

# Ancient history of X-ray crystal structure of B-DNA oligomers and its perspective

KAZIMIERZ GRZEŚKOWIAK<sup>1\*</sup>, HIROFUMI OHISHI<sup>2</sup>

<sup>1</sup>Molecular Biology Institute, University of California Los Angeles, USA

<sup>2</sup>Osaka University of Pharmaceutical Science, Osaka, Japan

\*Corresponding author: kaziu@mbi.ucla.edu

## Abstract

This is a retrospective review on the anniversary of the first X-ray crystal structure of B-DNA dodecamer CGCGAATTCGCG solved by R. Dickerson's team at UCLA/CIT. Timely established conclusions, drawn from the X-ray crystal structures of DNA oligomers, have explained some of the patterns of DNA oligomers interacting in complex with DNA binding drugs and proteins. The studies provide us with the insight into sequence mechanics and the sequence-dependent mode by which DNA double stranded helices interact with one another in a crystal. How to extrapolate information from DNA X-ray crystal structures to physical characteristics and function of these molecules in a solution remains a challenge. Nevertheless, these results are useful in the biophysical studies of DNA molecules and may shed some light on how to approach nanotechnology from a fundamental science perspective.

**Key words:** X-ray crystal structures, DNA decamers, base methylation, nanotubes

## Introduction

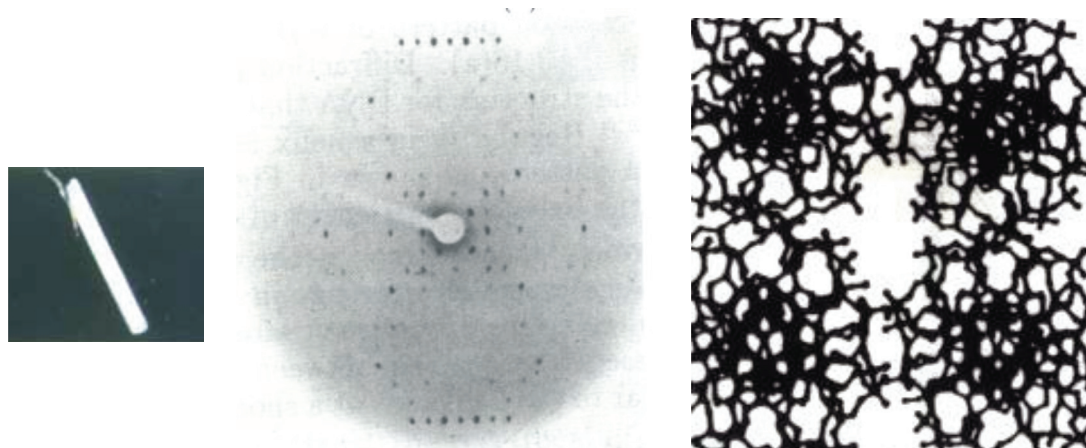
*Ancient history.* A little over 30 years ago, the X-ray crystal structure of CGCGCG Z-DNA hexamer was solved by A. Rich group (Wang et al., 1979). This discovery was followed by the publication of the crystal structure of a B-DNA dodecamer CGCGAATTCGCG, embedding the EcoRI restriction site (Dickerson and Drew, 1981). Since such 12-base-pair-long sequences are very rare in a genome, this piece of DNA was thought to be useful in the study of the structures of the complexes of anti-tumor agents targeting DNA (Kopka et al., 1985). In addition, DNA restriction sites were regarded especially useful for studying the recognition and structural patterns in DNA/restriction endonucleases complexes. The idea of our research was by analyzing a number of X-ray crystal structures of such complexes, we shall better understand the mechanism of the recognition of specific DNA sequences by proteins.

The work by John Rosenberg on X-ray crystal structure TCGCGAATTCGCG-EcoRI complex (Mc Clarin et al., 1986) spurred optimism that such a project could be persuaded. Thus the X-ray crystal structure of CGCGAATTCGCG was the impetus behind the further study of the variants of this sequence. All dodecamers

variants had CG ends on both sides of the sequence and all happened to be isomorphic. Noteworthy from this work is the conclusion drawn by A. Klug (Nelson et al., 1987) on the biological importance of oligo(A) .oligo(T) featuring of bifurcated hydrogen bonds.

In 1987, we shifted to a new class of B-DNA decamers of resolution structures ranging from 1.3 Å to 1.6 Å (Prive et al., 1987). These DNA decamers demonstrated interesting features. Ten base pairs provided one complete turn of the helix and thus the stacked helices one atop another would simulate an infinite DNA molecule. Then the contacts between entire lengths of helices could have been imposed by the sequence. This led us to view the influence of electrostatic forces at the edges of DNA bases on sequence dependent interaction mode between helices.

The presence of terminal CG/GC as well CC/GG base pairs, versus their absence at the ends in a DNA sequence would differentiate the intermolecular contacts between double stranded helices in the crystal. During nucleation and crystal formation, the terminal CG/GC and CC/GG base pairs are the major forces that produce crystal with the double stranded helices in parallel. If the CG/GC or GC/CG base pairs are in the middle of



**Fig. 1.** The crystal of CGATCGATCG is on the left and in the center is its X-ray survey photograph with the  $c$  axis vertical. It shows the helical phosphate backbone and the strong reflections noted at  $3.4\text{\AA}$  of stacked bases. In the 1980/90s, we used Buerger's precession camera to calculate cell dimension, crystal form and the space group. This is  $17^\circ$  precession hol photograph and film to crystal distance 100 mm. On the right is the view down  $c$  axis (length  $34\text{\AA}$ ) of parallel helices in crystal cell of  $40\text{\AA}$  square

the sequence without flanking CG/GC or CC/GG base pairs, the helices end up crossing each other at angles that vary within the sequence.

The importance of electrostatic forces on the edges of DNA CG bases as motors for imposing contacts between interacting double stranded helices was not discussed in the original publications on the X-ray crystal structures of DNA oligomers.

## Results

DNA structure of CGATCGATCG\* is probably the most classic example of a DNA molecule because of the sinusoidal widening and narrowing of a minor groove with an alternation of CG and AT base pairs. As we predicted, the double stranded helices stack one atop another in the orthorhombic cell  $P2_12_12_1$ . If the phosphate backbone of the helix can be compared to the threads of a screw, the two columns do not fit like a screw and nut. Instead, they only touch each other as of stacked columns packed side-by-side and parallel as is shown in Figure 1 (NDB ID; bdj025, Grześkowiak et al., 1991).

Since the methylation of the bases is an important control mechanism for rendering the sequence resistant to cleavage by restriction endonucleases, we analyzed the structural features of CGATCG<sup>me-6</sup>ATCG\*. This decamer is an example of just how sensitive the intermolecular interactions are upon base modification. Methylation of merely one amino group of adenine ( $N^6$ ) is sufficient to induce the new charge density on bases and forces the decamer to adopt an entirely new packing

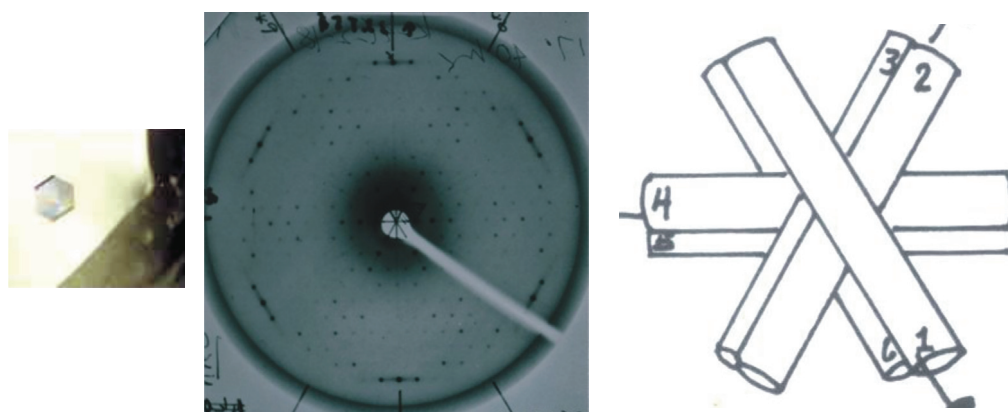
scheme, trigonal  $P3_221$ . In the trigonal lattice, only two logs of double helices are parallel and these two face the minor groove. Then, a set of two rods is rotated against another pair of rods by  $120^\circ$  – Figure 2 (NDB ID: bdj048 Baikalov et al., 1993).

CCAAGCTTGG\* helices build up the crystal cell as a ring of six columns with hexagonal morphology. Such hexagonal 6-fold arranged tube (Fig. 3) has a hollow inside with a volume capable of accommodating an additional double helix. Stacked helices exhibit  $23^\circ$  bend at T-G-G C-C-A junction across the non-bonded helices (NDB ID: 052, Grześkowiak et al., 1993). Such bending at GGCC segment is also observed in structures of CATGGCCATG (NDB ID; bdj048 Goodsell et al., 1993).

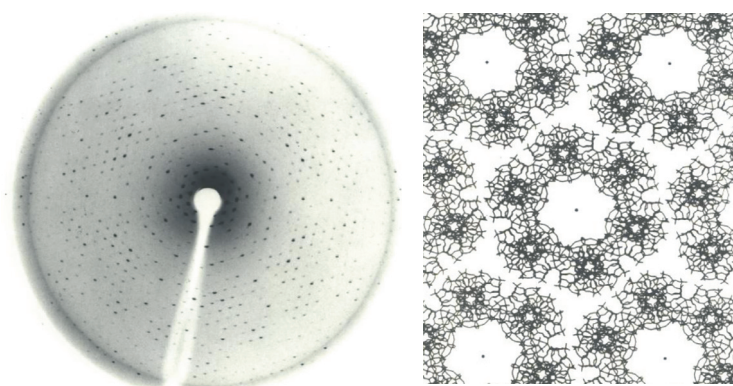
The next two decamers CTCTCGAGAG\* and CTCAGCTGAG\* create the crystals of a new class of structural features. They are related to the absence of flanking C-G base steps in each DNA sequence and the presence of only one C-G or G-C dimer in the center. In this division, the double stranded helices cross each other at the angles between helices axes varying from  $42$  to  $120^\circ$ .

Models of the Holliday junctions may be generated from the structures of the two DNA decamers presented below as well from the above-mentioned CGATCG<sup>me-6</sup>ATCG.

For CTCTCGAGAG, the X-ray survey photograph (Fig. 4 on the left) provides a cell dimension of  $a = 62.7\text{\AA}$ ,  $b = 24.48\text{\AA}$ ,  $c = 40.89\text{\AA}$ ,  $\beta = 127^\circ$ . In a monoclinic crystal, the helices interact side-by-side making an  $ab$  plane.



**Fig. 2.** The crystal on the left is in trigonal form, space group  $P3_221$ . In the center is a copy of archival photograph of the original film of X-ray survey photograph. From such patterns, Dick Dickerson always faultlessly drew the schematic sketch of packed helices in unit cell (picture on the right). He deduced that only two logs of parallel helices are formed and they cross another pair of logs at the angle of  $120^\circ$ . Cell dimension:  $a = b = 33.38 \text{ \AA}$ ,  $c = 98.3 \text{ \AA}$  and  $\gamma = 120^\circ$



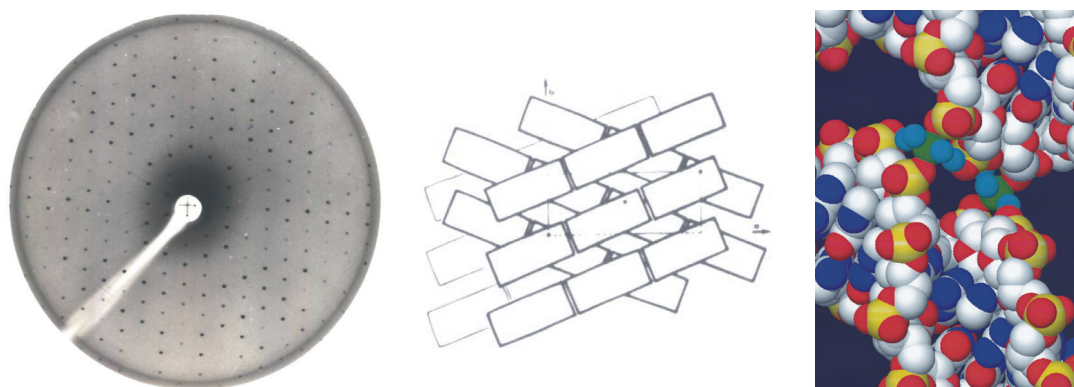
**Fig. 3.** On the left is an X-ray survey photograph from Buerger's precession camera. Calculated from this, the cell dimensions are:  $a = b = 53.08 \text{ \AA}$ ,  $c = 34.32 \text{ \AA}$  and  $\gamma = 120^\circ$ , space group  $P6$ . The picture on the right illustrates the packing of stacked helices in an arrangement of the hexagonal tube viewed down the  $c$  axis

These layers are then stacked along  $c$  in such a manner that individual helices from the top layer and from the bottom layer form an angle of  $42.5^\circ$  between helix axes (Fig. 4, picture in the center). Two of six  $\text{Ca}^{+2}$  ions connect two double stranded helices via phosphate backbones, (Fig. 4, picture on the right). Four others cations are within a double helix. The complete structure has been discussed by Goodsell et al., (NDB ID: bdj060 1993). On the right is the structure of two interacting helices.

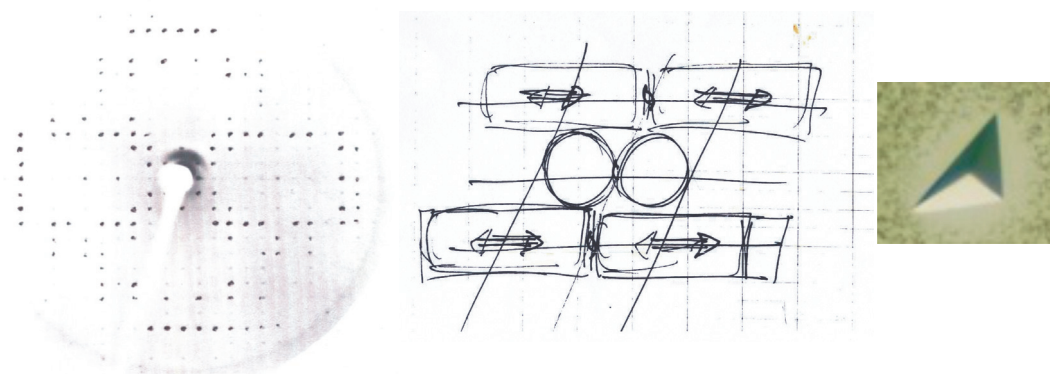
In the late 1990s, projects on X-ray crystal structure determination flat-lined and this happened at the time when crystals of DNA oligomers of more fascinating sequences appeared. The next four oligomers are examples of DNA crystals where structures have not been completely solved.

**CTCAGCTGAG\*** is one of the more potentially interesting sequences, which can be considered as a model structure for the strands exchanging at the DNA recombination site. Calculations from the X-ray survey photographs indicate that the crystal of CTCAGCTGAG creates its structure in a monoclinic array in the  $P2_1$  space group and cell dimensions,  $a = 34 \text{ \AA}$ ,  $b = 34 \text{ \AA}$ ,  $c = 40 \text{ \AA}$ .  $B = 99^\circ$  with the helices crossed at  $\sim 90^\circ$ . It also suggests that an asymmetric unit accommodates two helices. Patterson's map and the rotation function calculation provided a better solution for one double stranded helix and a half of duplex. Ironically, the molecular replacement method did not yield any further reasonable solution (Grzeškowiak K., unpublished data).





**Fig. 4.** Symmetry of X-ray patterns is related to the arrangement of helices in a crystal lattice (center) and is reflected in the structure showing crossed helices at  $42.5^\circ$  bound by two  $\text{Ca}^{+2}$  ions



**Fig. 5.** On the left is X-ray procession camera survey of CTCAGCTGAG in its monoclinic form and space group  $P2_1$ , Perpendicular reflections indicate helices cross one another at  $90^\circ$ . Thus the helical phosphate backbones are very distinct from the X-ray survey photograph of CGATCGATCG in Figure 1. An example of Dick Dickerson's hand drawing based on an X-ray survey photograph is depicted in the center and on the right is the crystal of the bromo-derivative  $\text{Br}^5\text{CTCAGCTGAG}^{**}$

The sequences of the next two DNA oligomers share a common motif with the previous CTCAGCTGAG. This is CAG trinucleotide, known to be unstable in the chromosome. When CAG triplet extends over the length of 40 consecutive repeats, the symptoms of precipitation among proteins occur due to the prolongation of the polyglutamine chain encoded by  $(\text{CAG})_n$  length.

For  $\text{CG}(\text{CAG})_4\text{CG}/\text{CG}(\text{CTG})_4\text{CG}^{**}$  (Fig. 6, crystal in yellow), preliminary calculations indicate that in a crystal lattice the helices of length of  $63 \text{ \AA}$  are packed parallel in a monoclinic array with  $3.9 \text{ \AA}$  helix rise/bp in contrast to  $3.4 \text{ \AA}$  in B-DNA or  $3.7 \text{ \AA}$  in Z-DNA, (Grzeskowiak et al., 2004).

$\text{CTCAGCTGAGCTCAGCTGAG}^{**}$  is 20-base pairs palindromic oligomer, which produces the hammer-rod shape crystal shown Figure 6 in blue color.

The last DNA decamer  $\text{CTCAATTGAG}^{**}$  has a context of 60% AT base pairs and such DNA sequences are

known in hyperthermophilic archaea (*Pyrococcus horikoshii OT3*) existing in deep underwater vents. Its optimal growth temperature is nearly  $100^\circ\text{C}$ , so is its DNA thermal stability. A diamond-shaped crystal of unsolved yet crystal structure of CTCAATTGAG is shown in Figure 6.

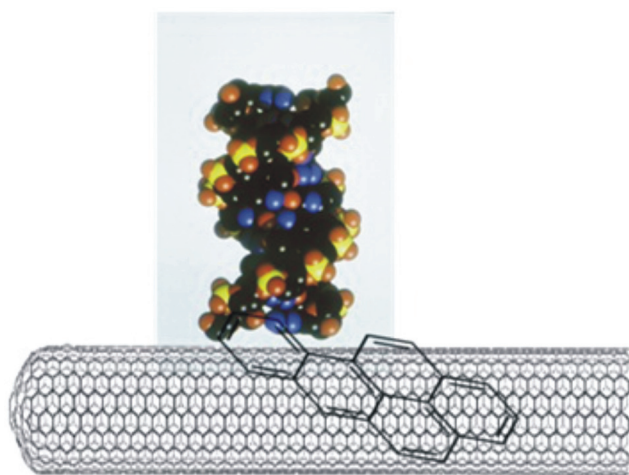
Studies on X-ray crystal structures, originally initiated by three major architects in DNA crystallography: Olga Kennard (A-DNA), Richard Dickerson (B-DNA), Alex Rich (Z-DNA) and others, revealed details of structural features of DNA molecules that raised a new perspective for the future research projects.

We synthesized benzo(a)pyrene-DNA adducts (BP-CGCGCG and BP-CGCGAATTCGCG) in a reaction of epoxy-benzo(a)pyrene with its corresponding DNA oligomer.

Because of the high hydrophobicity of aromatic polycyclic hydrocarbons, the adduct such as PB-DNA can be



**Fig. 6.** Crystals of DNA oligomers pending for the crystal structure determination are shown in clockwise direction: CG(CAG)<sub>4</sub>CG/CG(CTG)<sub>4</sub>CG, CTCAGCTGAGCTCAGCTGAG and CTCAATTGAG



**Fig. 7.** Schematic presentation of BP-DNA adduct irreversibly adsorbed onto the SWNT. Benzo(a)pyrene is attached to carbon SWNT while DNA molecule is capable to stick out of the surface and bind proteins or grooves binding components

adhesively placed on the surface of single wall nanotubes (SWNT), for example AC and CC conductivity measurements (Fig. 7).

Interestingly, BP-d(CGCGCG) adduct like its unmodified d(CGCGCG) undergoes a transition from B-form to Z-form accordingly, with an increasing salt concentration (Grześkowiak, unpublished results). In a solution of 3.0 M MgCl<sub>2</sub> its CD spectrum is virtually identical to the CD spectrum of dG-BP adduct (Jeffrey et al., 1977), whose structure shows the glycosidic bond in a syn position, a feature observed in Z-DNA. This adds to the evidence that in 3.0 M MgCl<sub>2</sub>, the guanine bases of BP-d(CGCGCG) are indeed in syn position similar to those observed in Z-form of the crystal of d(CGCGCG). In 0.1 M MgCl<sub>2</sub> solution, the BP-d(CGCGCG) returns to the B-form.

## Discussion

To assess the significance of a sequence in the interaction occurring between double stranded helices in a crystal lattice, we used the same crystallization conditions except that DNA oligomers of unsolved structures yielded better diffracting crystals when Mg<sup>+2</sup> ion was replaced by Ca<sup>+2</sup>. Crystals were grown from standing drops at +5 °C from solution of 0.37-1.34mM DNA duplex, 100-300 mM magnesium chloride or calcium acetate and 15-20% (v/v) 2-methyl-2,4-pentanediol (MPD) for a duration of 4-6 weeks.

In order to grow well-diffracting crystals, it is essential to find equilibration conditions in a well (solution of DNA oligomer, salt and MPD). The state of equilibrium depends upon the DNA sequence. It is noteworthy that the size of the magnesium atom compared to calcium and correspondingly the coordination numbers of six for Mg<sup>+2</sup> versus seven for Ca<sup>+2</sup> have an effect on bridging the two interacting DNA duplexes and packing in the crystal.

It seems to be the case that G-C dimers in the middle of DNA sequence favor Ca<sup>+2</sup> over Mg<sup>+2</sup> to produce better aligned helices in the crystals. Nevertheless, the mutual affinity of C-G dimers and the charge densities on the edges of DNA bases are the most important factors that bring double stranded helices together during lattice formation. The dipole moments for guanine  $\mu = 6.5$  and cytosine,  $\mu = 6.3$  (in D) point in the opposite direction. Dipole moments for adenine is  $\mu = 2.6$  and for thymine  $\mu = 4.0$  (Sponer et al., 2001). Respectfully, the diagonal vector for G-C bp is  $\mu = 0$  and for A-T bp is  $\mu = \sim 3.4$ .

The data of X-ray crystal structures are essential for defining the local helix parameters of DNA bases and the sugar-phosphate conformation in DNA helix. However, they also provide convenient framework to study biophysical properties of DNA, i.e. when measuring the elastic energy of DNA strands for the analysis of the dynamics of molecular recognition (Qu, 2011).

An existing challenge is to find how the results from DNA crystal structures can be extrapolated to the function of these molecules in a solution or embedded into a defined sequence context. For instance, to find the bending of a DNA segment in a crystal state is essential but needs to be compared with where double stranded DNA is bent in a solution. This would help to functionalize DNA strands stiffness for the creation of ordered structures over extended lengths of DNA in order to im-

plement a programmable charge transfer through DNA bases (O'Neill and Barton, 2005), and to fabricate designed molecular self-assembly of DNA on SWNT surface to the art of nanotechnology. A separate issue that may be addressed is as to how the X-ray diffractions patterns observed in the crystals would comply with the technology of a programmable molecular computing machine composed of enzymes and DNA molecules.

X-ray analysis of DNA crystals depict static structures that provide an insight into the mechanics of DNA helices by showing how to assess the competing influences of a nucleotide sequence on the three-dimensional DNA structure. The question is how to use such data in the determination of the biophysical characteristics and the function of those molecules in a solution. As it often happens, something that sounds impossible today becomes a reality the next day. Discoveries appear rapidly and the DNA Data Banks may offer attractive tools in this regard.

DNA oligomers were crystallized at: \*UCLA, \*\*OUPS.

### Acknowledgments

This work was supported by the Grant from the Japanese Society for the Promotion of Science.

This article is dedicated to Professor Richard E. Dickerson on the occasion of his 80<sup>th</sup> birthday.

### References

- Baikalov I., Grześkowiak K., Dickerson R.E. (1993) *The crystal structure of the trigonal decamer CGATCGmATCG: a B-DNA helix with 10.6 base pairs per turn*. J. Mol. Biol. 231: 768-784.
- Dickerson R.E., Drew R.H. (1981) *Structure of a B-DNA dodecamer II. Influence of base sequence on helix structure*. J. Mol. Biol. 149: 761-786.
- Goodsell D.S., Kopka M.L., Cascio D., Dickerson R.E. (1993) *Crystal structure of CATGGCCATG and its implication for A-tract bending models*. Proc. Natl. Acad. Sci. USA 90: 2930-2934.
- Goodsell D.S., Grześkowiak K., Dickerson R.E. (1995) *Crystal structure of CTCTCGAGAG. Implications for the structure of the holiday junction*. Biochemistry 34: 1022-1029.
- Grześkowiak K., Yanagi K., Prive G.G., Dickerson R.E. (1991) *Structure of B helical CGATCGATCG and comparison with CCAACGTTGG*. J. Biol. Chem. 266: 8861-8883.
- Grześkowiak K., Goodsell D.S., Kaczor-Grześkowiak M., Cascio D., Dickerson R.E. (1993). *Crystallographic analysis of CAAAGCTTGG and its implication for bending B-DNA*. Biochemistry 32: 8923- 8931.
- Grześkowiak K., Sawaya M.R., Ohishi H., Tomoo K., Ishida T. (2004) *Function of AGCA tetrads in (CAG)<sub>n</sub> repeats*. Nucleic Acids Symp. Series 48: 291-292.
- 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.
- Jeffrey A.M., Weinstein I.B., Jennette K.W., Grześkowiak K., Nakanishi K., Harvey G.R., Autrup H., Harris C. (1977) *Structure of benzo(a)pyrene-nucleic acid adducts formed in human and bovine bronchial explants*. Nature 269: 348-350.
- Mc Clarin J.A., Frederick C.A., Rosenberg J.M. (1986) *Structure of the DNA-EcoRI endonuclease recognition complex at 3Å resolution*. Science 234: 1526-1541.
- Nelson H.C.M., Finch J.T., Luisi B.F., Klug A. (1987) *The crystal structure of an oligo(dA).oligo(dT) track and its biological implications*. Nature 330: 221-226.
- O'Neill M.A., Barton J.K. (2005) *Sequence-dependent DNA Dynamics: the regulator of DNA-Mediated charge transport*. In: *Charge transfer in DNA: from mechanism to application*, ed. Wagenknecht H.A., Wiley-VCH, pp. 27-75.
- 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.
- Sponer J., Leszczynski J., Hobza P. (2001) *Electronic properties, hydrogen bonding, stacking and cation binding of DNA and RNA bases*. Biopolymers 61: 3-31.
- Qu H., Zocchi G. (2011) *The completed bending energy function for nicked DNA*. Europhys. Lett. 94: 18003.