keep the DNA in a superspiralled state. It turns out that this role is played by enzymes from the topoisomerase class: DNA-gyrase. These enzymes use ATP energy to put the relaxed, nonsuperspiralled DNA in a negatively superspiralled state.

The discovery of DNA-gyrase dispelled the remaining doubts that the superspiralling state was very important to the functioning of DNA. Many biological experiments have shown that superspiralling is necessary for replication, recombination, and the correct regulation of transcription. In order to determine the reasons for such a major biological role of superspiralling, it was necessary to determine how the structure of superspiralled DNA differs from that of the relaxed molecule, i.e., from that of the B form of DNA, discovered by Watson and Crick.

With this goal in mind, an effort was undertaken to determine the effect of superspiralling on the structure of DNA in Yu. S. Lazurkin's laboratory at the Institute of Molecular Genetics, Academy of Sciences of the USSR. The question was first subjected to a theoretical analysis.[4] By that time, the spiral-ball transition in DNA had already been studied in some detail. It was thus possible to determine theoretically whether uncovered regions could form during superspiralling. After the corresponding theory was derived, and calculations were carried out for the DNA molecule, whose sequences were not known at the time, we reached the conclusion that at the physiological values of the superspiralling density uncovered regions do not form, although the probability for uncovering increases significantly in comparison with that in the relaxed double helix. We then incorporated in the theory the possible formation of cross-shaped structures in regions in which the sequence has twofold symmetry axis (palindromes). It turned out that in the real sequences one encounters some rather long palindromes, which may convert into cross-shaped structures at the physiological stresses prevailing during superspiralling.

The theoretical conclusions were quickly confirmed experimentally. The most successful method for studying the formation of alternative structures (structures other than the regular B form—crosses, etc., which we will be discussing below) turned out to be the method of two-dimensional gel electrophoresis.[5]

This method can be summarized as follows. If a set of topoisomers, i.e., DNA molecules which are chemically identical but which differ in the value of Lk, are placed in a gel (a polymer network saturated with solvent), and if a static electric field is applied, the topoisomers can be separated. As the superspiralling increases, the molecules become progressively more compact, and they move more rapidly in the electric field. However, if a structural transition (to a cross, for example) occurs in highly superspiralled topoisomers, these topoisomers began to move more slowly than they are meant to do, since the uncoiling of the double helix in the region where the transition occurred involves a partial removal of the superspiralling stress. As a result, we find the complicated pattern of bands shown at the left in Fig. 1. To determine what is happening, the electric field is reversed. Now the molecules move from left to right. We add some special additives to the gel to reduce the superspiralling stress by a certain amount. As a result, the existence of the alternative structures becomes unfavorable, and for the second direction of the field all the topoisomers move in accordance with the number of superturns. The resulting two-dimensional gel electrophoresis pattern is shown in Fig. 1. The formation of an alternative structure can be clearly seen from the discontinuity in the mobility of the topoisomers for the first field direction.

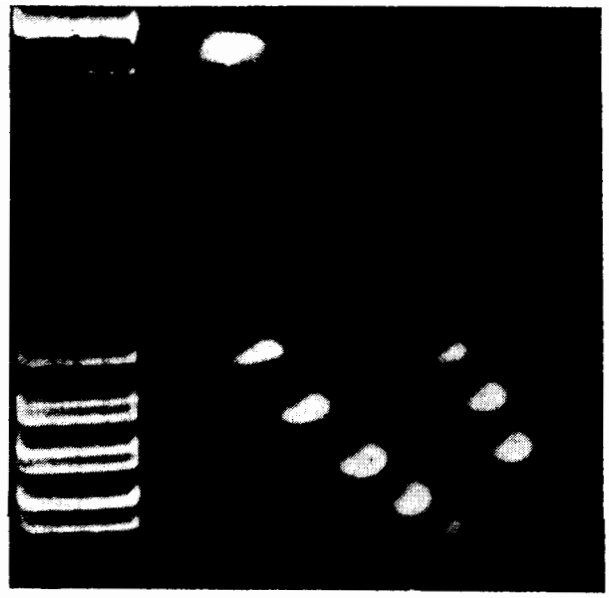

FIG. 1.

In the late 1970s, genetic engineering presented researchers involved with the structure of DNA some completely new opportunities. Before this time we had been forced to content ourselves with those DNA preparations which could be extracted from cells and viruses. We did not know the sequences of nucleotides in them. We now became able to obtain from the genetic engineers some specially "cut" DNA molecules with sequences of nucleotides which were known exactly and which carried some insertion or other in whose structure we were interested. This new opportunity could not have failed to lead to new discoveries regarding the structure of DNA, and these discoveries were not long in coming.

In 1979, A. Rich *et al.* (Massachusetts Institute of Technology) reported the discovery of a new form of DNA: the Z form. In many regards, the Z form differs from the classical B form, although it does retain the Watson-Crick complementarity principle. The primary distinction is that the double helix in it is left-handed, rather than right-handed.[6] This result quickly led to the suggestion that the formation of the Z form should strongly promote negative superspiralling. It was shown by the method of two-dimensional gel electrophoresis that insertions in which purine and pyrimidine neucleotides alternate, $(GC)_n \cdot (GC)_n$ and $(GT)_n \cdot (AC)_n$, actually convert into the Z form in the course of negative superspiralling. It was the study of the insertions in ring DNA by two-dimensional gel electrophoresis which made it possible—along with the corresponding theory—to determine the basic energy parameters of the B–Z transitions.[7]

We recently took up the study of the structure which arises under the influence of negative superspiralling in DNA regions in which there are only purine nucleotides in one chain and only pyrimidine nucleotides in the other.[8] For this purpose, we inserted a region containing the sequence $(GA)_{16} \cdot (TC)_{16}$ in a ring DNA convenient for study by two-dimensional gel electrophoresis. The two-dimensional gel electrophoresis revealed a structural transition in this region. A distinctive feature of this structural transition, which sets it apart from the transitions to a cross and to the Z form, which had been studied previously, is a strong dependence on the pH of the medium. As a result, the sequence studied acquires a completely different structure stabilized by protons. We called this new DNA structure the H form.

What is the structure of the H form? At this point, we can only guess. Our data do provide extremely definite quantitative characteristics of the H form, which impose some severe restrictions on possible models. Specifically, the strands in the H form are not interwoven, as in the case of crosses, and there are four pairs of nucleotides per protonization site. The structure which seems the most probable to us is shown in Fig. 2. The basic structural element is a ternary complex including a normal Watson-Crick double helix, in whose large groove there is a pyrimidine chain which forms Hoogsteen pairs with a purine chain. In these pairs, the cytosine is protonated. There are data which show that ternary complexes of this sort are formed by model polynucleotides at acidic values of the pH (Ref. 9).

We thus now know that negative superspiralling can lead to the formation of at least three types of alternative structures in DNA: crosses in palindrome regions, Z forms in alternating purine-pyrimidine sequences, and H forms in homopurine-homopyrimidine sequences. We will soon learn whether these structures exhaust the set of alternative structures whose formation results from topological constraints. An active effort is also being made to resolve the question of a possible biological role of the alternative structures.

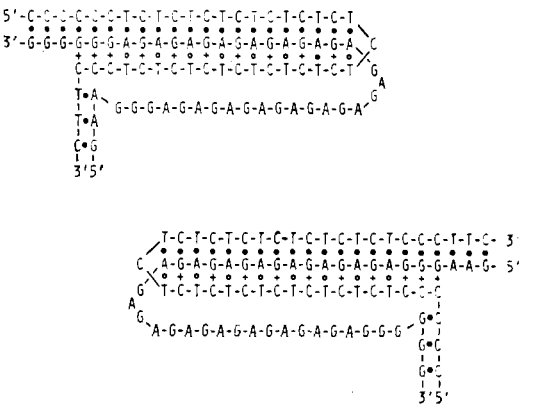

FIG. 2.

[1]M. D. Frank-Kamenetskiĭ and A. V. Vologodskiĭ, Usp. Fiz. Nauk **134**, 641 (1981) [Sov. Phys. Usp. **24**, 679 (1981)].
[2]M. D. Frank-Kamenetskiĭ, A. V. Lukashin, and A. V. Vologodskiĭ, Nature **258**, 398 (1975).
[3]A. V. Vologodskiĭ, V. V. Anshelevich, A. V. Lukashin, and M. D. Frank-Kamenetskiĭ, Nature **280**, 294 (1979); M. D. Frank-Kamenetskiĭ, A. V. Lukashin, V. V. Anshelevich, and A. V. Vologodskiĭ, J. Biomol. Struct. Dyn. **2**, 1005 (1985).
[4]A. V. Vologodskiĭ, A. V. Lukashin, V. V. Anshelevich, and M. D. Frank-Kamenetskiĭ, Nucl. Acids Res. **6**, 967 (1979); V. V. Anshelevich, A. V. Vologodskiĭ, A. V. Lukashin, and M. D. Frank-Kamenetskiĭ, Biopolymers **18**, 2733 (1979); A. V. Vologodskiĭ and M. D. Frank-Kamenetskiĭ, FEBS Lett. **143**, 257 (1982).
[5]J. C. Wang, L. J. Peck, and K. Becherer, Cold Spring Harbor Symposium on *Quantitative Biology*, Vol. 47, 1983, p. 85; V. I. Lyamichev, I. G. Panyutin, and M. D. Frank-Kamenetskiĭ, FEBS Lett. **153**, 298 (1983).
[6]A. Rich, A. Nordheim, and A. H. Wang, Ann. Rev. Biochem. **53**, 791 (1984).
[7]L. J. Peck and J. C. Wang, Proc. Nat. Acad. Sci. USA., **80**, 6206 (1983); M. D. Frank-Kamenetskiĭ and A. V. Vologodskiĭ, Nature **307**, 481 (1984); A. V. Vologodskiĭ and M. D. Frank-Kamenetskiĭ, J. Biomol. Struct. Dyn. **1**, 1325 (1984).
[8]V. I. Lyamichev, S. M. Mirkin, and M. D. Frank-Kamenetskiĭ, J. Biomol. Struct. Dyn. **2**, 327 (1985); **3**, 667 (1986); V. I. Lyamichev, S. M. Mirkin, and M. D. Frank-Kamenetskiĭ, Biopolimery i Kletka **2**, 115 (1986).
[9]J. S. Lee, D. A. Johnson, and A. R. Morgan, Nucl. Acids Res. **6**, 3073 (1979).

Translated by Dave Parsons