

DNA structure: Yet another avatar?

Manju Bansal

Everytime the story of DNA structure seems to reach a conclusion, it bounces back to centre stage by appearing in yet another incarnation. The latest avatar to manifest itself is a stretched and overwound form of DNA reported recently by a French group¹, working with single DNA molecules. When a moderately large stretching force (of about 3 pico Newtons) is applied, the DNA molecule apparently becomes highly twisted and extended, but even more amazingly it takes up an 'inside-out structure' in which the phosphodiester chain is on the inside and the bases are exposed (Figure 1 *a*). Precisely such a structure was proposed in 1953 by Pauling and Corey² (albeit with three strands), making use of 'the general principles of molecular structure', which they had applied with such spectacular success in predicting the α -helix and β -structure for proteins³. While they were convinced that the DNA structure was a helix, they did not make use of the already available chemical data of Chargaff, which demonstrated that although the base composition of DNA varies widely, certain bases were always present in equal numbers (viz. Adenine = Thymine and Guanine = Cytosine), clearly suggesting some kind of pairing between the bases. So in their model building studies, when they considered the question of arranging more than one

chain about the helix axis, they made the wrong choice. As shown schematically in Figure 2, the helix axis can be located either to the left or to the right of the polynucleotide backbone. An axis to the right of the chain, as shown in the top right of Figure 2, results in a structure with phosphates near the helix axis and the bases farther away from the axis and 'fanning' out. Watson and Crick on the other hand placed the helix axis on the left side of the chain, resulting in the arrangement shown in the top left of Figure 2, with the phosphates on the outside and the bases inside. This arrangement for a two chain molecule readily explains Chargaff's data, if a specific base-base interaction is postulated between the bases A and T, as also between G and C. The two chains can be either parallel or antiparallel to each other. The antiparallel arrangement was chosen, rather arbitrarily, since it was consistent with the pseudo-two-fold symmetry of the basepairs, but has subsequently been confirmed by biochemical studies. Thus, the canonical Watson-Crick structure for hydrated DNA, which gives the B-form X-ray fibre pattern, a two-strand right-handed helix, with ten planar A:T or G:C basepairs of nearly equivalent shape and size arranged in a spiral arrangement, almost perpendicular to the fibre axis, came into existence⁴. The currently accepted

detailed structure⁵ for the B-form of DNA retains the essential features of the original Watson-Crick model and is shown in Figure 1 *b*. Since this structure was observed when the molecule was fully hydrated and it also immediately suggested a possible copying mechanism for the genetic material and the semiconservative replication of DNA, it was readily accepted as being the 'biologically relevant structure' for DNA. This corollary was repeatedly emphasized, since even before the structure of DNA was known, X-ray fibre diffraction and spectroscopic data had clearly shown that it is quite a pliable molecule, readily undergoing interconversion between various forms (arbitrarily classified as A, B, C, etc.) when ionic or humidity conditions are changed. All these related structures were considered as relatively unimportant minor variants, or distorted versions, of the B-form, which occupied the centre-stage for nearly a quarter century after it was first proposed.

Once the Watson-Crick model for DNA became an accepted fact, the double helix came to be regarded merely as a safe storage device for the genetic information encoded in the nucleotide sequence, that could only be accessed after unwinding of the helix. Hence the double stranded DNA molecule was treated as being a very stable, intrinsi-

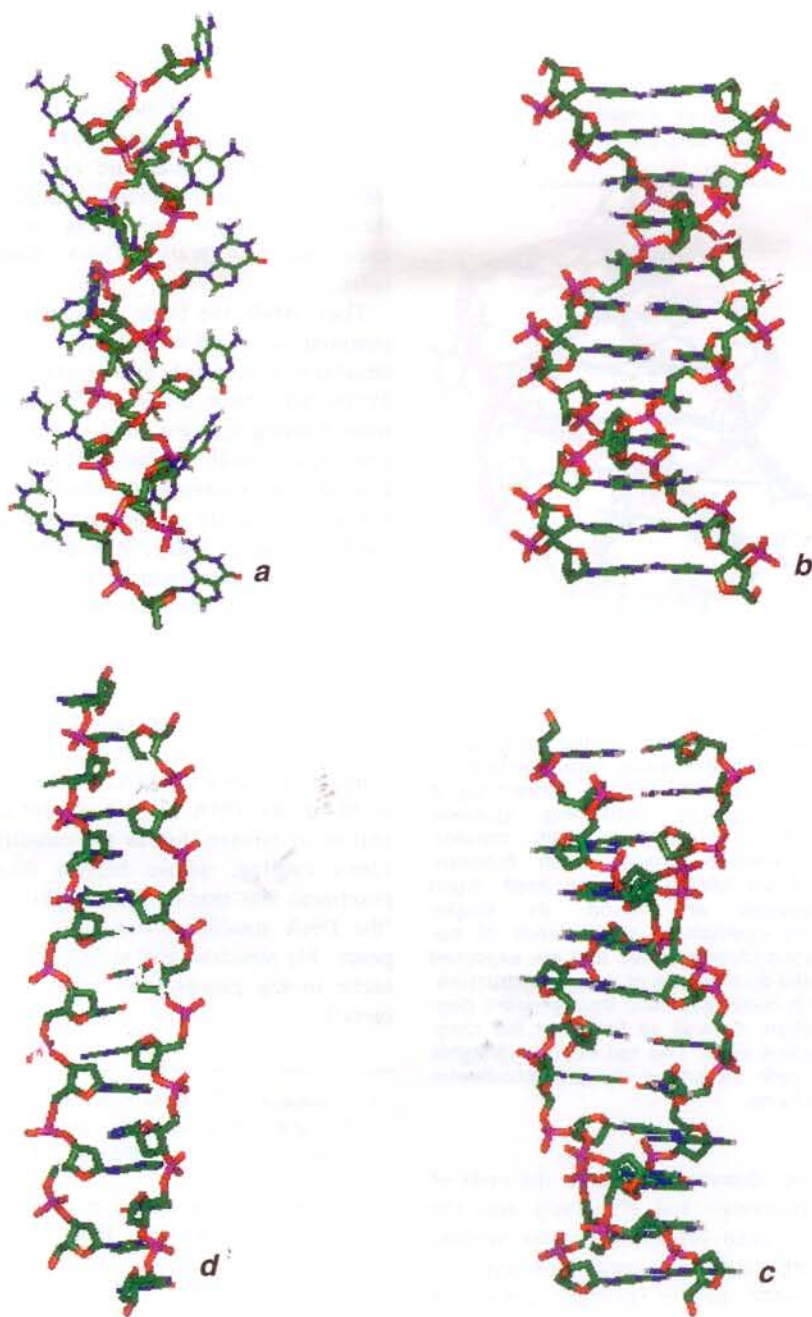


Figure 1. Line diagrams showing four different two-stranded structures for DNA. *a*, The Pauling-like anti-parallel duplex structure (P-DNA) with phosphate backbone in the centre and bases exposed on the outside; *b*, The canonical, Watson-Crick B-DNA structure, which apparently undergoes transformation to P-DNA, on stretching without writhing; *c*, Z-DNA, a left-handed zig-zag structure, which has been observed for alternating purine-pyrimidine sequences, under conditions of high salt; *d*, An alternative model to P-DNA, which has also been proposed as a structure for the single-stranded circular DNA in the Pfl filamentous phage, with bases being in the centre and intercalated. A 12-mer fragment of basepaired duplex DNA is shown for B and Z DNA structures (Figures *b* and *c*) while an 8-mer fragment is shown for the two stretched structures (Figures *a* and *d*).

cally uniform entity, requiring strong interaction with proteins or some other ligand, in order to undergo any struc-

tural change in the cellular environment. This cozy 'uniform B-DNA centric' world received a severe jolt, with the

arrival on the scene of single crystal X-ray structures of DNA oligomers. A dramatic sequence-dependent structural microheterogeneity was observed in the crystal structures of DNA oligomers, with the alternating sequences TATA (ref. 6) and (CG)₃ (ref. 7). The latter, in particular, had only the Watson-Crick G:C basepairs as a common feature with the B-DNA structure, otherwise it was completely different – first of all, it had a left-handed helical twist and secondly, it had quite different values for the torsion angles, defining the orientation about the various single bonds, in the guanine and cytosine nucleotides. This gives the phosphodiester backbone a distinctly zig-zag appearance (Figure 1*c*) and led to the structure being dubbed the Z-form DNA: This structure was solved using very high resolution X-ray data, so it could not be ignored, which was the fate of many theoretically model-built structures, but since it was only observed under conditions of high salt, its biological relevance remained questionable and it has also been relegated to the domain of 'interesting curiosities'. However, its formation may be facilitated under the stimulus of torsional stress induced by negative supercoiling.

The faith in the ubiquitous nature of B-DNA was restored when the next crystal structure⁸, for a 12-mer fragment, turned out to be very similar to the fibre B-form, even though it had alternating CGCG sequences, flanking an AATT tetramer at both ends. However, a closer look at this structure opened another Pandora's box, since it showed considerable local variability at each dinucleotide step and appeared to be slightly curved overall. This and subsequent oligonucleotide crystal structures have clearly blasted the myth of 'uniform B-DNA' structure and in fact sequence-dependent variability is now routinely invoked to explain phenomena as diverse as intrinsic bending in mini-circles of kinetoplast DNA, to recognition of promoter regions and replication origins in genomic DNA, by polymerases. It has now become common to talk about A, B or C type dinucleotide steps rather than assigning a particular type to the whole structure!

The next development in the DNA structure story came with the availability of large genomic sequences. These

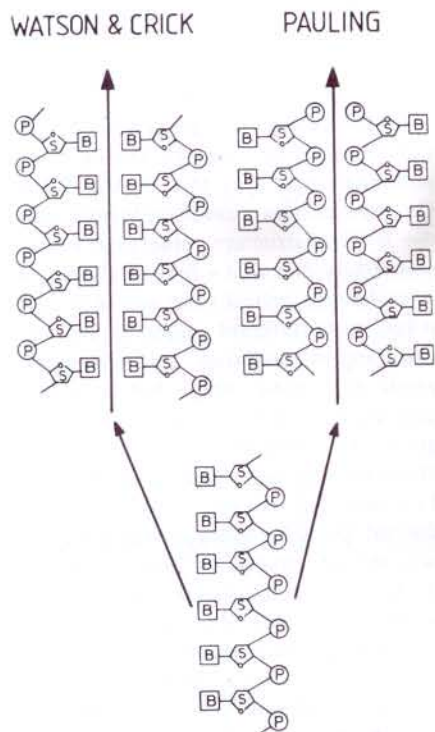


Figure 2. A schematic diagram showing two possible arrangements of polynucleotide chains about a common helix axis. The symbols B, S and P represent the bases, sugars and phosphates, respectively. In the scheme proposed by Pauling, the sugar phosphate backbone occurs inside and a third chain (not shown here) was also present, thus forming a triple-helical structure. In the recently proposed P-DNA, only two chains are involved and each of the chains is highly stretched. In the Watson-Crick scheme also two chains are present, but the bases occur on the inside and can be involved in inter-chain hydrogen bonds.

indicated the presence of several repeat motifs, with sequences which had earlier been shown to take up a variety of unusual structures, for example a single chain helix for poly (C), three chain or triplex structure for [poly (A) + poly 2(U)] and a parallel four chain assembly or quadruplex structure for poly (G). One by one, they have all come out of the cold-storage and found their rightful place in the pantheon of DNA structures. A variant of the G-quadruplex structure, formed by a single guanine-rich fragment folding on itself (Figure 3) has, in particular, evoked great interest since it has been postulated to occur

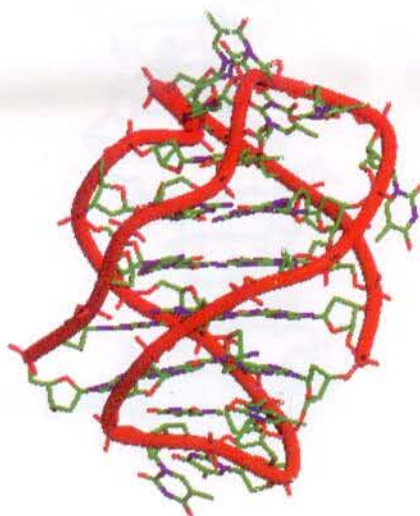


Figure 3. A four-stranded quadruplex structure, stabilized by hydrogen-bonded guanine tetrads, can be formed by a DNA fragment containing guanine stretches, interspersed with thymine and adenine bases which facilitate the chain folding back on itself. Such sequences are found, as single-strand overhangs, at the ends of eukaryotic chromosomes and are expected to take up this type of compact structure, which could seal and thus prevent degradation, as well as fusion of the chromosome ends. The red ribbon highlights the path traced by the phosphodiester backbone.

in the telomeric regions at the ends of chromosomes and effectively seal the ends⁹. Also the loss of these regions during successive replication cycle is attributed as the primary cause for 'ageing'. Retain the telomeres and remain 'young forever' seems to be the current thinking.

To return to the beginning, the Pauling-like structure with exposed bases (currently termed as P-DNA)¹ for a highly stretched and overwound form of DNA, with rise per residue along the helix axis of about 5.85 Å and 2.62 units per turn, as compared to 3.4 Å rise and 10 units per turn for B-DNA, had also been proposed as a model for a circular single-stranded DNA in the Pfl filamentous phage¹⁰. An alternative structure for this DNA¹¹, packaged in-

side the virus protein coat, is shown in Figure 1d. In this structure the phosphate backbone remains on the outside, with its negative charge being neutralized by the basic groups in the coat protein assembly, but the bases in the two antiparallel strands are merely intercalated, rather than basepaired through hydrogen bonds, as in the complementary Watson-Crick double helix.

Thus, while the basic principles and essential structural elements for protein structure, elucidated during the early 1950s, viz. the α -helix and β -structure from Pauling's group³ and the coiled-coil triple-helical structure for collagen from Ramachandran's laboratory¹² have remained virtually unchanged and unchallenged even today, with only their permutations, combinations and linking regions varying, in the more than 300 unique sequence protein structures currently known, the structure for a single DNA molecule seems to be able to adapt itself to its environment by twisting, turning and stretching into completely different 'avatars'. The last word in the DNA story has probably still to be written, but in the meantime Linus Pauling, whose biggest disappointment was that he did not discover 'the DNA structure', can now rest in peace, his structure has at last found a niche in the pantheon of DNA structures!!

1. Allemand, J. F., Bensimon, D., Lavery, R. and Croquette, V., *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 14152-14157.
2. Pauling, L. and Corey, R. B., *Proc. Natl. Acad. Sci. USA*, 1953, **39**, 84-97.
3. Pauling, L., Corey, R. B. and Branson, H. R., *Proc. Natl. Acad. Sci. USA*, 1951, **37**, 205-211.
4. Watson, J. and Crick, F. H. C., *Nature*, 1953, **171**, 737-738.
5. Chandrasekaran, R. and Arnott, S., *J. Biomol. Struct. Dyn.*, 1996, **13**, 1015-1027.
6. Viswamitra, M. A., Kennard, O., Jones, P. G., Sheldrick, G. M., Salisbury, S., Falvello, L. and Shakked, Z., *Nature*, 1978, **273**, 687-688.
7. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A., *Nature*, 1979, **282**, 680-686.
8. Dickerson, R. E. and Drew, H. R., *J. Mol. Biol.*, 1981, **149**, 761-786.

9. Zakian, V. A., *Science*, 1995, **270**, 1601-1607.
10. Liu, D. J. and Day, L. A., *Science*, 1994, **265**, 671-674.
11. Marvin, D. A., Nave, C., Bansal, M., Hale, R. D. and Salje, E. K. H., *Phase Transitions*, 1992, **39**, 45-80.

12. Ramachandran, G. N. and Kartha, G., *Nature*, 1955, **176**, 593-595.

Mr Anirban Ghosh for assistance in making the figures.

ACKNOWLEDGEMENTS. I thank Drs R. Lavery and D. A. Marvin for providing the coordinates of the stretched DNA structure and

Manju Bansal is in the Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India.