Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A glossary of DNA structures from A to Z

Anirban Ghosh[a,b]† and Manju Bansal[a,b]*

[a]Institute of Bioinformatics and Applied Biotechnology, ITPL, Bangalore 560066, India, and [b]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

† Present address: Infosys Technologies Limited, Electronics City, Hosur Road, Bangalore 561229, India.

Correspondence e-mail: mbansal@ibab.ac.in, mb@mbu.iisc.ernet.in

Manju Bansal obtained her PhD degree from the Molecular Biophysics Unit, Indian Institute of Science, Bangalore in 1977, for her work on the structure of the fibrous protein collagen. She is currently a Professor at the same Institute, as well as Director of Institute of Bioinformatics and Applied Biotechnology, Bangalore. Her research interests include computer modelling and structural analysis of proteins and nucleic acids. Sequence dependent variability of DNA structure and its role in transcription is a particular area of interest. Anirban Ghosh is a consultant in Life Sciences Practice at Infosys Technologies Limited. His research involves model building studies and database analysis to identify sequence dependent structure variation in free and bound forms of DNA. His technological interests include biological and chemical data analytics, and integrated informatics for drug discovery. Anirban received a PhD in Molecular Biophysics from the Indian Institute of Science, Bangalore.

Received 4 December 2002
Accepted 10 February 2003

The right-handed double-helical Watson–Crick model for B-form DNA is the most commonly known DNA structure. In addition to this classic structure, several other forms of DNA have been observed and it is clear that the DNA molecule can assume different structures depending on the base sequence and environment. The various forms of DNA have been identified as A, B, C *etc.* In fact, a detailed inspection of the literature reveals that only the letters F, Q, U, V and Y are now available to describe any new DNA structure that may appear in the future. It is also apparent that it may be more relevant to talk about the A, B or C type dinucleotide steps, since several recent structures show mixtures of various different geometries and a careful analysis is essential before identifying it as a 'new structure'. This review provides a glossary of currently identified DNA structures and is quite timely as it outlines the present understanding of DNA structure exactly 50 years after the original discovery of DNA structure by Watson and Crick.

1. Introduction

Watson and Crick's postulation in 1953 of a double-helical structure for DNA (shown schematically in Fig. 1*a*), with the two strands coiling around a common axis and being linked together by a specific hydrogen-bond scheme, heralded a revolution in our understanding of biology at the molecular level (Watson & Crick, 1953). A striking feature of the structure was the pairing proposed between the purine and pyrimidine bases (Fig. 1*b*): *viz.* adenine with thymine and guanine with cytosine, which led the authors to conclude with the sentence

> It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

This line, which embodied the seminal idea about how the DNA molecule could code and transmit genetic information, aroused the maximum interest. However, the structure itself (often referred to as the B-DNA structure, by association with the corresponding X-ray fibre pattern) has also attained an almost iconic status and was for some time regarded as being the only biologically relevant structure, though Watson and Crick had themselves pointed out that the structure could readily undergo changes depending on the environment.

The canonical Watson–Crick DNA structure is a right-handed double helix with ten nucleotides per turn, separated by a 3.4 Å translation along the helix axis, in each of the two chains and the two chains aligned in mutually antiparallel orientations (Fig. 1*a*). Subsequent studies on synthetic polynucleotides, as well as naturally occurring DNA sequences

with certain repeat patterns, have established that the DNA molecule could have structural polymorphism that is important for its biological function. With the availability of the human genome sequence, it is expected that the three billion base pairs in the genome will exhibit a variety of structural polymorphs of DNA. Initially, the DNA molecules were thought to interconvert between only two well defined right-handed double-helical structures, *viz.* A and B. Now it is clear that the DNA molecule exhibits a chameleon-like property of adapting itself to its environment by twisting, turning and stretching into completely different 'avatars', leading to a pantheon of DNA structures (Bansal, 1999). Several of these structural polymorphs of DNA have been experimentally observed using X-ray diffraction, NMR or other spectroscopic studies and are not confined to the Watson–Crick-type structure.

The structural differences range from minor variations in some local parameters (illustrated in Fig. 2) of the Watson–Crick paired duplex structures to structures that are completely different, even in their essential features such as handedness, base-pairing scheme or number of strands. This wide range of structural variability is possible owing to the inherent conformational flexibility of the polynucleotide backbone, including the puckering of the five-member sugar ring and rotation about the glycosyl bond (Fig. 3). Also, in addition to the three-hydrogen-bonded GC and two-hydrogen-bonded A·T base pairs proposed by Watson and Crick (Fig. 1*b*), there are 27 other distinct possibilities for forming at least two hydrogen bonds between any two bases (Saenger, 1983). Of these, nine can form between purine–pyrimidine base pairs, seven between homo-purine base pairs, four between hetero-purine base pairs and seven between pyrimidine–pyrimidine base pairs. One of the alternative hydrogen-bond schemes, known as the Hoogsteen base pairing (Hoogsteen, 1959) and often observed between adenine and thymine bases, is also shown in Fig. 1(*b*). The keto/enol and amino/imino tautomerism of some purine and pyrimidine bases, as well as protonation at low pH, can also lead to unusual DNA duplex, triplex or quadruplex structures (Lavery & Zakrzewska, 1999). There has been a practice of naming or identifying each DNA structure by associating a letter of the English alphabet with it and occasionally this has led to some confusion owing to two groups assigning the same letter to quite different structures. Hence, it seems worthwhile to record the currently used nomenclature for various forms of DNA in alphabetical order, as well as to describe some other structural forms that do not fit into this nomenclature.

2. The different forms of DNA with a one-letter name

The term A-DNA was first used to describe the X-ray pattern recorded for fibres of the sodium salt of calf thymus DNA under conditions of low humidity and was subsequently used to describe the structure which gives rise to this pattern (Franklin & Gosling, 1953; Fuller *et al.*, 1965). It is a right-handed double helix with 11 residues per turn; the base pairs are considerably displaced from the helix axis as well as being

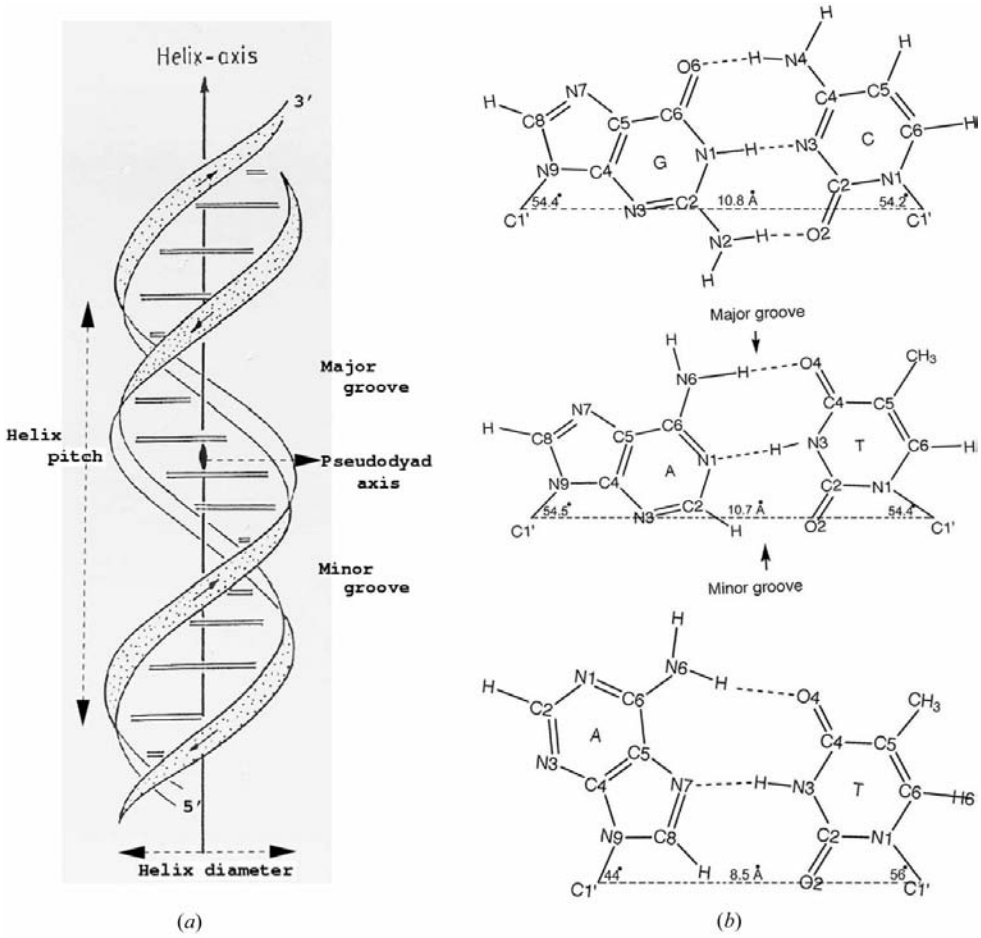

Figure 1
(*a*) A schematic diagram of the Watson–Crick double helix. The base pairs are represented by horizontal bars and the sugar–phosphate backbones of the two chains, related by a twofold rotation axis perpendicular to the helix, are represented by ribbons running in opposite directions. The 5′ and 3′ ends are labelled for the ascending strand. The helix axis, its pitch and diameter and the major and the minor grooves have been identified. (*b*) Base pairs G·C and A·T with Watson–Crick and A·T with Hoogsteen hydrogen-bond schemes are shown as line diagrams. The base atoms are numbered according to the standard nomenclature and the hydrogen bonds between them are represented by dotted lines. The C1′—C1′ distance and the C1′—C1′—N1/N9 angles are indicated in each case. The minor-groove and major-groove sides of the Watson–Crick base pair is indicated in the case of the A·T base pair.

inclined to it (Fig. 4). The B-form, the most celebrated member of the DNA structure family, as mentioned above, is the closest to the original Watson–Crick model. It is observed under conditions of high relative humidity and is characterized by a near-perfect 'ten' units per turn and the base pairs being located nearly astride the helix axis and normal to it (Langridge, Marvin *et al.*, 1960; Langridge, Wilson *et al.*, 1960) (Fig. 4). It is interesting to note that a random-sequence DNA in solution has a helical repeat intermediate between these two forms *viz.* ∼10.5 units per turn (Wang, 1979).

B′-DNA is a variant of the B-DNA structure, thought to be assumed by poly-d(A)·poly-d(T) as well as stretches of A·T base pairs, and is characterized by a large propeller twist (Fig. 2) and a narrow minor groove (Fig. 1*a*) compared with the normal B-DNA structure (Chandrasekaran & Radha, 1992; Lipanov & Chuprina, 1987).

C-form DNA was first observed for the lithium salt of calf thymus DNA and has about 9.3 residues per turn of helix (Marvin *et al.*, 1961), while D-DNA has eight residues in one turn and is observed for sodium salts of poly[d(A-T)]·poly[d(A-T)] as well as poly[d(A-A-T)]·poly[d(A-T-T)] sequences (Arnott *et al.*, 1974). Both C and D forms are also double-stranded helices with right-handed twists, but with slightly different structural parameters. All the four forms (A–D) are stable within a range of ionic or humidity conditions.

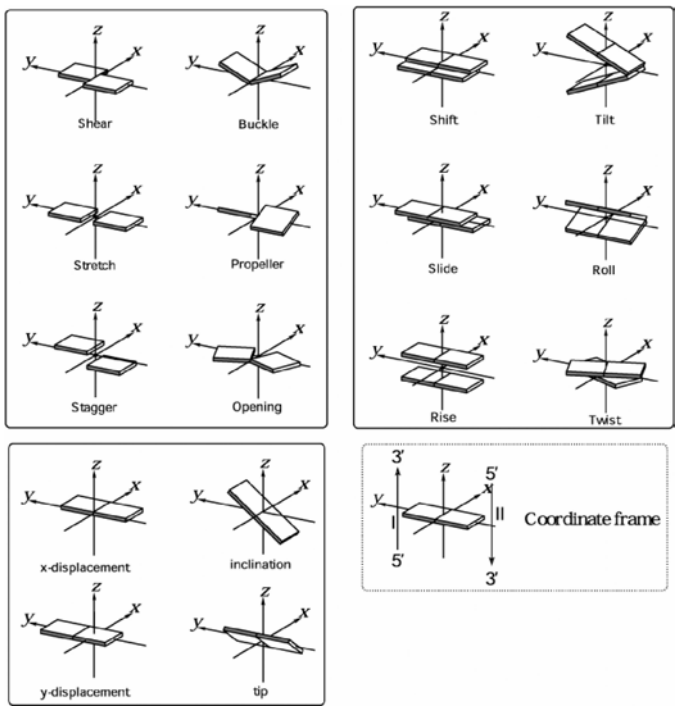

Figure 2
Schematic diagram showing the intra-base pair and other geometrical parameters, as suggested in the Cambridge convention (Dickerson *et al.*, 1989). The reference frame chosen is shown in the lower right-hand box with the minor groove facing the viewer. The translation and rotation parameters, which describe dinucleotide-step geometry, are shown in the top right-hand box. The commonly used helical parameters are included in the lower left-hand box, while the intra-base pair parameters are shown in the top left-hand box. (Adapted from Olson *et al.*, 2001.)

Another polymorphic form of right-handed double-helical DNA structure, with 15 residues in two turns of the helix or a twist of 48° per unit, was termed E-DNA (Leslie *et al.*, 1980). Recently, a structure observed for an oligonucleotide with methylated or brominated cytosine was also designated E-DNA, but had no similarity to the earlier structure. The structure, which was called E-DNA by the authors in order to recognize the extended nature of the helix and the eccentric trace of the backbone (Vargason *et al.*, 2000), raised some basic questions about DNA structure nomenclature which are discussed later in this review.

G-DNA describes a family of four-stranded quadruplex DNA structures with Hoogsteen-type hydrogen bonding between the four guanines in each of the stacked G-tetrads (Mohanty & Bansal, 1993). They can be formed with a parallel arrangement of four strands, as observed for poly(G) and some G-rich oligonucleotides (Horvath & Schultz, 2001), and also by folded-back chains of oligonucleotides (Fig. 5), as observed for some naturally occurring telomeric DNA sequences which have runs of Gs with A/T interruptions (Blackburn & Szostak, 1984; Balagurumoorthy *et al.*, 1992; Haider *et al.*, 2002).

H-DNA is an intramolecular triple-helical structure, formed under low-pH conditions, by DNA sequences containing long stretches of poly(purine)·poly(pyrimidine). In it, the pyrimidine-rich chain partly dissociates from its complementary strand and folds back parallel to the purine-rich strand onto the major groove of the Watson–Crick duplex (Fig. 5). This type of structure is believed to play a role in transcriptional control of gene expression (Lyamichev *et al.*, 1986; Htun & Dahlberg, 1989). Similar structures have also been reported for synthetic polynucleotides and oligonucleotides (Arnott *et al.*, 1976; Radhakrishnan & Patel, 1993). A naturally occurring $(dT\text{-}dC)_{18}\cdot(dA\text{-}dG)_{18}$ repeat which takes up the H conforma-

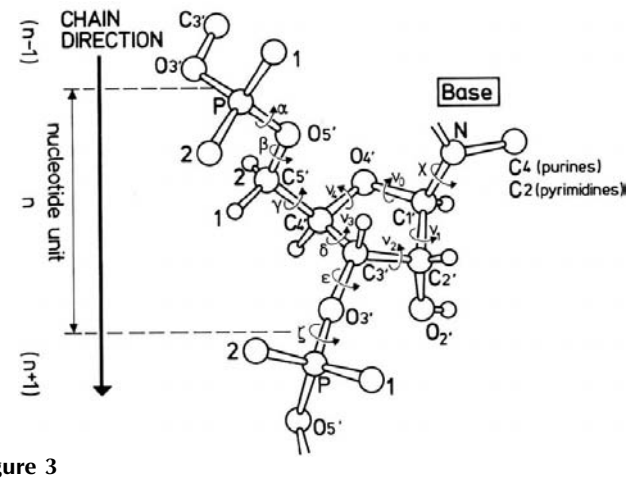

Figure 3
A ball-and-stick model showing the seven backbone torsion angles α, β, γ, δ, ε, ξ and the glycosidic torsion angle (χ) in the repeating unit of a polynucleotide chain. Most of the backbone torsion angles can take up any of the three staggered orientations (*trans*, *gauche*$^-$, *gauche*$^+$). The torsion angle δ, owing to the constraints of sugar-ring closure, is restricted to near-*trans* (C2′-*endo* pucker geometry) and near-*gauche*$^+$ (C3′-*endo* pucker) conformations.

tion at low pH has been shown to assume, under mildly alkaline pH conditions, a novel conformation termed J-DNA (Htun & Dahlberg, 1988).

The I-DNA structure results from the interdigitation of two parallel stranded duplexes of oligo(C), stabilized by hemiprotonated $C-C^+$ base pairing, in an antiparallel orientation (Chen *et al.*, 1994; Gehring *et al.*, 1993). An example of this unusual four-stranded DNA structure is shown in Fig. 5.

The term K-DNA is used to describe the kinetoplast DNA, which contains short adenine tracts in its sequence. It is characterized by anomalously slow gel electrophoretic mobility, a feature attributed to the sequence-dependent intrinsic curvature of the molecule (de Souza & Goodfellow, 1998).

L-DNA refers to DNA containing the sugar L-deoxyribose as part of its backbone and can act as an antisense DNA, resistant to nucleases (Urata *et al.*, 1992; Urata, 1999).

M-DNA is a complex between divalent metal ions and B-DNA. The imino proton in each base pair of the duplex is substituted by a metal ion under low-pH conditions; thus, the duplex behaves as a molecular wire (Aich *et al.*, 1999).

N-, O- and R-DNA have no structural connotations associated with them but are related to some location or functional aspect of DNA. Thus, while N-DNA describes 'nuclear' DNA, O-DNA refers to *ori*-DNA sequence, the site of origin of replication in bacteriophage λ (Alberts *et al.*, 1994). R-DNA refers to a recombinant DNA molecule formed by incorporating a non-homologous DNA fragment into the host molecule. This method of transformation of DNA has led to the emergence of the whole new field of biotechnology (Griffiths *et al.*, 1999).

P-DNA is the original model of DNA as proposed by Pauling, with three sugar–phosphate chains coiled around each other to form a close-packed core and the bases exposed on the outside (Pauling & Corey, 1953).

S-DNA describes a stretched form of DNA, with 1.6–1.65 times the normal rise (Cluzel *et al.*, 1996; Konrad & Bolonick, 1996). It has highly inclined and unwound bases, stabilized through inter-strand stacking interactions and phosphates located near to the central axis, as in 'P' DNA (Allemand *et al.*, 1998).

T-DNA is geometrically analogous to D-DNA (with an eight residues per turn right-handed helical structure) and is observed for phage T2 DNA, with 5-glucosylated cytosine residues, at <60% relative humidity (Mokul'skii *et al.*, 1972).

W-DNA or Z (WC)-DNA is a structure with a left-handed zigzag double-helical backbone but with base-pair orientation and chain directions as in B-DNA (Ansevin & Wang, 1990).

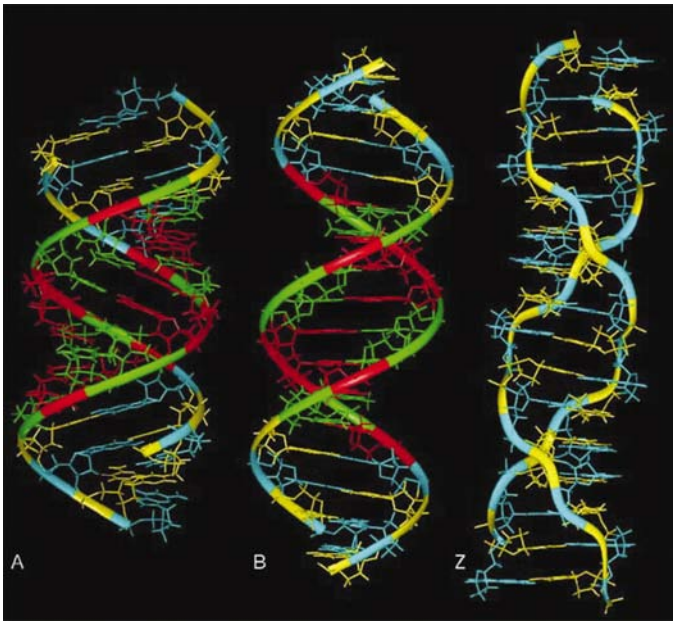

Figure 4
The currently accepted fibre model structures for A-, B- and Z-DNA are shown here using the ball-and-stick representation (Chandrasekaran *et al.*, 1989; Chandrasekaran & Arnott, 1996; Arnott *et al.*, 1980). The nucleotides are colour-coded (cytosine in yellow, guanine in cyan, thymine in green and adenine in red) and a ribbon is superposed on the backbones connecting the P atoms. A-DNA and B-DNA are both right-handed uniform double-helical structures, while Z-DNA is a left-handed double helix with a dinucleotide repeat and the backbone follows a zigzag path.

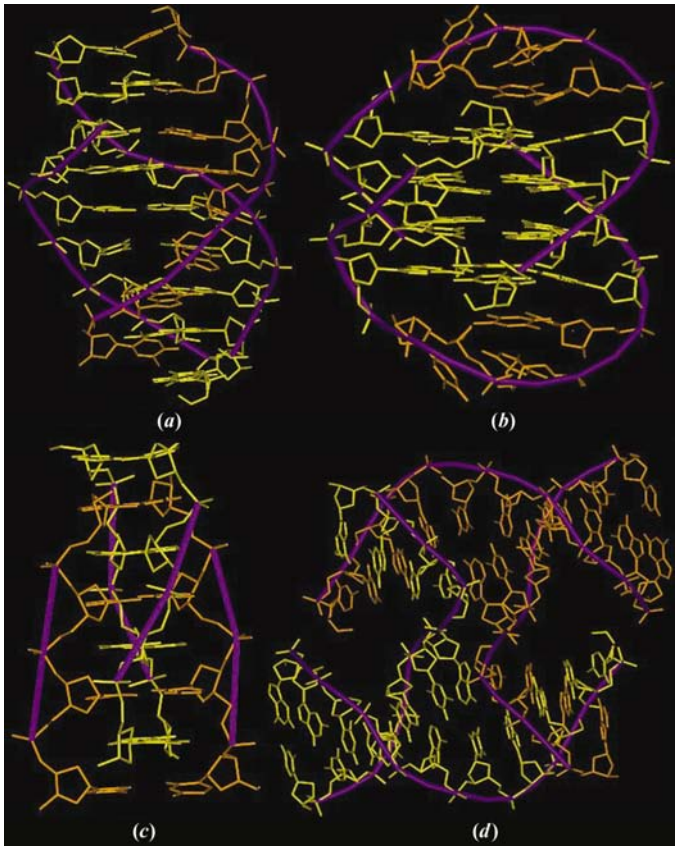

Figure 5
Some unusual DNA structures, with the molecules shown in stick representation and ribbons tracing the backbone. (*a*) A Py·Pu·Py triple helix reported from NMR, with the Watson–Crick duplex shown in yellow and the third strand shown in orange (Radhakrishnan & Patel, 1993). (*b*) An antiparallel G-quadruplex observed in the crystal structure of the oligonucleotide d(GGGGTTTTGGGG) with the loops shown in orange (Haider *et al.*, 2002). (*c*) The i-motif, a four-stranded intercalated structure observed for oligo-d(C4) (Chen *et al.*, 1994). (*d*) A Holliday junction structure for a decamer with the inverted repeat sequence d(CCGGTACCGG) (Eichman *et al.*, 2000).

X-form is an unusual form of duplex DNA adopted by poly (dA-dT) or $(dT\text{-}dA)_4$ at high concentrations of CsF and stabilized by aliphatic substituents, larger than a methyl group, in position 5 of the pyrimidine base. It has only been characterized by CD and NMR studies (Vorlickova *et al.*, 1991; Kypr *et al.*, 1996).

Z-DNA is a left-handed duplex structure with a dinucleotide repeat unit, generally confined to alternating purine (G) and pyrimidine (C) sequences. It has six dinucleotides per turn and exhibits a characteristic zigzag backbone. This is a consequence of distinctly different geometries for the two residues in the dinucleotide repeat which arise from alternating sugar puckers and *syn*/*anti* conformations about the glycosyl bond (Arnott *et al.*, 1980; Wang *et al.*, 1979).

Thus, only the letters F, Q, U, V and Y are now available to describe any new forms of DNA structure that may appear in the future.

3. The other generic forms of DNA

Apart from the structures with a one-letter 'name', there are several other generic descriptions of DNA. For example, form V DNA is used to describe supercoiled DNA, which may contain regions of right-handed B-DNA and left-handed Z-DNA, as suggested for the polypurine/polypyrimidine-rich sequences in the pBR322 plasmid (Brahmachari *et al.*, 1987; Stettler *et al.*, 1979).

During genetic recombination, one of the strands from each of two duplex DNA molecules exchange to form a four-way junction structure known as a Holliday junction (Holliday, 1964; Lilley & Norman, 1999). It has recently been observed by crystal structure analysis of a protein–DNA complex as well as in a free oligonucleotide (shown in Fig. 5) with an inverted repeat sequence (Ariyoshi *et al.*, 2000; Eichman *et al.*, 2000).

When DNA transforms from a coiled state to a globular structure, induced by high concentrations of polymer and inorganic salts, the condensate is known as ψ-DNA (Maniatis *et al.*, 1974). The term 'sticky' DNA is used to describe an intramolecular triplex formed by direct repeats of $(GAA{\cdot}TTC)_n$ (n = 9–150) in a plasmid or gene sequence that acts as a potential transcriptional blocker (Sakamoto *et al.*, 1999).

Alternative structures known as the side-by-side models (S-B-S) have also been proposed, wherein the two strands are not wound/twisted around each other, but contain helical fragments of five base pairs each, with alternating right and left-handedness, potentially facilitating uncoiling of the double helix during DNA replication (Sasisekharan *et al.*, 1978; Rodley *et al.*, 1976).

There are also other types of nucleic acids which are chemically distinct from DNA. They can differ by a simple substitution of the exocyclic group of the furanose sugar; the most commonly and biologically important example being ribonucleic acid (RNA) (Voet & Rich, 1970; Kim *et al.*, 1974). There can also be partial or complete replacement of the PO_4 group in the sugar–phosphate backbone by phosphorothioate (Cruse *et al.*, 1986) or methylphosphonate (Chacko *et al.*, 1983). Another recent addition is a polymer in which the nucleic acid backbone is replaced by a peptide backbone and which was therefore termed a peptide nucleic acid (PNA) (Nielsen, 1988; Rasmussen *et al.*, 1999). Duplexes and triplexes containing DNA/PNA chains have been reported. Interestingly, a DNA·PNA·PNA triple-helical structure determined by X-ray diffraction has been christened 'P-form helix' (Betts *et al.*, 1995). This should not be confused with the P-DNA structure proposed by Pauling and discussed above. Then there is the conformationally locked nucleic acid (LNA), wherein the sugar pucker is restricted through a 2′-*O*,4′-*C*-methylene bridge connection (Nielsen *et al.*, 1999). Finally, a 2′-5′ linked parallel-stranded double-helical model structure has been constructed using the cytidyl-2′,5′-adenosine crystal structure (Krishnan *et al.*, 1991) as a repeat unit.

It is instructive to remember this elaborate nomenclature for DNA, in view of the recent assignment of the term E-DNA (mentioned above) to describe an eccentric and extended DNA structure, when the same nomenclature was already being used to describe a quite different structure. However, apart from the confusion over nomenclature, it is also worth mentioning that the unusual parameters reported for the new 'E' structure (average residue rise of 3.6 Å and slide value of −2.3 Å) were subsequently attributed to the structural analysis program used. The rise and slide values were recalculated as 3.2 Å and −2.0 Å which, along with the observed C3′-*endo* sugar pucker, puts this structure within the A-DNA family (Ng & Dickerson, 2001; Vargason & Ho, 2001). Overall, there is general agreement that the sugar-pucker geometry and the disposition of the base pairs with respect to the helix axis are the parameters that best discriminate between the various right-handed duplex structures. Even so, several structures have now been reported as being intermediate between the well characterized A and B forms (Ng *et al.*, 2000; Vargason *et al.*, 2001).

4. Conclusion

The above discussion clearly highlights the need for caution in assigning new nomenclature to DNA polymorphs without detailed structure analysis as well as prior literature search. In addition, an examination of all available nucleic acid/oligonucleotide structural data indicates that it may be more relevant to talk about A, B or C type dinucleotide steps, rather than assigning a type to the whole structure, for DNA fragments with varying base sequences (Lu *et al.*, 2000).

New structures with non-Watson–Crick base pairing are also being reported on a regular basis. While this has become an accepted fact for multi-stranded triplex or quadruplex structures and isolated base pairs in duplexes, a hexamer with the sequence d(ATATAT) has been recently reported which takes up an antiparallel double-helical structure with all six A·T base pairs showing the Hoogsteen (Fig. 1*b*) hydrogen-bond scheme (Abrescia *et al.*, 2002). The authors have abstained from assigning a name to this structure, probably because the Hoogsteen hydrogen-bond scheme for A·T base

pairs has an illustrious past (Voet & Rich, 1970). However, the plethora of novel DNA structures does raise the question of whether we should abandon the rather arbitrary but historically relevant and commonly used nomenclature for nucleic acid structures and strive to arrive at a more rational scheme. However, until that takes place this brief review may serve as a ready reference guide for currently used DNA nomenclature.

Note added in proof: In addition to the recent confusion over use of the name E-DNA, it may be mentioned that the description L-DNA has also been used earlier to describe a ladder-type duplex structure, formed by calf thymus DNA saturated with intercalating $[(bipy)_2Pt(en)]^{2+}$ (Saenger, 1983).

This work was supported by financial assistance from CSIR, India. The authors are greatful to Mr Rahul Banerjee for help in the preparation of the manuscript.

References

Abrescia, N. G., Thompson, A., Huynh-Dinh, T. & Subirana, J. A. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 2806–2811.

Aich, P., Labiuk, S. L., Tari, L. W., Delbaere, L. J., Roesler, W. J., Falk, K. J., Steer, R. P. & Lee, J. S. (1999). *J. Mol. Biol.* **294**, 477–485.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J. D. (1994). *Molecular Biology of the Cell*. New York: Garland Publishing.

Allemand, J. F., Bensimon, D., Lavery, R. & Croquette, V. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 14152–14157.

Ansevin, A. T. & Wang, A. H. (1990). *Nucleic Acids Res.* **18**, 6119–6126.

Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H. & Morikawa, K. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 8257–8262.

Arnott, S., Bond, P. J., Selsing, E. & Smith, P. J. C. (1976). *Nucleic Acids Res.* **3**, 2459–2470.

Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. & Ratcliff, R. L. (1980). *Nature (London)*, **283**, 743–745.

Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C. & Watts, L. (1974). *J. Mol. Biol.* **88**, 523–533.

Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M. & Sasisekharan, V. (1992). *Nucleic Acids Res.* **20**, 4061–4007.

Bansal, M. (1999). *Curr. Sci.* **76**, 1178–1181.

Betts, L., Josey, J. A., Veal, J. M. & Jordan, S. R. (1995). *Science*, **270**, 1838–1841.

Blackburn, E. H. & Szostak, J. W. (1984). *Annu. Rev. Biochem.* **53**, 163–194.

Brahmachari, S. K., Shouche, Y. S., Cantor, C. R. & McClelland, M. (1987). *J. Mol. Biol.* **193**, 201–211.

Chacko, K. K., Lindner, K., Saenger, W. & Miller, P. S. (1983). *Nucleic Acids Res.* **11**, 2801–2814.

Chandrasekaran, R. & Arnott, S. (1996). *J. Biomol. Struct. Dyn.* **13**, 1015–1027.

Chandrasekaran, R. & Radha, A. (1992). *J. Biomol. Struct. Dyn.* **10**, 153–168.

Chandrasekaran, R., Wang, M., He, R. G., Puigjaner, L. C., Byler, M. A., Millane, R. P. & Arnott, S. (1989). *J. Biomol. Struct. Dyn.* **6**, 1189–1202.

Chen, X., Ramakrishnan, B., Rao, S. T. & Sundaralingam, M. (1994). *Nature Struct. Biol.* **1**, 169–175.

Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J. L., Chatenay, D. & Caron, F. (1996). *Science*, **271**, 792–794.

Cruse, W. B. T., Salisbury, S. A., Brown, T., Cosstick, R., Eckstein, F. & Kennard, O. (1986). *J. Mol. Biol.* **192**, 891–905.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Kitzing, E. V., Lavery, R., Nelson, H. C. M., Olson, W. K, Saenger, W., Shakked, Z., Skelnar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J. & Zhurkin, V. B. (1989). *J. Mol. Biol.* **205**, 787–791.

Eichman, B. F., Vargason, J. M., Mooers, B. H. & Ho, P. S. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3971–3976.

Franklin, R. E. & Gosling, R. (1953). *Acta Cryst.* **6**, 673–677.

Fuller, W., Wilkins, M. H. F., Wilson, H. R. & Hamilton, L. D. (1965). *J. Mol. Biol.* **12**, 60–80.

Gehring, K., Leroy, J. L. & Gueron, M. (1993). *Nature (London)*, **363**, 561–565.

Griffiths, A. J. F., Gelbart, W. M., Miller, J. H. & Lewontin, R. C. (1999). *Introduction to Genetic Analysis*. New York: W. H. Freeman & Co.

Haider, S., Parkinson, G. N. & Neidle, S. (2002). *J. Mol. Biol.* **320**, 189–200.

Hoogsteen, K. (1959). *Acta Cryst.* **12**, 822–823.

Holliday, R. (1964). *Genet. Res.* **5**, 282–304.

Horvath, M. P. & Schultz, S. C. (2001). *J. Mol. Biol.* **310**, 367–377.

Htun, H. & Dahlberg, J. E. (1988). *Science*, **241**, 1791–1796.

Htun, H. & Dahlberg, J. E. (1989). *Science*, **243**, 1571–1576.

Konrad, M. W. & Bolonick, J. L. (1996). *J. Am. Chem. Soc.* **118**, 10986–10994.

Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H., Seeman, N. C. & Rich, A. (1974). *Science*, **185**, 435–440.

Krishnan, R., Seshadri, T. P. & Viswamitra, M. A. (1991). *Nucleic Acids Res.* **19**, 379–384.

Kypr, J., Chladkova, J., Arnold, L., Sagi, J., Szemzo, A. & Vorlickova, M. (1996). *J. Biomol. Struct. Dyn.* **13**, 999–1006.

Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960). *J. Mol. Biol.* **2**, 38–64.

Langridge, R., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960). *J. Mol. Biol.* **2**, 10–37.

Lavery, R. & Zakrzewska, K. (1999). *Oxford Handbook of Nucleic Acid Structure*, edited by S. Neidle, pp. 39–76. Oxford University Press.

Leslie, A. G., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. (1980). *J. Mol. Biol.* **143**, 49–72.

Lilley, D. M. & Norman, D. G. (1999). *Nature Struct. Biol.* **6**, 897–899.

Lipanov, A. A. & Chuprina, V. P. (1987). *Nucleic Acids Res.* **15**, 5833–5844.

Lu, X.-J., Shakked, Z. & Olson, W. K. (2000). *J. Mol. Biol.* **300**, 819–840.

Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1986). *J. Biomol. Struct. Dyn.* **3**, 667–679.

Maniatis, T., Venable, J. H. J. & Lerman, L. S. (1974). *J. Mol. Biol.* **84**, 37–64.

Marvin, D. A., Spencer, M., Wilkins, M. H. F. & Hamilton, L. D. (1961). *J. Mol. Biol.* **3**, 547–565.

Mohanty, D. & Bansal, M. (1993). *Nucleic Acids Res.* **21**, 1767–1774.

Mokul'skii, M. A., Kapitanova, K. A. & Mokul'skaya, T. D. (1972). *Mol. Biol. (Moscow)*, **6**, 714–731.

Ng, H. & Dickerson, R. E. (2001). *Nature Struct. Biol.* **8**, 107–108.

Ng, H. L., Kopka, M. L. & Dickerson, R. E. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 2035–2039.

Nielsen, C. B., Singh, S. K., Wengel, J. & Jacobsen, J. P. (1999). *J. Biomol. Struct. Dyn.* **17**, 175–191.

Nielsen, P. E. (1988). *Nord. Med.* **113**, 268–271.

Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C. & Berman, H. M. (2001). *J. Mol. Biol.* **313**, 229–237.

Pauling, L. & Corey, R. (1953). *Proc. Natl Acad. Sci. USA*, **39**, 84–96.

Radhakrishnan, I. & Patel, D. J. (1993). *Structure*, **1**, 135–152.

Rasmussen, H., Kastrup, J. S., Nielsen, J. N., Nielsen, J. M. & Nielsen, P. E. (1999). *Nature Struct. Biol.* **4**, 98–101.
Rodley, G. A., Scobie, R. S., Bates, R. H. & Lewitt, R. M. (1976). *Proc. Natl Acad. Sci. USA*, **73**, 2959–2963.
Saenger, W. (1983). In *Principles of Nucleic Acid Structure*, edited by C. R. Cantor. New York: Springer–Verlag.
Sakamoto, N., Chastain, P. D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J. D. & Wells, R. D. (1999). *Mol. Cell*, **3**, 465–475.
Sasisekharan, V., Pattabiraman, N. & Goutam, G. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 4092–4096.
Souza, O. N. de & Goodfellow, J. M. (1998). *J. Biomol. Struct. Dyn.* **15**, 905–930.
Stettler, U. H., Weber, H., Koller, T. & Weissmann, C. (1979). *J. Mol. Biol.* **131**, 21–40.
Urata, H. (1999). *Yakugaku Zasshi*, **119**, 689–709. In Japanese.
Urata, H., Ogura, E., Shinohara, K., Ueda, Y. & Akagi, M. (1992). *Nucleic Acids Res.* **20**, 3325–3332.
Vargason, J. M., Eichman, B. F. & Ho, P. S. (2000). *Nature Struct. Biol.* **7**, 758–761.
Vargason, J. M., Henderson, K. & Ho, P. S. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 7265–7270.
Vargason, J. M. & Ho, P. S. (2001). *Nature Struct. Biol.* **8**, 107.
Voet, D. & Rich, A. (1970). *Prog. Nucleic Acid Res. Mol. Biol.* **10**, 183–265.
Vorlickova, M., Sagi, J., Hejtmankova, I. & Kypr, J. (1991). *J. Biomol. Struct. Dyn.* **9**, 571–578.
Watson, J. D. & Crick, F. H. C. (1953). *Nature (London)*, **171**, 737–738.
Wang, J. C. (1979). *Proc. Natl Acad. Sci. USA*, **76**, 200–203.
Wang, A. H. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. & Rich, A. (1979). *Nature (London)*, **282**, 680–686.