Sequence-dependent structural variation in B-DNA

Kazimierz Grzeskowiak

Though fiber diffraction originally led to the belief that the structure of DNA would be a simple regular helix, X-ray crystallography of synthetic oligomers has shown that both deformability and structure depend on sequence. But the rules that determine these factors remain mysterious.

Address: Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095-1570, USA.

Chemistry & Biology October 1996, **3**:785–790

 ISSN 1074-5521

X-ray crystal structure analysis first proved itself to be a useful tool in the study of DNA conformation in 1979, with the first report of a single-crystal X-ray structure of the hexamer CGCGCG [1]. The results themselves were something of a shock, as the helix was observed to be in neither of the canonical A or B forms anticipated by fiber diffraction analysis (which uses stretched fibers of DNA rather than the crystals used in more recent studies). Instead, the CGCGCG oligomer adopted an entirely new, left-handed zigzag Z-DNA structure, dramatically illustrating the structural diversity of DNA. Simultaneously, a group at Cambridge led by Olga Kennard was determining a structure that turned out to be A-DNA [2] and Dickerson and colleagues were studying the structure of the dodecamer CGCGAATTCGCG [3,4] (known as the 'Drew sequence'), an embedding of the *Eco*RI restriction site. At that time, the hypothesis was (understandably) that restriction sites might be especially promising candidates for crystallization, because endonuclease recognition sequences might be DNA sequences that were especially regular or well ordered. Dickerson and others hoped that by analyzing local helix parameters of different DNA sequences, and combining this with an understanding of the patterns of residue–base contact in DNA–protein complexes, it would be possible to make general rules for how DNA sequence determines structure, and how proteins recognize DNA. As it turns out, simple rules of this kind either do not exist or are much harder to find than would have been predicted at the time, due to both the intrinsic flexibility of DNA and the lack of predictability of amino acid–base interactions. Nevertheless, the work inspired by this initial optimism has taught us much about the dependence of structure and deformability on DNA sequence.

Lessons from dodecamer structures

During the decade following the single-crystal structure analysis of CGCGAATTCGCG, 22 variants of this sequence were analyzed [5]. The information from these crystal structures established two important principles: first, that variation of the width of the minor groove is determined by the DNA sequence, and, second, that bending of the DNA helix is also sequence-dependent.

It was the structure of Drew's dodecamer [6] that originally suggested that regions with runs of AT base pairs had intrinsically narrower minor grooves compared to those with runs of GC. In this structure, the very narrow (3.5 Å) AATT region of the minor groove in CGCGAATTCGCG was filled with a highly ordered zigzag spine of hydration [7], which seems to be important in the stability of the narrow groove structure. The discovery of sequence-dependent deformability came from the remarkable observation (see [5] for review) that several DNA dodecamers have a sharp bend (10°–20°) at one end, at the junction between the G–C and A–T regions (GC/AT junction) [4,8–14]. The bend is best seen at the 5'-end of CGCAAAAAAGCG. It is produced by one base pair rolling over the next along the long axis of the helix in a direction that compresses the major groove.

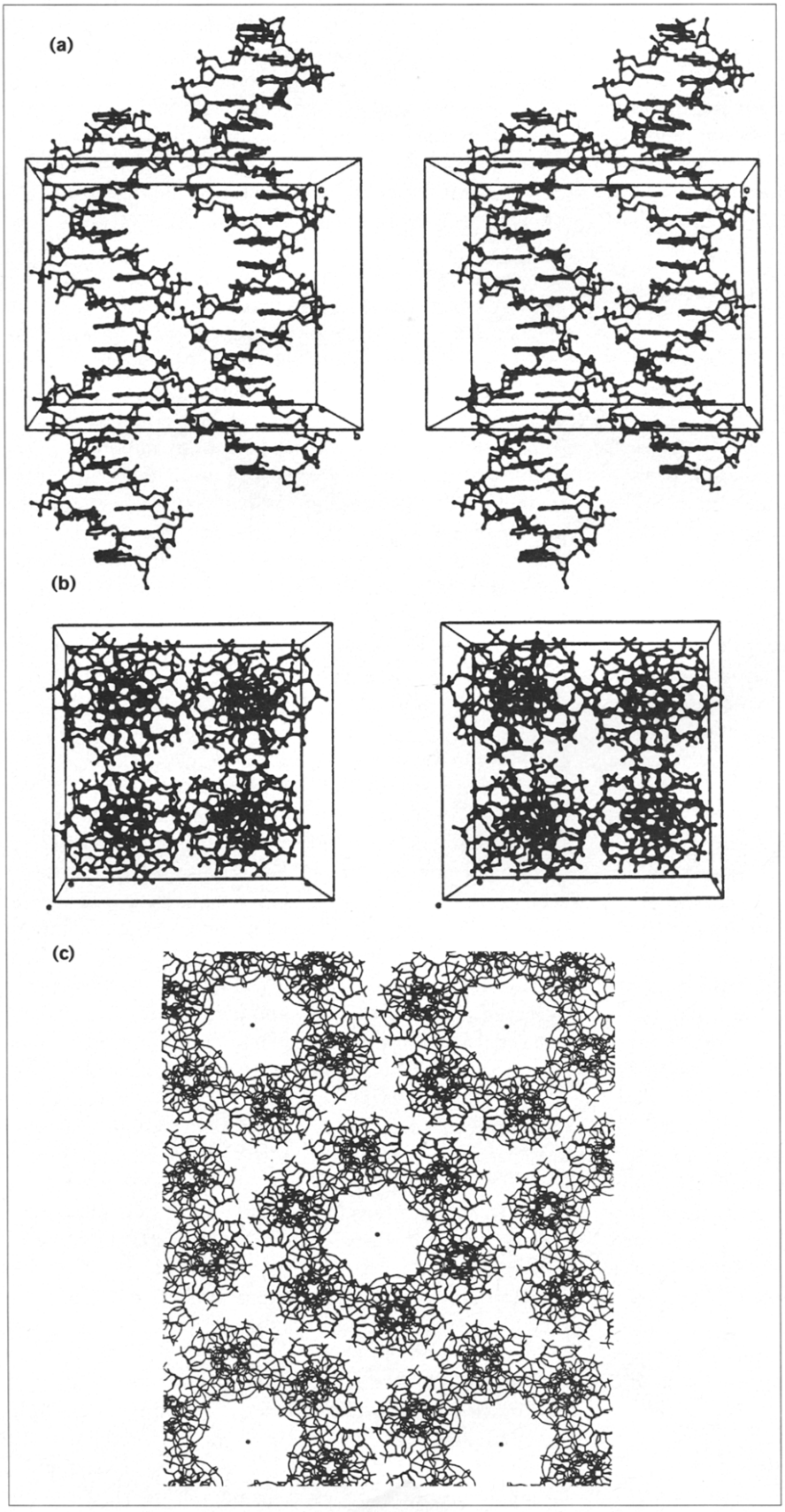
The move to decamers

In 1987 the strategy of DNA crystallographic analysis changed. A new class of B-DNA structures, the high resolution decamers, appeared, and it began to be possible to study the conformational effects of mismatched base pairs [15]. Crystals of decamers turned out to have far better diffraction patterns than those of dodecamers (resolutions of 1.3–1.6 Å), because of the orderly way in which helices were stacked within the crystals (Fig. 1). Because a decamer provides a single complete turn of the DNA helix, decamers can stack endlessly on top of each other, forming continuous columns, whereas dodecamers must overlap the last two bases. But the same characteristics that made the decamer crystals diffract to high resolution also made the results derived from them difficult to interpret. The packing of the helices in these crystals is so close that intermolecular forces may well be strong enough to distort the overall structure of the DNA helix. These intermolecular contacts are clearly important in determining whether the columns lie parallel to each other or cross at at an angle (Table 1).

Predispositions of sequence towards structure

During this period, Dickerson was working on his self-assigned task of establishing rules for predicting the behavior of a DNA helix based solely on its sequence. Calladine [16] had proposed a model in which the orientation of one base pair was mainly determined by stacking

Figure 1



Crystal packing of decamers. **(a)** Stereoview of the decamer CGATCGATCG, which packs in an orthorhombic cell, with endless vertical columns of stacked helices simulating an infinite helix in an extremely ordered manner. **(b)** Stereoview of the same crystal viewed along the c axis, the axis of the columns, shows the helices packed in a square array. **(c)** CCAAGCTTGG viewed down the c axis. Six columns of helices pack in a hexagonal array around a 6-fold rotation axis to form a hollow tube. These tubes are packed in a triangular lattice to build the crystal.

interactions with its nearest neighbor. But it became clear that many DNA sequences could be influenced by nearby bases to exhibit more than one state of local variables such as twist, rise, and roll [17]. Dickerson [18] therefore corrected the simple base-step model [16], improving it to take the influence of the neighboring nucleotide bases into account. By 1991 he had concluded [19] that this analysis needed to be extended to all 136 possible tetrad sequences involving regions of four successive base pairs. Even this, in light of our current knowledge, does not go far enough. To achieve the goal of understanding the factors that determine the structure an isolated oligomer will adopt, we need to study much longer segments of DNA than tetrads. Nevertheless, the study of tetrads has value. The propensity of an amino acid sequence to form α-helix or β-sheet may have predictive value even though a given sequence may sometimes adopt a different fold in the context of the whole protein. Similarly, the propensity of a tetrad to adopt a particular conformation may be important for the overall structure of a helix whether that tetrad indeed adopts its favored conformation in the context of the whole molecule or not.

But before an analysis of this kind can be helpful, one must establish which of the observed differences between the structures of different molecules could be due to crystal packing forces. There are two ways to look at this in the crystal structures of DNA decamers. One is to examine many isomorphous sequences in the same crystal setting, with similar crystal packing. The other is to look at the same sequence in several different crystal settings. Several sets of isomorphic crystal structures of decamers exist, in which all members of the set share a crystal form (see Table 1). But there are two possible reasons for two decamers to have the same structure. The sequence of the DNA molecule may directly determine its molecular shape, in turn determining the packing options available. Alternatively, the intermolecular contacts that the molecule can make may be the determining factor in the crystal form chosen, which then dictates the shape of the molecule. In fact, of course, both the intermolecular contacts and the intrinsic predisposition of a molecule to adopt a particular shape must be important, and it is very hard to disentangle these two factors. For example, three laboratories have analyzed the structure of unmethylated and methylated oligomers in an attempt to learn what structural changes the addition of

Table 1

B-form crystal structures discussed in this paper.

Sequence	Space group	Crystal form	Helix crossing angle (deg.)	Reference
CGCGAATrCGCG	$P2_12_12_1$	orthorhombic	180	2,3
CGATCGATCG	$P2_12_12_1$	"	180	20
CGATTAATCG	$P2_12_12_1$	"	180	27
CGATATATCG	$P2_12_12_1$	"	180	28
CGATATATCG	$P2_12_12_1$	"	180	28
CATGGCCATG	$P2_12_12_1$	"	180	34
CCAAGATTGG	C2	monoclinic	180	36
CCAACGTTGG	C2	"	180	31
CCAGGCCTGG	C2	"	180	22
CCAACITTGG	C2	"	180	26
CTCTCGAGAG	C2	"	43	30
CTCAGCTGAG	$P2_1$	"	90	–
CGATCG^{m}ATCG	$P3_221$	trigonal	60	21
CGATGC^{m}ATCG	$P3_221$	"	60	–
CCAACITTGG	$P3_221$	"	60	26
CCACTAGTGG	$P3_221$	"	60	37
CCATTAATGG	$P3_221$	"	60	38
CCAAGCTTGG	P6	hexagonal	180	32
CCAGGC^{m}CTTGG	P6	"	180	23

the methyl group may cause. Methylation of adenine or cytosine is an important control mechanism, for example, making a sequence resistant to cleavage by a restriction enzyme, and it seemed reasonable that this altered DNA–protein interaction might result from a change in the DNA structure. Indeed, the structure of CGATCGATCG-180 [20] is different from that of its methylated analog CGATCG^{m}ATCG-60 [21]; small changes in molecular structure result in a large change in intermolecular association, with consequent changes in local helix parameters. However, two other structures, CCAGGCCTGG-180 [22] and CCAGGC^{m}CTGG [23], have virtually identical local helix parameters even though they occur in two different crystal forms. Furthermore, methylation of the *Eco*RI recognition site in CGCGAATTCGCG does not alter the conformation of this oligomer [24]. Thus, it appears that the resistance to cleavage that results from methylation is not, or not always, due to major changes in the structure of the DNA. But the presence of a methyl group may be responsible for the dislocation of magnesium cations, whose accurate positioning within the helix is required for enzymatic cleavage of DNA phosphodiester bonds.

Our study of the factors that determine DNA conformation is in its infancy. Few conclusions can be made from the analysis of isomorphous crystal structures, and only about a third of the 136 tetrads have been examined to date, including a few crystal structures of nonpalindromic sequences. Any general conclusions concerning sequence effects at the tetrad level must be viewed cautiously, because a number of known structures appear to exhibit variation in local helix parameters that are not clearly related to the sequence.

Effect of crystal packing on structure

What of the other approach, examining the same sequence or segment of sequences in different crystal settings? Examples of oligomers for which this has been possible include GGGCGCCC (A-DNA) [2,25] and CCAACITTGG (B-DNA) [26], which have both been examined in two different space groups. Surprisingly large differences are found in the structure of the helix under differing crystallization conditions, including large changes in the width of the minor groove and in the helical twist of the C–A step. In the monoclinic structure of CCAACITTGG, the twist of the C–A step is about 50°, whereas this large twist is not encountered in the trigonal form of the same sequence.

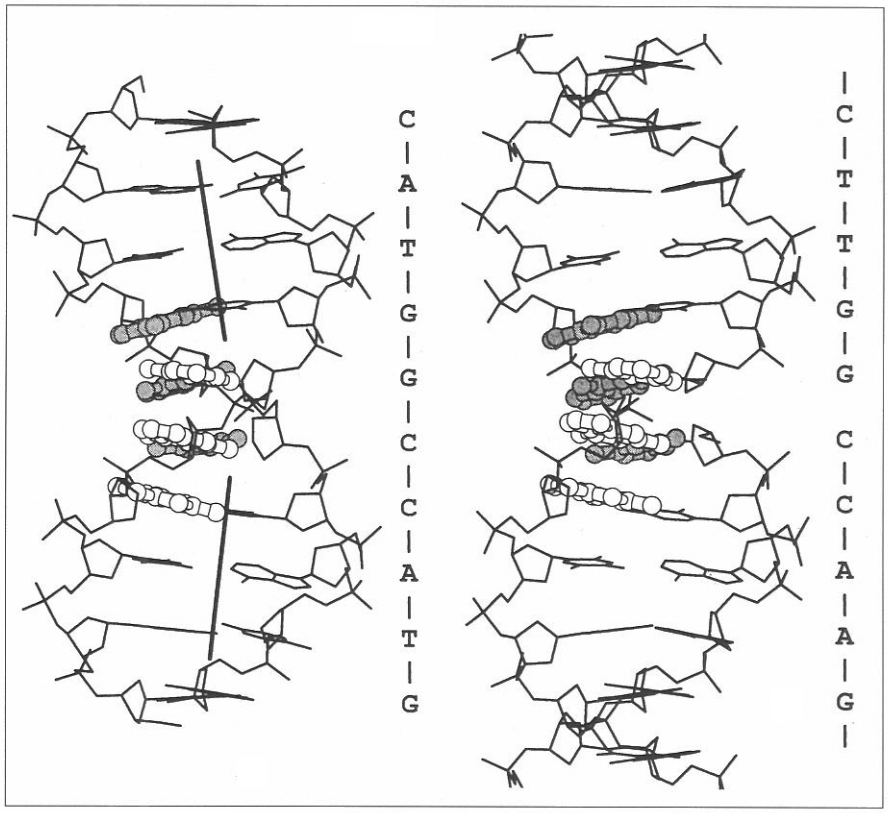
Minor groove width in the structure of CGATCGATCG [20] behaves in accordance with Drew's observation, showing periodicity of groove width with alternation of GC and AT base pairs. Yet the structures of CGATTAATCG [27] and CGATATATCG [28] both show exactly the same oscillation of minor groove width as CGATCGATCG, even though these oligomers include six consecutive AT base pairs. These apparent contradictions were later clarified by recognizing that the width of the minor groove in B-DNA is influenced by the phosphate conformation. The minor groove is widest when the opposing phosphate across the groove has the BII conformation [5,29], in which the main chain torsion angles ϵ (C-4'-C-3'-O-3''-P) and ζ (C-3'-O-3'-P-)-5') are (*gauche*$^{-}$, *trans*) rather than the more common (*trans*, *gauche*$^{-}$). But this still does not tell us how BII phosphate conformation is related to the base sequence. The identity of the cations present may also be a factor. It seems likely that the width of the minor groove is more variable in AT than GC base pairs. For example, CTCTCGAGAG [30] adopts a regular B-helical structure with a wide minor groove, instead of the oscillating structure that might be expected from the alternation of GC and AT pairs along the length of the helix. The width of the minor groove is such that it is clear that the AT pairs are adopting a GC-like conformation, instead of the other way around.

Bends in crystal forms and in solution

Despite all the problems of interpretation of crystal structures, it seems clear that some findings are not the result of crystal contacts. For example, at the C–A step in the monoclinic form of CCAACGTTGG [31], the twist angle is ~50° (the average helical twist of B-DNA oligomers is 36°, with a standard deviation of 4°). Such large twist angles have been observed at C–A steps in several structures [22,23,32,33] and in several different crystal forms. Thus it appears that the large twist at the C–A step is a consequence of base sequence. Furthermore, the examination of decamer structures CCAACGTTGG, CCAGGCCTGG, CCAAGCTTGG, CCAGGC^{m}CTGG and CATGGCCATG [22,23,31–33] reveals that in these decamers the sequence TGGCCA is bent, either when it is found in the center of the sequence or across the boundary between two decamers. Both types of bending are represented in Figure 2. These five structures also all show low twist at the three steps within G–G–C–C and high twist at flanking T–G and C–A steps. This example illustrates the dynamic nature of DNA structure and the long range of base-pair interactions along the DNA helix.

Even these findings have attracted their share of controversy, however. It has always been a source of concern whether the packing of DNA helices into a lattice might introduce aspects of structure that would not be present when the helices are floating free in solution. Organic solvents including 2-methyl 2,4-pentadiol (MPD) have been commonly used for oligonucleotide precipitation. The oldest model of macromolecule crystallization was that organic solvents acted by dehydration, removing water that allowed the DNA to remain in solution. However, the reverse can also be true. Upon addition of MPD, a thicker shell of water molecules is formed around the surface of the DNA oligonucleotides, and the thicker shell of water excludes cosolvent molecules.

Figure 2



Bending produced by the central TGGCCA sequence in CATGGCCATG (left) and across the interhelix junction TGG–CCA in CCAAGCTTGG (right). In the four central bases, the GGC bases in the near stand are shown in open space-filling atoms, and the GGC bases of the far strand are in dark space-filling atoms. Bending is produced by preferential stacking of guanine bases, as described by Goodsell *et al.* [33].

Gel mobility studies of DNA bending have contradicted many of the conclusions from X-ray structural analysis; such studies can only detect bends in some A-tract DNA. Concern about the effects of MPD were the subject of gel electrophoresis experiments with DNA oligomers that alternate short runs of adenine bases (A-tract) with general sequence DNA. Adding MPD to the gel medium at a concentration comparable to that used in growing DNA crystals (~25–30 %) seems to remove roughly half of the curvature of the DNA oligomer as measured by gel retardation.

Dickerson *et al.* [34] pointed out that high MPD concentrations, in both solution and crystals, decrease local bending somewhat, without removing it altogether. But in any case, all of the arguments concerning DNA bending are based on studies with short oligomers (decamers and dodecamers) and do not take into account the interactions between the bases and sugar-phosphate backbone. Thus, even when we understand the parameters involved in the degree of bending seen in a short sequence of DNA, it may be necessary to study much longer sequences to understand fully the relationship between sequence and molecular shape.

Towards biological significance

The most recent trend is towards detailed understanding of the structures of DNA sequences of particular biological relevance, instead of attempting to derive general rules from the study of many structures of relatively unimportant sequences. One potentially interesting sequence is CTCAGCTGAG, which can be considered to be a model for the sequence at the site of DNA recombination. Preliminary results (K.G., D.S Goodsell and R.E. Dickerson, unpublished data) suggest that the structure of this sequence is unusual, with cell dimensions consistent with an asymmetric unit that accommodates either two double-stranded helices or one double-stranded helix and a half duplex of CTCAGCTGAG. The X-ray survey photographs suggest that the helices cross at a right angle.

Another situation in which the structure of a DNA sequence may be important is in the gene for Huntington's disease, in which a stretch of CAG repeats within the affected gene expands on the Huntington's disease chromosome to produce an abnormally long polyglutamine region within the gene product, huntingtin, which eventually causes disease [35]. It is plausible that the structure of the DNA in the CAG tract may contribute to the mechanism that causes the number of repeats to increase in affected individuals. A model sequence studied in our laboratory is the 20mer CTCAGCTGAGCTCAGCTGAG, which contains a total of four CAGs as a dimer, although the CAGs alternate on the two strands.

Preliminary observations indicate that the structure of CTCAGCTGAGCTCAGTCGAG is quite different from its shorter parent CTCAGCTGAG.

The previously rate-limiting steps of synthesis, purification and crystallization of DNA oligomers have now been overcome by improved chemical techniques. The new challenge in this area is to study longer DNA sequences of particular biological interest.

Acknowledgements

This article is dedicated to Professor Richard E. Dickerson on the occasion of his 65th birthday. I would like to thank Dr David S. Goodsell for helping with figure preparation.

References

1. Wang, A.H.-J., *et al.*, & Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686.
2. Kennard, O. & Hunter, W.H. (1989). Oligonucleotide structure: a decade of results from single crystal X-ray diffraction studies. *Q. Rev. Biophys.* **22**, 327–379.
3. Wing. R., *et al.*, & Dickerson, R.E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758.
4. Dickerson, R.E. & Drew, H.R. (1981). Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J. Mol. Biol.* **149**, 761–786.
5. Dickerson, R.E. (1990). What do we really know about B-DNA? In *Structure and Methods, Vol. 3; DNA and RNA. Proceedings of the Sixth Conversation in Biomolecular Stereodynamics.* (Sarma, R.H. & Sarma, M.H., eds), pp. 1–38, Adenine Press, Schenectady, NY.
6. Drew, H.R. & Dickerson, R.E. (1981). Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **151**, 535–556.
7. Kopka, M.L., Fratini, A.V., Drew, H.R. & Dickerson, R.E. (1983). Ordered water structure of a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* **163**, 129–146.
8. Larsen, T.A., Kopka, M.L. & Dickerson, R.E. (1991). Crystal structure analysis of the B-DNA dodecamer C-G-T-G-A-A-T-T-C-A-C-G. *Biochemistry* **30**, 4443–4449.
9. Narayana, N., Gibell, S.L., Russu, I.M. & Berman, H.M. (1991). Crystal and molecular structure of a DNA fragment d(C-G-T-G-A-A-T-T-C-A-C-G). *Biochemistry* **30**, 4449–4455.
10. Coll, M., Frederick, C.A., Wang, A.H-J. & Rich, A. (1987). A bifurcated hydrogen-bonded conformation in the d(A–T) base pairs of the DNA dodecamer d(C-G-C-A-A-A-T-T-T-G-C-G) and its complex with distamycin. *Proc. Natl. Acad. Sci. USA* **84**, 8385–8389.
11. Edwards, K.J., Brown, D.G., Spink, N., Skelly, J.V. & Neidle, S. (1992). Molecular structure of the B-DNA dodecamer d(C-C-G-A-A-A-T-T-T-G-C-G))$_2$. An examination of propeller twist and minor-groove water structure at 2.2 Å resolution. *J. Mol Biol.* **226**, 1161–1173.
12. DiGabriele, A.D., Sanderon, M.R. & Steitz, T.A. (1989). Crystal lattice packing is important in determining the bend of a DNA-dodecamer containing an adenine tract. *Proc. Natl. Acad. Sci. USA* **86**, 1816–1820.
13. Nelson, H.C.M., Finch, J.T., Luisi, B.F. & Klug, A. (1987). The crystal structure of an oligo(dA)•oligo(dT) tract and its biological implications. *Nature* **330**, 221–226.
14. DiGabriele, A.D. & Steitz, T.A. (1993). A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.* **231**, 1024–1039.
15. Kennard, O. & Salisbury, S.A. (1993). Oligonucleotide X-ray structure on the study of conformation and interactions of nucleic acids. *J. Biol. Chem.* **268**, 10701–10704.
16. Calladine, C.R. (1982). Mechanics of sequence-dependent stacking of bases in B DNA. *J. Mol. Biol.* **161**, 343–352.
17. Dickerson, R.E. (1989). Definitions and nomenclature of nucleic acid structure parameters. *J. Biomol. Struct. Dyn.* **6**, 627–634.
18. Dickerson, R.E. (1983). Base sequence and helix structure variation in B- and A-DNA. *J. Mol. Biol.* **166**, 419–441.
19. Yanagi, K., Prive, G.G. & Dickerson, R.E. (1991). Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J. Mol. Biol.* **217**, 201–214.
20. Grzeskowiak, K., Yanagi, K., Prive, G.G. & Dickerson, R.E. (1991). Structure of B helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G. *J. Biol. Chem.* **266**, 8861–8883.
21. Baikalov, I., Grzeskowiak, K. & Dickerson, R.E. (1993). The crystal structure of the trigonal decamer C-G-A-T-^{6m}A-T-C-G: a B-DNA helix with 10.6 base-pairs per turn. *J. Mol. Biol.* **231**, 768–784.
22. Heinemann, U. & Alings, C. (1989). Crystallographic study of one turn of G/C rich B-DNA. *J. Mol. Biol.* **210**, 369–381.
23. Heinemann, U. & Alings, C. (1991). Conformation of a B-DNA decamer is mainly determined by its sequence and not by crystal environment. *EMBO J.* **10**, 35–43.
24. Frederick, C.A., Quigley, G.J., van der Marel, G.A., van Boom, J.H., Wang, A.H-J. & Rich, A. (1988). Methylation of the *Eco*RI recognition site does not alter DNA conformation. The crystal structure of d(CGCGA^{m6}ATTCGCG) at 2.0-Å resolution. *J. Biol. Chem.* **263**, 17872–17879.
25. Rabinovich, D., Haran, T., Eisenstein, M. & Shakked, Z. (1988). Structures of the mismatched duplex d(GGGTGCCC) and one of its Watson–Crick analogues d(GGGCGCCC). *J. Mol. Biol.* **200**, 151–161.
26. Lipanov, A., Kopka, M.L., Kaczor-Grzeskowiak, M. & Dickerson, R.E. (1993). Structure of the B-DNA decamer C-C-A-A-C-I-T-T-G-G in two different space groups. *Biochemistry* **32**, 1373–1389.
27. Quintana, J., Grzeskowiak, K., Yanagi, K. & Dickerson, R.E. (1992). Structure of a B-DNA decamer with a central T•A step: C-G-A-T-T-A-A-T-C-G. *J. Mol. Biol.* **225**, 379–395.
28. Yuan, H.S., Quintana, J. & Dickerson, R.E. (1992). Alternative structure for alternating poly(dA–dT) tracts: the structure of the B-DNA decamer C-G-A-T-A-T-A-T-C-G. *Biochemistry* **31**, 8009–8021.
29. Dickerson, R.E. (1992). DNA structure from A to Z. *Methods Enzymol.* **211**, 67–111.
30. Goodsell, D.S., Grzeskowiak, K. & Dickerson, R.E. (1995). Crystal structure of C-T-C-T-C-G-A-G-A-G. Implications for the structure of the holliday junction. *Biochemistry* **34**, 1022–1029.
31. Prive, G.G., Yanagi, K. & Dickerson, R.E. (1991). Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *J. Mol. Biol.* **217**,177–199.
32. Grzeskowiak, K., Goodsell, D.S., Kaczor-Grzeskowiak, M., Cascio, D. & Dickerson, R.E. (1993). Crystallographic analysis of C-C-A-A-G-C-T-T-G-G and its implications for bending in B-DNA. *Biochemistry* **32**, 8923–8931.
33. Goodsell, D.S., Kopka, M.L., Cascio, D. & Dickerson, R.E. (1993). Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl. Acad. Sci. USA* **90**, 2930–2934.
34. Dickerson, R.E., Goodsell, D.S. & Kopka, M.L. (1996). MPD and DNA bending in crystals and in solution. *J. Mol. Biol.* **256**, 108–125.
35. Huntington's Disease Collaborative Research Group (1989). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983.
36. Prive, G.G., Heineman, U., Chandrasegaran, S.S., Kan, L., Kopka, M.L. & Dickerson, R.E. (1987). Helix geometry, hydration and G–A mismatch in a B-DNA decamer. *Science* **238**, 498–504.
37. Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. & Sigler, P.B. (1994). Determinants of repressor/operator recognition from the structure of the *trp* operator binding site. *Nature* **368**, 469–473.
38. Goodsell, D.S., Kaczor-Grzeskowiak, M. & Dickerson, R.E. (1994). The crystal structure of C-C-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at T-A steps. *J. Mol. Biol.* **239**, 79–96.