Pure & Appl. Chem., Vol. 56, No. 8, pp. 989–1004, 1984.
Printed in Great Britain.

0033-4545/84 $3.00+0.00
Pergamon Press Ltd.

DNA FROM A-Z - A SURVEY OF OLIGONUCLEOTIDE STRUCTURES

Olga Kennard
University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, UK.

Abstract - The paper discusses the application of single crystal X-ray analysis to synthetic DNA fragments and surveys oligonucleotide structures published between 1978 and 1984. It focuses on the effects of base sequence on local conformation, on the influence of the base stacking and the role of water in stabilising global conformations.

INTRODUCTION

Periodically, in different areas of science, a single brilliant idea emerges, which reduces the complexity of experimental observations to one basic concept and provides an insight into the mechanism of a wide range of disparate phenomena. A state of euphoria usually follows such a discovery while the new concept is applied to an even wider range of phenomena only to reveal, ultimately, a higher level of complexity, needing perhaps a new simplification.

The double helical structure of DNA proposed by Watson and Crick in 1953 [1] provided such a unifying concept to molecular biology. It lead to the discovery of the genetic code, to the very simple mechanism of the transmission of genetic information by complementary hydrogen bonding of base-pairs and to an explanation of mutations as a result of errors in the reading of the genetic code. During the next thirty years, as these ideas were explored in depth, more and more became known about the complexity of the biological mechanisms controlling these events. Many of the control functions were found to be mediated by proteins and enzymes which appear to recognise and interact with specific regions of DNA, characterised by conserved or semi-conserved base sequence.

Indeed, even without a full understanding of the molecular basis of these interactions it has been possible to use enzymes, particularly those involved in DNA cleavage and repair to determine the primary structure of DNA's of a variety of organisms and ultimately to create artificial mutations. A whole new field of genetic engineering was opened up with widespread scientific, medical and social implications, of a complexity hardly anticipated in the early days following Watson's and Crick's discovery. One of the keys to understanding this new complexity lies in the detailed secondary structure of DNA and the influence of specific base sequence on local conformation.

The original model for the double helix was based on X-ray fibre diffraction patterns which, because of the inherent disorders of the molecules, and lack of resolution, only indicate gross structural features, and certainly not conformational details at an atomic level. An added limitation of the method is that it can not be used to derive a structure directly, but only to test hypotheses and models. Even then there may be difficulties in rejecting a wrong model. Although improved techniques for preparing fibres and computer fitting of models gave improved structural parameters [2] there was little change in our view of DNA as a monotonous right-handed helix, with conformational versatility limited to a few distinct families, principally the A and B-types.

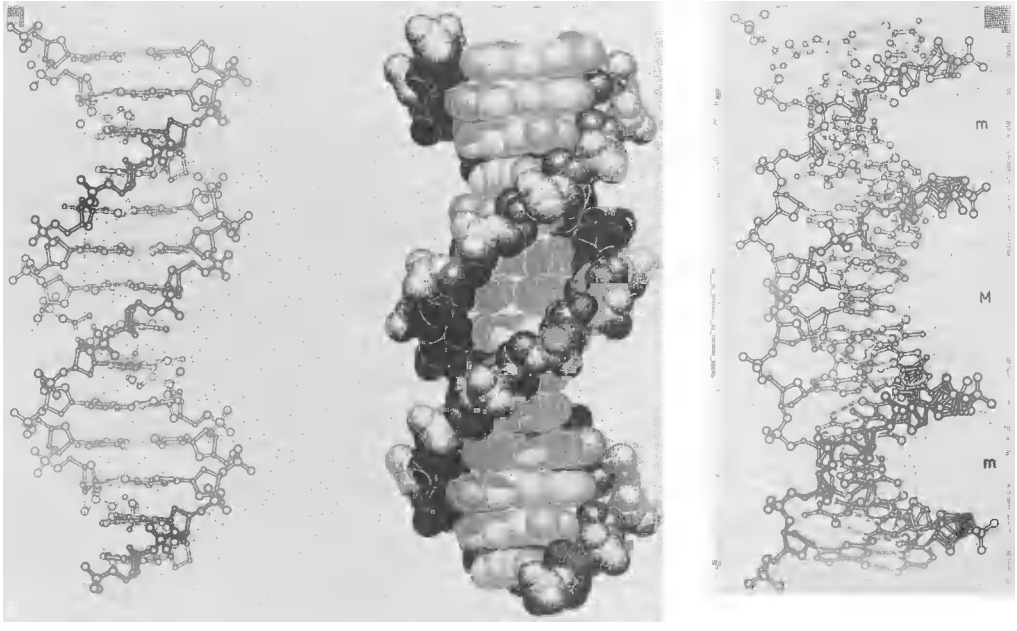
These same limitations do not apply to X-ray diffraction from single crystals of homogeneous molecules. With single crystals it is possible to derive the structure of molecules at atomic or near atomic resolution. The technique was used with outstanding success for the analysis of the

structure of many thousands of organic and biologically important molecules including several hundred proteins and a few viruses and transfer RNAs. Results of these studies have brought, about a profound change in our view of the structure and function of macromolecules and contributed greatly to our understanding of the chemistry of biological processes.

Unfortunately it has not been possible, so far, to obtain single crystals from either natural DNA fragments or synthetic polynucleotides. In the mid-1970's however, the development of the triester method of oligonucleotide synthesis [3,4] enabled the preparation of short lengths of DNA in quantities and purities required for crystallisation experiments. Further improvements in chemical syntheses, particularly the introduction of solid support techniques [5] allowed the relatively rapid production of defined sequences of up to approximately 20 base pairs. The way was opened for crystallographers, in collaboration with physical and organic chemists, to attempt high resolution studies of the structure of DNA fragments. The work to be described was carried out by three groups of laboratories and I am grateful to Professor R.E. Dickerson at UCLA and Professor A. Rich of MIT and their colleagues who shared with us their results over the years, often in advance of publication.

STRUCTURAL MODELS - B AND A DNA

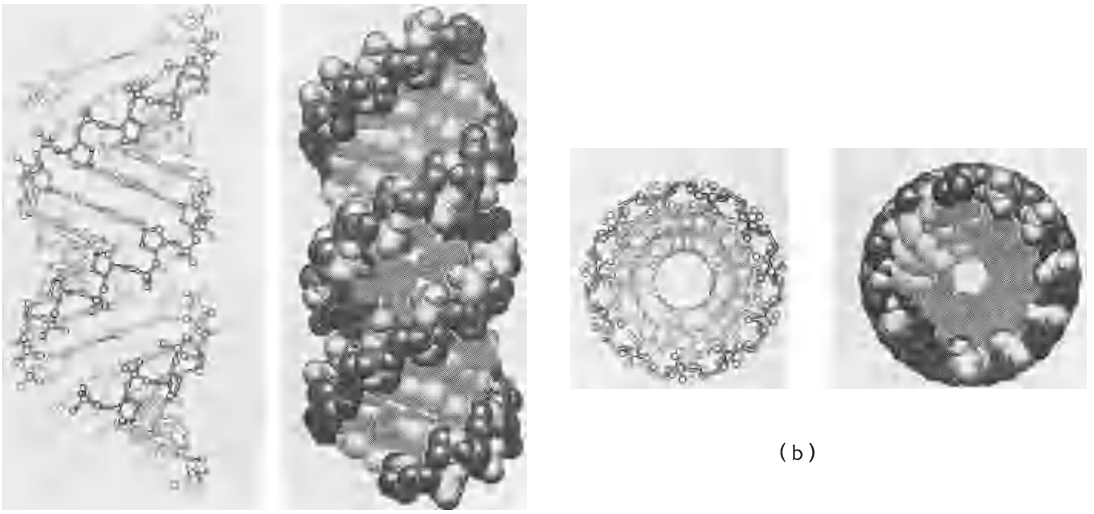
The original model for the continuous B type helix, the conformation associated with the high humidity form of DNA and DNA in vivo, is characterised by base pairs which are perpedicular to, and located on the helix axis. There are 10 base pairs for a full helical form with an axial rise per nucleotide of 3.3-3.4A and a phosphate phosphate separation of 7.0A. The sugar conformation is uniformly 2' endo. The resultant helix is relatively squat, with major and minor grooves of nearly equal depth (Fig. 1).



(a) (b)

Fig. 1. Structure of B-DNA from fibre coodinates. (a) side view (b) view showing major (M) and minor (m) grooves.

B-DNA is transformed to A-DNA when the humidity of the fibre environment is reduced or, in solution, by addition of dehydrating agents. In the fibre model of A-DNA the base pairs are displaced from the helix axis and inclined to the axis by about 20°. The complete helical turn consists of 11 base pairs with an axial rise per residue of 2.56Å. The sugar conformations are all assigned as C3' endo. The resultant helix illustrated in Fig. 2, has an open central core, a deep and narrow major groove and a relatively wide and shallow minor groove.



(a) (b)

Fig. 2. Structure of A-DNA from fibre coordinates. (a) side view (b) top view

Base stacking is one of the main forces stabilising DNA conformation. There are marked difference in the stacking patterns in A and B-DNA even when these patterns are restricted to monotonous helices, without regard to the actual purine-pyrimidine primary sequences. In B-DNA (Fig.3B) only intrastrand stacking is observed: base overlap in purine-pyrimidine and purine-purine steps is good and contrasts with the poor stacking seen at pyrimidine-pyrimidine and pyrimidine-purine. In A-DNA (Fig.3A) there is almost no pyrimidine-pyrimidine stacking but significant interstrand overlap of purines occurs at pyrimidine-purine steps.

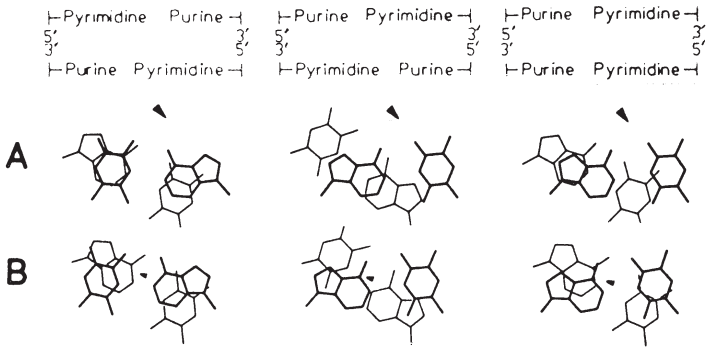


Fig. 3 Base stacking overlaps in A and B type double helix.

Until the advent of single crystal analysis there was considerable emphasis on the differences between A and B forms in terms of the C2' endo, C3' endo sugar conformation. It should, however, be emphasised that the sugar conformation in deoxyoligonucleotides is relatively flexible, with shallow energy barriers estimated to be as low as one Kcal [6]. The fundamental difference is more likely to be related to the relative orientation and stacking of bases in the two forms. The flexible backbone then adopts the lowest energy conformation including changes in the sugar base orientation, and, for steric reasons, in the sugar conformation. The importance of base stacking, the flexibility of the backbone and variable sugar conformation have indeed been confirmed by the single crystal X-ray analyses described in the following sections.

SOME CONSTRAINTS OF SINGLE CRYSTAL ANALYSIS

Experimental limitations

As discussed in the introduction, single crystal X-ray analysis can be carried to the limit of atomic resolution and it is possible to deduce the structure directly from X-ray data without chemical assumptions other than the existence of discrete atoms. For such an analysis, however, it is essential to measure the diffracted intensities to a resolution around 1Å or better. This requires :

- Single crystals of sufficient size to give measurable intensities to the limit of resolution - usually at least 0.2x0.1x0.1 mm^3.
- Moderate thermal vibration of atoms since high thermal motions limit resolution.
- The effect of thermal vibrations can be minimised by taking measurements at temperatures well below the melting point of the crystal.
- Highly ordered arrangement of atoms in the crystal structure, including, if the crystal is solvated, a well ordered solvent structure.

If the intensities can only be measured to a resolution of around 2.5Å the structure can still be solved, usually either by isomorphous replacement, which requires the introduction of one or more heavy atoms without serious perturbation of the crystal structure, or by phase measurements using anomalous scatterers. There are also more or less sophisticated trial and error methods which are based on trying to fit a hypothetical model structure to the experimental observations to obtain a starting solution.

Structure refinement

Once the phase problem is solved it is possible to use all measured experimental intensities to refine the positions and thermal vibrations of all atoms in the structure. With high resolution structures of moderate size the individual positional parameters can be located with an accuracy of around 0.01-0.001Å since there are usually about 10 experimental points for each variable being refined. For large molecules and especially for large molecules of low (~2Å) resolution there are proportionally fewer data points and it may be necessary to refine the structure as if it were composed of individual rigid molecular fragments. There are several different approaches to such constrained refinements and great care must be taken not to inhibit the refinement of a stuctural feature - not to impose, for example, planarity on a group of atoms which may in reality not be planar.

Problems with oligonucleotides

Early experiments with oligonucleotides soon indicated that these compounds are extremely difficult to crystallise. In the three laboratories mainly involved in this work many thousands of crystallisation experiments were carried out on a large variety of oligomers between 4-12 base pairs long. Success has so far been confined to the oligomers listed in Table 1, a few others whose structures have been solved but not yet published, and a handful for which experimental measurements are available but whose structures could not, as yet, be solved. All the crystals so far examined are hydrated, (up to 60% by weight) and contain spermine and various counterions including Mg^{2+}, which aid crystallisation.

With the exception of two structures, intensity data could not be measured below 1.8-2Å probably because of disordered solvents and very small crystal size. The solution of oligonucleotide structures also presented its own problems with considerable restrictions on the heavy atom method, which had been so successful in the analysis of proteins. New trial and error methods were introduced [7] with spectacular success in favourable cases and sad failure, so far, in the case of oligomers which crystallise with several helices in the asymmetric unit.

TABLE 1 OLIGONUCLEOTIDES X-RAY STRUCTURES 1978-1984

Compound	Resolution A	Sp. Gp.	Z	Structure	Reference
d(pATAT)	1.04	$P2_1$	1	B	[8,9]
d(CGCGAATTCGCG)	1.90	$P2_12_12_1$	2	B	[12]
d(CGCGCG)	0.90	$P2_12_12_1$	2	Z	[26,28]
d(CGCG)	1.50	$P6_5$	3	Z	[27]
d(GGTATACC)	1.80	$P6_1$	2	A	[17,18]
d(GGGGCCCC)	2.40	$P6_1$	2	A	[20]
d(^{I}CCGG)	2.10	$P4_32_12_1$	1	A	[22]
d(GGCCGGCC)	2.25	$P4_32_12_1$	1	A	[21]
r(GCG)d(TATACGC)	1.90	$P2_12_12_1$	1	A	[23]

The structures which had been solved yielded, after careful refinement, reliable parameters defining the local and global helical conformations. A good deal of information could also be derived about hydration since a high proportion of the water molecules in the crystal structure were located. However, with the exception of the high resolution $d(CG)_3$ structure no reliable assignment of the counterions could be made. There is scope for technical innovations such as the use of high intensity synchotron sources to obtain stronger diffraction from the very small crystals, and improved methods of structure solution. Above all there is scope for preparing better crystals and this, in some degree, is dependent on even closer collaboration between chemists and crystallographers.

THE DEMISE OF THE UNIFORM DNA STRUCTURE - "ALTERNATING B-DNA"

The first base paired oligonucleotide structure solved was the tetramer d(pATAT) [8, 9]. It is different from others listed in Table 1 in having only one molecule in the asymmetric unit of the crystal. A short, two base pair fragment was, however, formed by Watson-Crick type base pairing between the adenine and thymine bases of neighbouring molecules. The structure was solved to atomic resolution and indicated, for the first time, the precise geometry of this base pairing.

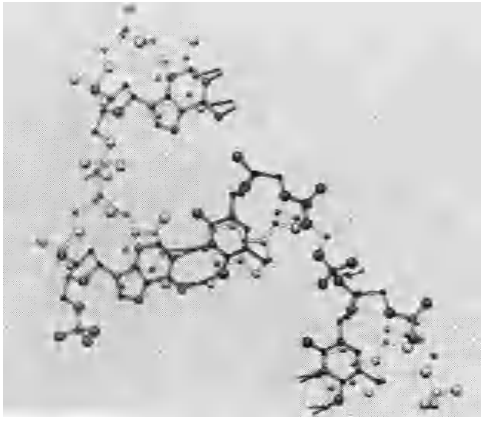


Fig. 4 Two base pair mini helix with alternating sugar conformation observed in the crystal structure of d(pATAT).

The most important aspect of the structure was, however, the presence of two different sugar conformations associated with the pyrimidine and purine moieties : C2' endo and C3' endo respectively. The crystal structure led to the suggestion that a continuous B-DNA helix could be generated not, as had hitherto been supposed, from a single base-pair unit but from a two base-pair unit with alternating sugar conformation [10, 11]. The resultant helix has sequence dependent properties reflected in the base stacking, backbone geometry and sugar conformation. This "alternating B" model was used to interpret some puzzling specroscopic and enzymatic observations and suggested the existence of DNA helices with more than one base pair in the repeat unit.

FULL TURN OF A B-DNA HELIX

There is only one example, so far, of a single crystal study of a B-type helix which was found in the dodecamer d(CGCGAATTCGCG). The analysis, by Dickerson and co-workers [12] was a very careful and extensive one and involved several heavy atom derivatives and measurements at a number of different temperatures [13]. The helix in the crystal structure is right handed with average global helical parameters very similar to those derived from fibre diffraction. The helix is flexible and was observed curved (Fig. 5) or straight, depending on conditions [14].

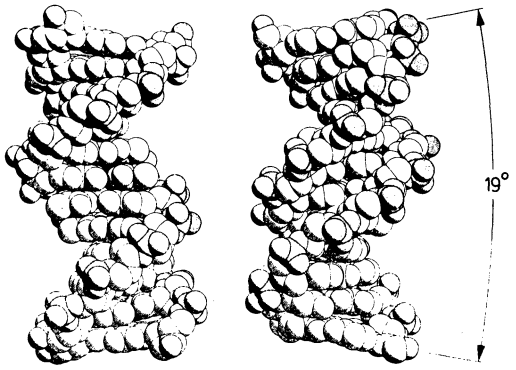


Fig. 5 Two views of the B-dodecamer d(CGCGAATTCGCG).

The local helical parameters calculated from the atomic positions, show twist angles which range from 22-42 degrees, and sugar conformations which varies from C2' endo through C3' endo to O4' exo. The positions of the bases vary with respect to the helix axis with a marked displacement in the central GAATTC region towards the minor groove side, characteristic of A-DNA.

An interesting feature of the structure is the pronounced "propeller twist" of individual base pairs, which has the effect of improving the stacking of bases along each strand of the double helix. The variations in local helical parameters and propellor twist with base sequence have been successfully correlated by Calladine's rules [15, 16] which are based on the principle of elastic beam mechanics. According to these rules a balance is achieved between the propeller twist angles, which maximise base overlaps and the helical turn and base plane roll angles which relieve the steric clash between the large purine bases. Fig. 6 illustrates the agreement between local helical parameters predicted from Calladine's rules and those found experimentally.

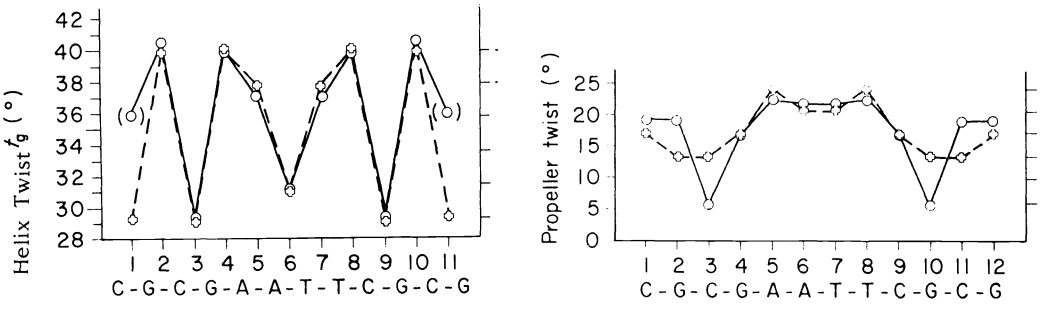


Fig. 6 Agreement between observed (—) and predicted (---) local parameters.

In the extended structure the symmetry related molecules are linked by hydrogen bonding between minor-groove atoms and water molecules, which will be discussed in the section on hydration.

A-DNA IN SINGLE CRYSTAL STRUCTURES

The A-type conformation was found in several oligomers including three octamers, a tetramer and a decamer RNA-DNA hybrid. The conditions for crystallising these compounds were similar to those used for the B-dodecamer and involved the addition of alcohol, thought to determine the A-conformation. It is not at all clear at present what the factors are which stabilise a particular conformation when double helices form an ordered crystalline array.

The structure of A-DNA in the octamer d(GGTATACC)

A brief description of the analysis of this structure and the results obtained [17, 18] is a good illustration of the scope and limitation of present day X-ray methods. The octamer was synthesised by the solid phase method [5] with yields of 8-10mg per preparation. Crystals were grown by vapour diffusion with 2-methylpentane-2,4-diol (MPD) from buffered solutions containing spermine hydrochloride, barium, magnesium or strontium chloride and MPD. The conditions of crystallisation were systematically varied until crystals large enough for diffraction measurement were obtained. Intensity data was collected on a diffractometer and later extended to higher resolution using a high intensity synchotron source.

The octamer crystallised with one double helix in the asymmetric unit of a hexagonal cell. The structure was solved by a specially developed, multidimensional search method which combines packing criteria and best fit (R-factor) calculations for low resolution X-ray data [7]. The uniform A-DNA helix postulated in the trial searches was refined by a constrained-restrained refinement technique [19], making maximal use of the experimental observations to determine the variations in the backbone conformation and the local helical parameters. The analysis is now being extended to more variables, using high resolution synchotron data.

Fig. 7 shows stereoscopic views of the refined octamer duplex. The most striking difference between the fibre model and the actual structure is the opening up of the major groove by nearly 2Å as a result of the symmetrical bending, by 15°, about the non-crystallographic two fold axis. This bend corresponds to a radius of ~80Å demonstrating, as did the B dodecamer structure, the intrinsic flexibility of the DNA helix. The width of the minor groove remains nearly constant.

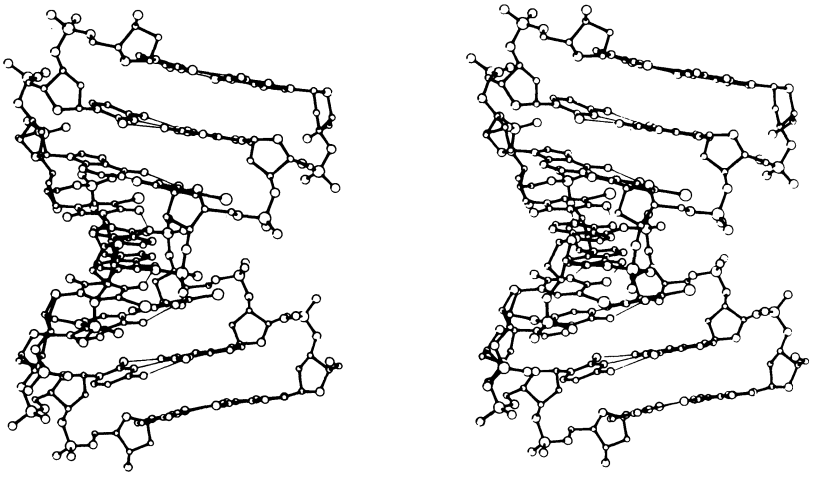


Fig. 7 Stereoscopic view of the octamer duplex

A detailed analysis of the local helical parameters shows systematic variations with base sequence which are most pronounced in the central TATA region. This sequence is of particular interest since it is found at the binding site of RNA polymerase prior to transcription. Fig. 8 compares the base overlaps in the first four steps of the refined octamer with those calculated from a uniform A-DNA fibre model. The overlaps at the first two G-C base pairs are very similar to the fibre model. The T-A step shows the interstrand overlap characteristic of A-DNA between the adjacent adenine bases. The greatest difference is at purine-pyrimidine steps G-T. Here the G.C base pair is shifted relative to the T.A base pair by nearly 1Å towards the major groove, resulting in a parallel arrangement of the keto group which is energetically more favourable. The variations in base stacking have the overall effect of evenly distributing the stacking energy along the helix axis, and can be correlated with the variations in the local helical parameters.

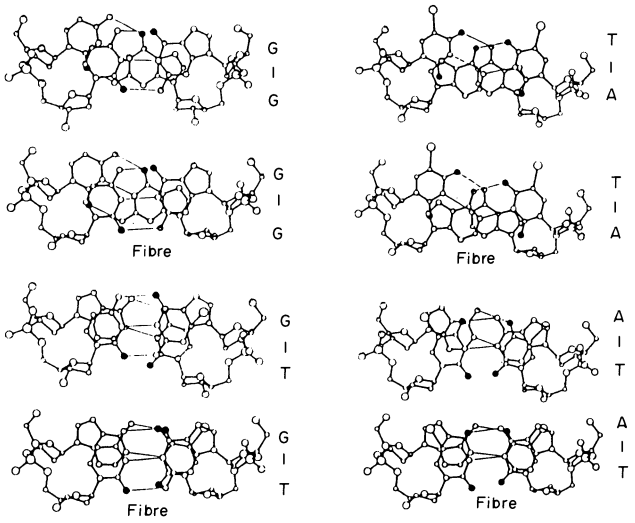


Fig. 8 Projected views of the first four steps of the octamer compared with the equivalent steps from the A-DNA fibre model. The view is perpendicular to mean plane through the upper bases

The extended crystal structure is formed by packing the octamer duplexes in infinite spirals about the crystallographic 6_1 axis leaving large central cavities filled with solvent molecules and (probably) with counterions. Adjacent spirals related by 2_1 screw axes interleave, forming continuous cylinders. The intimate packing of adjacent molecules is clearly shown in Fig. 9 with the end G.C base pairs parallel to, and in van der Waals contact with, the sugar phosphate backbone of symmetry related molecules. The structure is further stabilised by a pair of intermolecular hydrogen bonds. The only other intermolecular hydrogen bonds involve water molecules and will be discussed in the section on hydration.

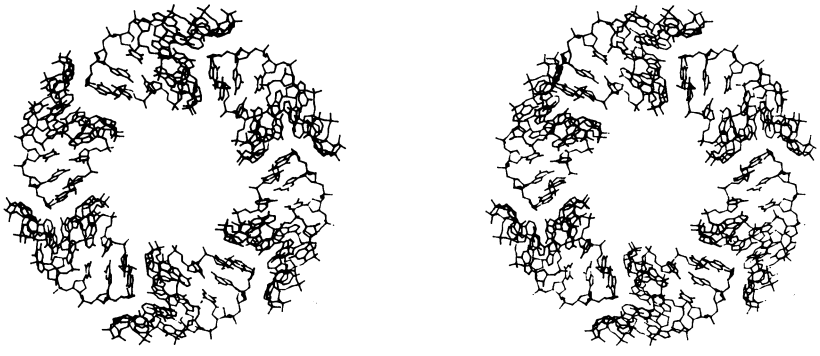


Fig. 9 Stereoscopic view of the packing of the octamer duplexes along the sixfold axis.

Comparison of the A structure

The A conformation was found in a number of crystal structures (Table 2). One of these, the octamer d(GGGGCCCC) [20] is isomorphous with d(GGTATACC). The octamer d(GGCCGGCC) [21] and the tetramer (d^{I}CCGG) [22] crystallise isomorphously in a tetragonal rather than a hexagonal space group but using the same packing motif as the hexagonal crystals with large cavities, accomodating solvent molecules. Interestingly the tetramer d(^{I}CCGG) behaves as if it were an octamer with two tetramers stacked on top of one another but with the connecting phosphate missing. It seems as if in these A type octamers the ability of the end base-pair to stack against the shallow minor groove is the dominant factor in stabilising the extended crystal structure.

TABLE 2 CRYSTAL DATA FOR A-TYPE STRUCTURES

Compound	Cell A			Volume A^3	Volume/base pair A^3
	a	b	c		
d(GGTATACC)	45.01	45.01	41.55	72899	1519
d(GG^{Br}UA^{Br}UACC)	45.08	45.08	41.72	73425	1530
d(GGGGCCCC)	45.32	45.32	42.25	75151	1566
d(^{I}CCGG)	41.10	41.10	26.70	45102	1409
d(GGCCGGCC) [-8°C]	42.06	42.06	25.17	44527	1392
d(GGCCGGCC) [-18°C]	40.51	40.51	24.67	40485	1265
r(GCG)d(TATACGC)	24.20	43.46	49.40	51956	1299

The A-type structures form a particularly interesting set for a detailed examination of sequence dependent structural variations. The sequence d(TATA) is contained in both d(GGTATACC) and in the RNA-DNA hybrid r(GCG)d(TATACGC) [23] thus allowing comparison of the same sequence in two different crystal environments. On the other hand the two isomorphous oligomers d(^{I}CCGG) and d(GGCCGGCC) contain the recognition sites CCGG and GGCC of the two restriction endonucleases Hpa II and Hae II respectively and can thus be compared in the same environment. Fig. 10 illustrates the overall changes in the double helix in these crystals compared with the fibre model. A very detailed comparison of the A-type structure has recently been carried out [24] and followed by a quantitative examination of base stacking using atom-atom potential energy calculation [25]. Such comparative studies are important in evaluating the relative contribution of intramolecular and crystal forces on the stability of helix conformations and in trying to understand the rules governing sequence dependent modulations.

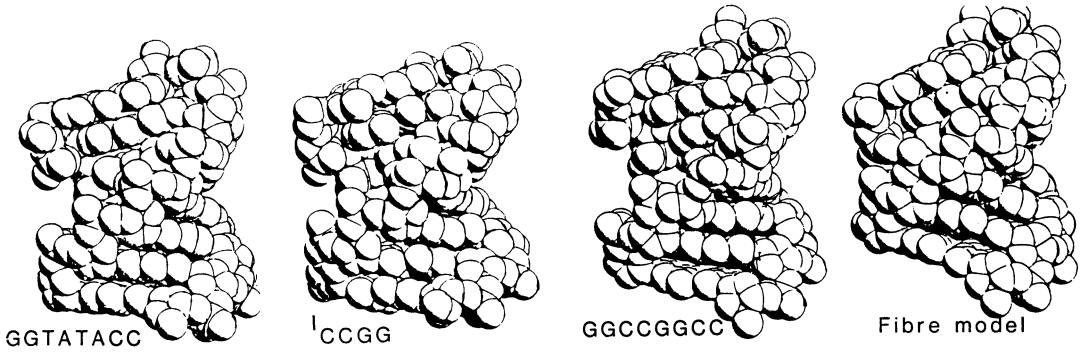


Fig. 10 Comparison of space filling drawings of the three A-DNA structures with the fibre model.

Z-DNA - A NEW LEFT HANDED HELIX

The most dramatic result to have emerged from single crystal analysis of oligonucleotides is the radically different, left handed helix. The Z helix was discovered by Rich and co-workers from a high resolution study of the hexamer d(CGCGCG) [26]. The same helix was found shortly after in two forms of the tetramer d(CGCG) by Dickerson and colleagues [27] and by Rich and his collaborators [28]. Left handed helices had previously been postulated on both theoretical and experimental grounds, particularly from solution studies of synthetic poly (dG-dC) [29] but could not be accepted with any confidence since no detailed structural model was available.

The repeat unit of the Z helix consists of two base pairs which differ in the sugar-base orientation. For purines the orientation is syn but for pyrimidines it is anti. The sugar conformations are different, like in the "alternating B" structure, and so is the orientation about the C4'-C5' bond (Fig. 11). Base stacking is unusual (Fig. 12), with no intrastrand stacking at the CpG step but only interstrand stacking of the cytosines.

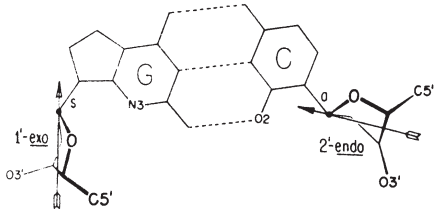


Fig. 11 Structure of a G.C base pair in Z-DNA.

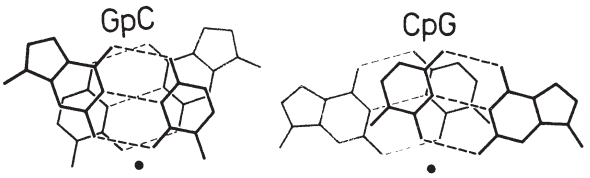


Fig. 12 Stacking of adjacent bases in left handed poly(dG-dC) poly(dG-dC).

The two base-pair unit is repeated six times to form a full turn of a left handed helix. The alternating syn and anti arrangement implies that the neighbouring sugar units point alternately up and down and as a consequence the line connecting the phosphorous atoms follows a zig-zag course (Fig. 13). As in "alternating B-DNA" the phosphates in sequence d(CpG) and d(GpC) have different chemical environments. The long and relatively thin Z-helix has no major groove but only a concave surface with the cytosine C5 and the guanine N7 and C1 atoms exposed. The minor groove side, which accommodates the helix axis is deep, narrow and lined with phosphate groups.

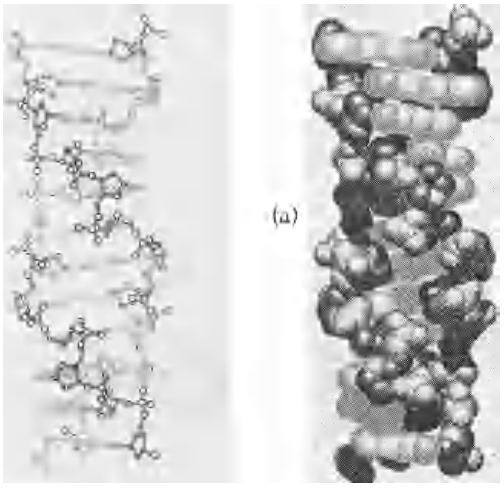


Fig. 13 Structure of Z-DNA extrapolated from the coordinates of the hexanucleotide d(CGCGCG) (a) side views (b) top views

Z-DNA has been shown also to exist in solution using CD and NMR measurements, and in fibres by X-ray diffraction. It is highly antigenic and can be detected in Drosophila chromosomes when extracted with 45% acetic acid [30], but its biological function is obscure [31]. Changing the helical sense of sections of DNA will markedly affect supercoiling in higher order assemblies and so could be involved in regulating replication and expression of genes.

HYDRATION IN SINGLE CRYSTALS OF OLIGONUCLEOTIDES

All the crystals that have been analysed contain approximately 40-60% DNA per cell volume, the remaining space being taken up by water and counterions. An examination of difference maps at the end of the refinement of the double helical molecule alone allowed the location of a substantial number of water molecules. Typical difference maps showing the location of water peaks in the B DNA [34] dodecamer are reproduced in Fig. 14.

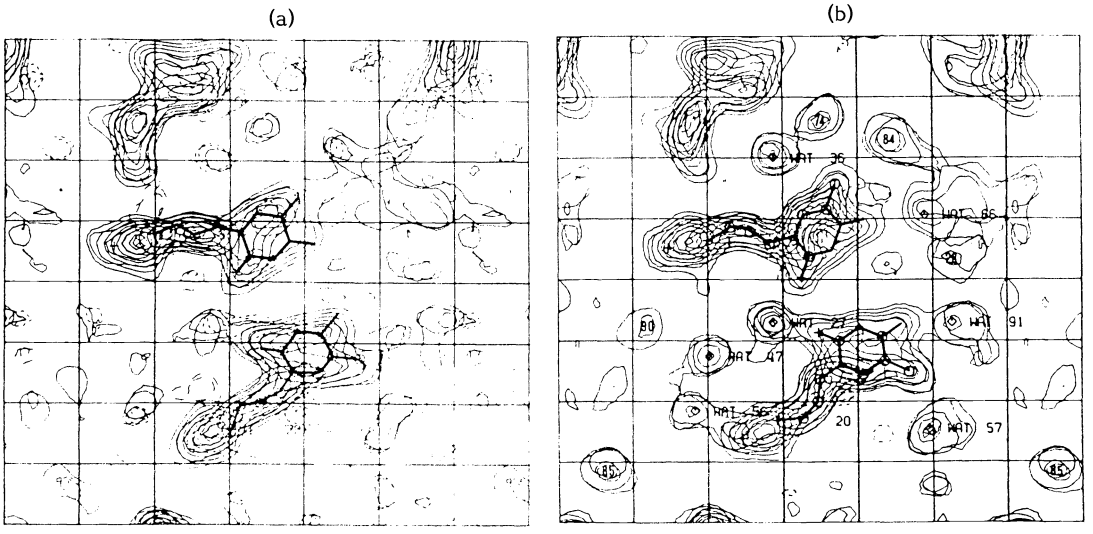


Fig. 14 Combined Fourier and Fourier difference maps for the dodecamer d(CGCGAATTCGCG). (a) first stage of refinement (b) last stage of refinement and addition of water molecules.

Water in crystal structures generally forms either monodentate, bidentate or three centred bonds. The hydrogen bondm being an electrostatic interactions, has variable geometry with O-H...X distances of 2.5-3Å and O-H...X angles of 150°-180°. When oxygen acts as a hydrogen acceptor the X-H...O bond tends to lie along the lone pair directions. Water has a strong tendency to form the maximum number of hydrogen bonds. Thus four-coordinated water is surrounded ny an approximate tetrahedral arrangement of donor and acceptor groups. These guidelines have been used to interpret the water peaks in the difference maps. In view of the relatively low resolution of the analysis, however, a tolerance of about 0.25Å has been allowed in bonf length thus limiting the distance between H-bonded atoms to about 3.5Å. Atoms at a distance greater than say 3.7Å can not, with any certainty, be regarded as part of an H-bonded network. It should also be remembered that the limited resolution precludes the identification of all molecules in the crystal structure especially the disordered water in, for example, the large cavities of the A-type crystal. In these spaces a number of alternative water networks may exist and the X-ray analysis gives a time averaged view of several superimposed structures.

Infra-red spectroscopic studies [32, 33] indicate that the primary hydration shell around double helical DNA contains at least 11-12 water molecules per nucleotide, rising to about 20 at a relative humidity of 80%. The phosphate oxygen atoms have the highest affinity for water, followed by the phosphodiester and sugar oxygen atoms. Above 65% relative humidity the functional groups of the bases are also hydrated, the entire primary hydration shell forming a relatively ordered betwork with about half the water oxygen atoms hydrogen bonded to the DNA molecule and the remaining forming cross links. The crystal structures offer an excellent opportunity of comparing the hydration of the various atoms in different types of helices and correlating these with solution studies. An examination of the extended water structure in the unit cell may also give an indication on the partial order imposed on an aqueous environment by the hydrophobic and hydrophilic atomic grouping in the double helix.

Hydration in B-DNA - the dodecamer

The number of water molecules which could be located with certainty in the B-dodecamer was dependent on experimental conditions. At room temperature 72 peaks could be assigned (of which some were thought to be spermine) [34] and a further 48 appeared in difference maps when the study was carried out at 16°K or when the MPD concentration in the crystallisation mixture was increased to about 60% [35]. Under these conditions between one and three water molecules are found around most phosphate oxygens, of which about 40% coordinate to both phosphates oxygens. Very few of the phosphate ester atoms are hydrated.

The most striking feature of the B-DNA hydration is the ordered water structure found in the minor groove and particularly in the central AT rich region. As illustrated in Fig. 15 water molecules link the functional groups of adjacent steps on the two strands. These water molecules are in turn linked by further water molecules and a chain of nine linked water molecules extends along the centre of the minor groove. The network is further stabilised by coordinating with the sugar O4' atoms. This primary hydration shell becomes less well ordered towards either end of the helix. A higher hydration shell links with the primary shell and involves some of the phosphate oxygen atoms.

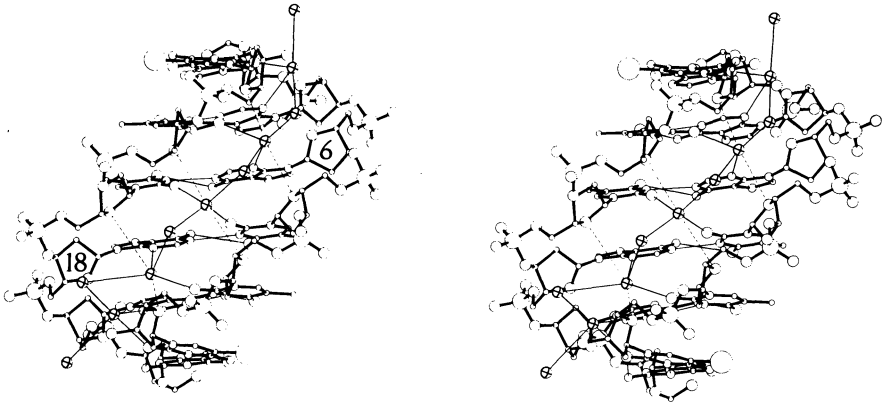


Fig. 15 Stereo diagram of the minor groove hydration in the GAATTC segment of the B-dodecamer

The water structure in the major groove is discontinuous and much less regular than in the minor groove (Fig. 16). Only relatively few water molecules were located at distances of 3.5Å or less from the functional groups and only two of these bridge functional groups of the bases between steps directly. A second and equally irregular hydration shell links some of these water molecules and also the phosphate oxygen atoms pointing into the major groove.

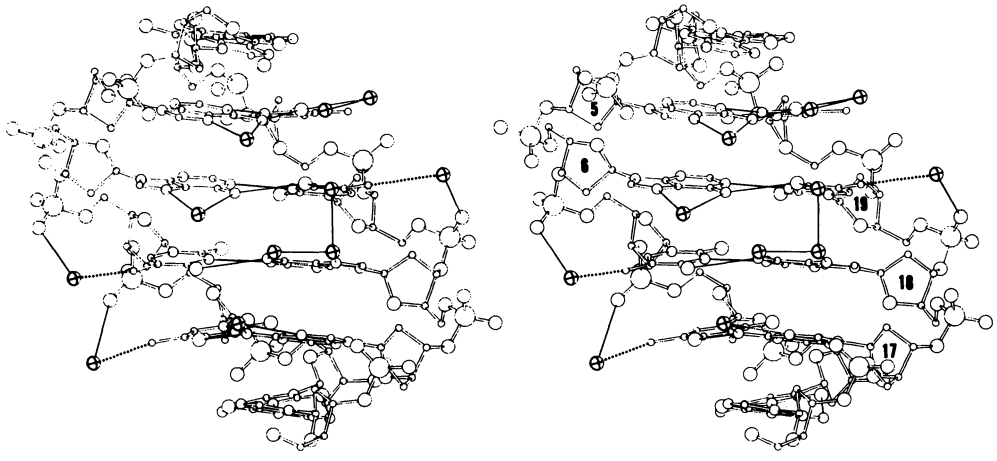


Fig. 16 Stereo diagram of the major groove hydration in the AT rich segment of the B-dodecamer.

Hydration in A-DNA

A large percentage of solvent molecules were located in the A-DNA structures. Some details of the water structure were published for the tetramer d(^{I}CCGG) [22] and the analysis of the hydration network in d(GGBrUABrUACC) has recently been completed [36].

In the octamer virtually all the phosphate O1 and O2 atoms are hydrated. Neighbouring O1 atoms along the backbone are frequently bridged by one water molecule but none of the water molecules are linked to the two oxygen atoms of the same phosphorous atom.

The most remarkable feature of the first hydration shell is the ribbon of water molecules extending along the centre of the major groove (Fig. 17). This ribbon is formed by pentagonal arrangements of five water molecules, each tetrahedrally coordinated either to the functional groups of the bases or to the water molecules which in turn are linked to the phosphate O1 atoms pointing into the major groove. The network has pronounced two-fold symmetry, particularly in the central TATA region. The network becomes less regular towards either end of the groove. Pentagonal arrangements involving water were observed in small molecule structures, notably in a complex of d(CpG) and proflavine [37] and in hydrated α cyclodextin [38, 39].

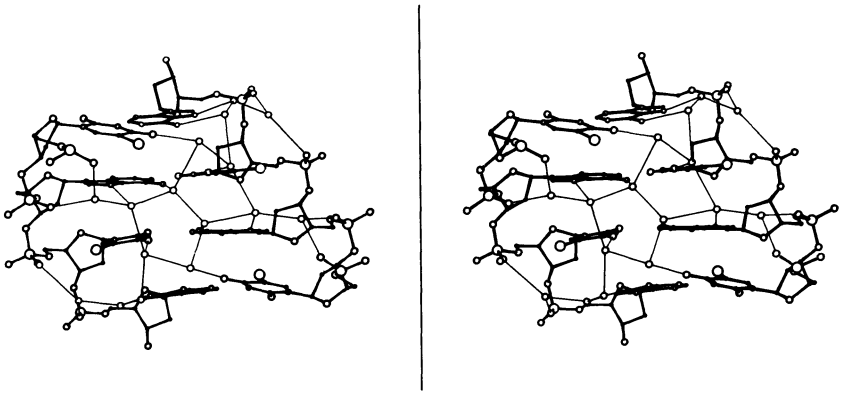


Fig. 17 Stereo diagram of the hydration in the major groove of the ATA region of the A-octamer.

The minor groove in d(GGBrUABrUACC) is less accessible to solvent molecules than the major groove because the end pairs of one helix are stacked in the minor groove of a symmetry related helix. Nevertheless, most of the functional groups in the minor groove are hydrated (Fig. 18). The network of water molecules is much less regular than in the major groove and the ordered chain of water molecules, observed in the AT rich region of the minor groove of the B-dodecamer, is not seen. However, there is a well defined chain of water molecules in the central ATA region, where hydrogen bonds are formed to the bases of opposite strands and also to the sugar O4' atom pointing into the minor groove. Less regular networks cross link the functional group of the bases at either end of the octamer.

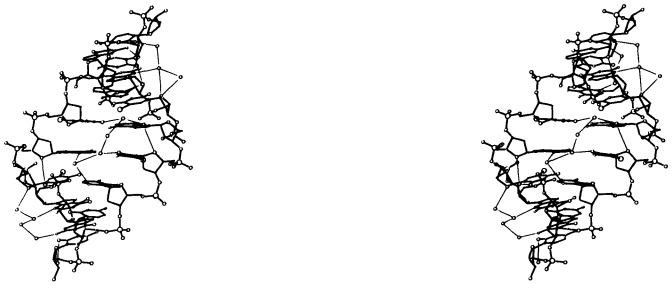


Fig. 18 Stereo diagram of the hydration in the minor groove of the A octamer.

In the other A-type structure d(ICCGG) [22, 40] where the water structure was analysed in some detail, the hydrogen bonded network in the major groove is much more irregular than in d(GGTATACC) possibly because of the intrusion of the terminal OH groups after only four base pairs along the helix. The primary hydration shell consists mostly of local clusters of water molecules coordinating the functional groups and providing intrastrand links. There is only one hydrogen bonded network stretching from the phosphate oxygen of one strand via various functional groups to a phosphate oxygen atom of a symmetry related molecule. In the minor groove very few of the functional groups are hydrated, possibly again because of end effects.

In the octamer d(GGCCGGCC) eighty water molecules were located mostly in the large solvent channels near the major groove. No detailed analysis of the water is as yet available. The position is similar for the hydration in the Z-structures.

It has been suggested that the regular water molecules in the minor groove of B-DNA stabilise the conformation. On dehydration this regular structure is disrupted and only the highly polar phosphate oxygen atoms remain hydrated. This induces a conformational transition and the A-DNA form is stabilised by the hydration in the major groove. The very regular ribbon of water molecules found in the A-octamer structure lends support to the view that the A-conformation is indeed maintained by the hydration of the major groove.

SUMMING UP

The behaviour of DNA in vivo is very complex and becomes apparently more complex as it is investigated in detail. The same seems to hold true for the structure of DNA. The handful of single crystal analyses of deoxyoligonucleotides published in the last few years have made us realise the structural flexibility of DNA, the existence of left handed as well as right handed helices, helices generated from a variable number of repeat units and variable local conformations. We are ready to accept that the classical B-structure is but one of the conformations adopted in vivo and have ourselves become more flexible when interpreting biochemical, as well as physiochemical, experiments on DNA in structural terms.

The search for unifying principles and predictive rules is one of the driving forces behind most scientific investigations. There are already tantalising glimpses of such principles, which might account for some of the observed structural variability. Local conformational changes in particular seem to be governed by base stacking and the tendency for equipartitioning the stacking energy along the entire helix. The different hydration patterns observed in the B and A type structures have suggested some theories about the role of water in mediating conformational transitions. However, these generalisations are based on structural data from just a few short, self-complementary DNA fragments. We need to analyse many more deoxyoligonucleotides of varying length and base composition, selected to answer specific questions. New information will also come from the rapidly expanding field of single crystal analysis of DNA-protein and DNA-drug complexes, and the wide variety of solution studies particularly high resolution nuclear magnetic resonance measurements. It seems a safe prediction that this wealth of data will lead to a new simplification in our understanding of the mode of action of DNA in biological systems.

ACKNOWLEDGEMENTS

I thank my colleagues Dr. T. Brown, Dr. G.G. Kneale, Dr. M.J. McCall, Dr. W.B.T. Cruse, Mr. J. Nachman, Professor D. Rabinovich, Dr. S. Salisbury, Dr. Z. Shakked and Professor M.A. Viswamitra for their collaboration during the period covered by this survey; Dr. M.J. McCall and Dr. S. Salisbury for their helpful comments on the manuscript and Mrs. S. Berry for all her work in preparing the camera ready copy. Financial support was provided by the Medical Research Council.

REFERENCES

1. J.D. Watson and F.H.C. Crick, Nature (Lond) 171, 737-740 (1953).
2. S. Arnott and D.W.L. Hukins, Biochemical and Biophysical Research Communications 47, 1504-1509 (1972).
3. J. Stravinsky, T. Hozumi, S.A. Narang, C.P. Bahl and R. Wu, Nucl. Acid Res. 4, 353-371, (1977).
4. a) S.L. Beaucage and M.H. Caruthers, Tet. Lett. 22, 1859-3661 (1981). b) R.L. Letsinger and W.B. Lunsford, J. Amer. Chem. Soc. 98, 3655-3661 (1976).
5. M.J. Gait, H.W.D. Mathes, M. Singh, B.S. Sproat and R.C. Titmas, Nucl. Acid Res. 10, 6243-6248 (1982).
6. M. Levitt and A. Warshel, J. Am. Chem. Soc. 100, 2607-2613 (1978).
7. D. Rabinovich and Z. Shakked, Acta Cryst. A. In press (1984)
8. M.A. Viswamitra, O. Kennard, P.G. Jones, G.M. Sheldrick, S.A. Salisbury and L. Falvello, Nature (Lond) 273, 697-688 (1978).
9. M.A. Viswamitra, Z. Shakked, P.G. Jones, G.M. Sheldrick, S.A. Salisbury and O. Kennard, Biopolymers 21, 513-532 (1982).
10. M.A. Viswamitra, O. Kennard, Z. Shakked, P.G. Jones, G.M. Sheldrick, S.A. Salisbury and L. Falvello, Curr. Sci. 9, 289-292 (1978).
11. A. Klug, A. Jack, M.A. Viswamitra, O. Kennard, Z. Shakked and T.A. Steitz, J. Mol. Biol. 131, 699-680 (1979).
12. R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, Nature (Lond) 287, 755-758 (1980).
13. H.R. Drew, S. Samson and R.E. Dickerson, Proc. Nat. Acad. Sci. USA 79, 4040-4044 (1982).
14. A.V. Fratini, R.L. Kopka, H.R. Drew and R.E. Dickerson, J. Biol. Chem. 257, 14686-14707 (1982).
15. C.R. Calladine, J. Mol. Biol. 161, 343-352 (1982).
16. R.E. Dickerson, J. Mol. Biol. 166, 419-445 (1983).
17. Z. Shakked, D. Rabinovich, O. Kennard , W.B.T. Cruse, S.A. Salisbury and M.A. Viswamitra, Proc. Roy. Soc. 213, 479-487 (1981).
18. Z. Shakked, D. Rabinovich, W.B.T. Cruse, E. Egert, O. Kennard, S.A. Salisbury and M.A. Viswamitra, J. Mol. Biol. 166, 183-201 (1983).
19. S.L. Sussman, S.R. Holbrok, R.W. Warrant, G.M. Church, S.A. Kim, Acta Crystal Org. Set A 33, 800-804 (1977).
20. M.J.M. McCall and O. Kennard. To be published (1984).
21. A.H.J. Wang, S. Fujii, J. van Boom and A. Ritch, Proc. Nat. Acad. Sci. U.S.A. 79, 3968-3972 (1982).
22 B.N. Conner, T. Takano, S. Tanaka, K. Itakura and R.E. Dickerson, Nature (Lond) 295, 294 299 (1982).
23. A.H.J. Wang, S. Fujii, J.H. van Boom, G.A. van der Marcel, S.A.A. van Boeckel and A. Rich, Nature (Lond) 299, 601-604 (1982).
24. Z. Shakked and O. Kennard, Structural Biology, Eds. A. McPherson and F. Journak, John Wiley and Sons, N.Y. In press (1984).
25. T.E Haran, Z. Berkovitch-Yellin and Z. Shakked. Submitted for publication (1984).
26. A.H.J. Wang, G.J. Quigley, F.J. Kolpak, J.L. Crawford, J.L. van Boom, G. van der Marcel and A. Rich, Nature (Lond) 282, 680-686 (1979).
27. H. Drew, T. Takano, S. Tanaka, K. Itakura and R.E. Dickerson, Nature (Lond) 286,567-573 (1980).
28. J.L. Crawford, F.J. Kolpak, A.H.J. Wang, G.J. Quigley, J.H. van Boom, G. van der Marcel and A. Rich, Proc. Nat. Acad. Sci. 77, 4016-4020 (1980).

29. F.M. Pohl and T.M. Jovin, J. Mol. Biol. 67, 375-396 (1972).
30. R.M. Sandella, D. Grunberger, I.B. Weinstein and A. Rich, Proc. Nat. Acad. Sci. U.S.A. 78, 1451-1455 (1981).
31. R.J. Hill and B.D. Stollar, Nature (Lond) 305, 338-340 (1983).
32. M.J. Falk, K.H. Hartman, Jr and R.C. Lord, J. Amer. Chem. Soc. 85, 387-391 (1963).
33. M. Falk, A.G. Poole and C.G. Goymour, Can. J. Chem. 48, 1536-1542 (1970).
34. H.R. Drew and R.E Dickerson, J. Mol. Biol. 151, 535-556 (1981).
35. M.L. Kopka, A.V. Fratini, H.R. Drew and R.E. Dickerson, J. Mol. Biol. 163, 129-146 (1983).
36. O. Kennard et. al. In preparation.
37. S. Neidle, H.M. Berman and H.S. Shieh, Nature (Lond) 288, 129-133 (1980).
38. W. Saenger, Nature 279, 343-344 (1979).
39. W. Saenger and K. Lindner, Agnew. Chem. Int. Ed. Engl. 19, 398-399 (1980).
40. R.E. Dickerson. personnal communication.

Figs. 1, 2, 3, 5, 11, 12, 13 were reproduced from W. Saenger Principles of Nucleic Acid Struture, Springer Verlag, New York, 1984 and Figs. 6, 14, 15, 16 from J. Mol. Biol. by kind permission of the Editors.