

ICA News letter, 2003-2004

DNA Crystallography

N. Gautham

Department of Crystallography and Biophysics

University of Madras, Chennai 600 025

Introduction

The characterization of sequence effects in DNA structure has received much attention from DNA crystallographers over the past two decades (Dickerson, 1998). These studies have led to the realization that within the framework of the double-helix proposed by Watson and Crick, the molecule can show a great deal of variation in its detailed structure (Calladine & Drew, 1997). It is not however clear whether many of the dramatic effects seen are only a result of crystal packing forces or whether they are in fact relevant to the biological processes which involve those sequences. Opinion has converged on exploring the sequence effects on the basis of the following two hypotheses: (a) DNA is not a passive molecule, just storing the information that is read by proteins. On the contrary it is an active player in the game of life. It plays its role by changing its three dimensional structure according to sequence, some of these changes being quite far from the Watson Crick double helix, e.g. left-handed helical structures, intercalated structures and quadruplexes (Wells & Harvey, 1988). (b) DNA is a plastic molecule with a sequence-dependent plasticity that changes the local structure of the DNA while retaining the overall double-helical form. It is this feature that allows some of the DNA sequences to act as regulators and modulators of transcription and participate actively in biological development processes. In both cases X-ray crystallography is expected to give information on the basis of which the theories can be elaborated and made more detailed and precise. In our laboratory we have been using this technique to obtain the structures of oligonucleotide sequences chosen to answer very specific questions. In the first place the sequences have been chosen to be non-self-complementary or as close as possible to natural sequence DNA. A look at the NDB (Berman, et al., 1992) reveals that most of the 700 odd structures deposited in the database are of self-complementary sequences. It is possible that the symmetry of the sequence hides effects that may be present in natural DNA sequences that are only rarely symmetrical. To elucidate these effects and to remove any artefacts caused by the symmetry,

we have used non-self-complementary sequences. With this design principle broadly underlying our efforts, we are currently working on two projects. The first one is a study of sequence effects in left-handed DNA. The second project is an attempt to find any correlation that may exist between the exact sequence of consensus regions of promoters, their structure and the promoter strength.

Sequence effects in left-handed DNA

We have analysed the effects of A.T base pairs in Z-DNA in a systematic manner. Left-handed Z-DNA was first noticed in crystal structures of the sequences CGCGCG and CGCG. We have chosen to first study the effects of replacements of the CG base pairs by AT base pairs in the hexamer. Due to the symmetry of the sequence there are only three ways in which a single CG base pair can be replaced by AT. Such a substitution would yield one of the three non-self-complementary duplex sequences d(TGCGCG).d(CGCGCA), d(CACGCG).d(CGCGTG) and d(CGTGCG).d(CGACAG). We obtained synthetic samples of these sequences, annealed them in appropriate pairs to form the duplexes mentioned above and then grew crystals of each of them. The first sequence crystallized in the C2 space group with four times the volume of the unit cell of the 'mother' sequence, namely d(CGCGCG)₂ (Wang, et al., 1979). Structure solution by the molecular replacement method using the program AMoRe showed that the duplexes form infinite columns along the 'c' axis that are packed together in a hexagonal close packed pattern. Refinement of this structure is in progress and we expect to uncover fine changes in the structure that might explain the different space groups that occur in spite of the similar packing. An interesting crystallographic curiosity arose in the experiments on d(TGCGCG).d(CGCGCA). In an attempt to grow crystals more amenable to X-ray studies, we tried to crystallise this sequence in the presence of Co(NH₂)₆Cl₃, known to be strong promoter of left handed Z DNA. To our surprise and delight we obtained ring crystals (Figure 1, Kumar & Gautham, 1999). These are the first reports of such crystals for any material. The crystals were however too small to be subject to X-ray diffraction analysis.

The other two duplexes in this series have been solved and refined (Sadasivan & Gautham, 1995). The sequence d(CACGCG).d(CGCGTG) packs in exactly the same cell as d(CGCGCG)₂. The structure was solved by molecular replacement using XPLOR and refined to an R factor of 19.9% for data to a resolution of 1.6 Å. Like the crystal packing, the molecular structure also showed only insignificant differences when compared to d(CGCGCG)₂. This sequence was also

crystallized in the presence of $\text{Ru}(\text{NH}_2)_6\text{Cl}_3$ (Karthe & Gautham, 1998). The use of this metal ion instead of BaCl_2 to stabilize the structure did not introduce any changes either to the crystal packing or to the molecular structure. Since in neither case the metal ion (i.e. neither Ba nor Ru) was observed in the electron density even at the relatively high resolution of these crystals, it was concluded that the metal, in these cases at least, plays a non-specific role in stabilizing the structure. The third duplex, viz. $d(\text{CGTGCG}).(\text{CGCACG})$ also had approximately the same packing as the others – the duplexes were stacked above each other in infinite columns which were then put together in a hexagonal close packed arrangement. However, as in the case of the first sequence, in this case too, the space group was different from that of $d(\text{CGCGCG})_2$. The space group was determined with great difficulty, which was due to two, probably related, reasons. The first reason for the difficulty arose from the fact that the crystal diffracted only to a resolution of 2.5 Å. At this resolution the molecule may be approximated to a cylinder and a single packing mode can be indexed equally well in a number of related space groups (Sadasivan, Karthe & Gautham, 1994). The second reason was that, as observed from the solved structure, changes in the molecular conformation led to subtle changes in the packing that gave rise to a different space group. We could overcome these problems and arrive at the correct space group by an analysis of the reciprocal lattice symmetry (Sadasivan, Karthe & Gautham, 1994). The structure was then solved by molecular replacement. The final structure was quite different from that of both $d(\text{CACGCG}).d(\text{CGCGTG})$ and $d(\text{CGCGCG})_2$ (Figure 2). Comparing our results with those obtained elsewhere from Raman spectroscopy of similar sequences (Wang, Thomas & Peticolas, 1987) we could conclude that the model of left handed Z-DNA obtained from the crystal structure of $d(\text{CGCGCG})_2$ requires a continuous stretch of four CG base pairs for stability.

In the same series we have obtained crystals of $d(\text{TGCGCA})_2$ and also of the sequence $d(\text{AGCGCT})_2$. These sequences have two AT base pairs each and the second one, in addition, has a perturbation in the pyrimidine-purine alternation thought to be necessary for Z-DNA. The former sequence crystallizes in the same space group as $d(\text{CGCGCG})_2$ and has an almost identical structure. The latter sequence yields very beautiful platy crystals that diffract very well but the pattern cannot be indexed satisfactorily. It is possible that the crystals are twinned. The crystal structure of $d(\text{TGCGCA})_2$ has been solved and refined at 296 K and at a resolution of 1.6 Å. The molecule adopts a left-handed Z type helical conformation, common for alternating pyrimidine-purine sequences. The presence of A.T base pairs at the two terminals does not perturb the structure

to any great degree. However, several sequence-specific micro-structural changes are noticeable. The structure of the identical sequence determined at 120° K has been reported previously (Harper et al., 1998). A comparison of the present structure with the low temperature structure shows that the effect of the sequence on the micro-structural variations is significantly greater than the effect of the temperature.

This effect of out-of-alternation sequences has been further probed in the structure of (CCCGGG)₂. This sequence, perhaps surprisingly, crystallizes in the same cell as d(CGCGCG)₂. It diffracts to a resolution of 2.5 Å. The refined structure is a left-handed helix with a conformation that cannot be classified strictly as Z-DNA. This crystal structure indicates that, like the right-handed helices, left-handed helices too may be polymorphic. Further characterization of these effects is in progress. Other sequences under study include brominated d(CCCGGG), in the expectation that a clearer and more convincing view of the structure may be obtained.

The structural basis of promoter activity

Another ongoing project on DNA crystallography in our laboratory seeks to elucidate the role that sequence-specific structure may play in the regulation of transcription. We have focussed our attention on promoter sequences, in particular the sequences at the -10 and -35 positions. The sequences in these regions are known to be highly consensual, with TATAAT (-10) and TTGACA (-35) as the consensus sequences. One of the determinants of promoter strength is the degree of homology of the promoter sequence with the consensus sequences (Youderian, Bouvier & Susskind, 1982). The aim of the project is to arrive at the answer as to how the sequence may modify the structure and how these effects may together affect promoter strength.

Towards this end we have designed and crystallized a few dodecamer sequences that contain either the consensus sequence or one of its variants. From the point of view of crystallography, the sequences had to satisfy the following criteria: They had to be capable of crystallizing easily; they had to pack in ways which would not affect the analyses of the results in terms of the DNA structure alone, ignoring crystal effects; they had to be as far as possible in the B-type structure, which is thought to be the most common form. We chose the Dickerson sequence as the basis on which we constructed the sequence. This is CGCGAATTCGCG, a self-complementary dodecamer, which was the first B-type structure to be solved (Dickerson, et al., 1982). In the crystal, contacts between neighbouring duplexes are only at the ends, leaving the

central six base pairs free to adopt unconstrained conformations (within the overall structure of a B-type double helix). We replaced the central six bases with the promoter consensus sequences and their variants, and obtained a set of non-self-complementary sequences that we used in our crystallization trials. We have obtained crystals of the following sequences.

- a) d(CGCTATAATGCG).d(CGCATTATAGCG). This is the consensus -10 sequence of the prokaryotic promoters.
- b) d(CGCTATGTTGCG).d(CGCAACATAGCG). This is the -10 region of the *lac* promoter.
- c) d(CGCTTTAATGCG).d(CGCATTAAAGCG). This is the -10 region of the artificially constructed *tacII* promoter.
- d) d(CGCTTGACAGCG).d(CGCTGTCAAGCG). This is the consensus sequence of the prokaryotic promoter -35 region.
- e) d(CGCTTAACTGCG).d(CGCAGTTAAGCG). This is the -10 region of the *trp* promoter.

We have also tried crystallization experiments on the following decanucleotides. Decamers have been shown to crystallize in a more ordered way and therefore diffract to higher resolution. However, crystal-packing effects also are greater and the true influence of the sequence on the structure is likely to be masked to a larger extent. Nevertheless, important information regarding the structure of the promoter sites is likely to accrue from their study, especially since they can be compared with the dodecamer structures to eliminate effects due to packing.

- a) d(CCTATAATGG).d(CCATTATAGG). This is the consensus -10 sequence of the prokaryotic promoter.
- b) d(CCTATGTTGG).d(CCATACAAGG). This is the -10 region of the *lac* promoter.
- c) d(CCTTAACTCG).d(CCAGTTAAGG). This is the -10 region of the *trp* promoter.

2'-5' linked oligonucleotides

Natural DNA has a 3' to 5' phosphodiester linkage between one nucleoside and the next. However several studies (Lalitha & Yathindra, 1995) have shown that 2' to 5' linkages can also take up double helical structures which are very similar to the A and B type helices of the more usual DNA. In collaboration with Professor Yathindra of our department we have grown crystals of a self-complementary dodecanucleotide with 2'-5' linkage. Macroseeding techniques were used to

improve the size of the crystal. Data up to a resolution of 2.8 Å have been collected at the National Area Detector Facility in Bangalore. The space group is P6₁22 with a = b = 46.94 Å and c = 126.65 Å. Structure solution has been hampered by lack of an appropriate model and by the large unit cell and low resolution data.

References:

- Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, J., Hsieh, S.-H., Srinivasan, A.R. & Schneider, B. (1992). *Biophys. J.*, **63**, 751-759
- Calladine, C.R. & Drew, H.R. (1997). *Understanding DNA. The Molecule and How It Works*. Academic Press, San Diego, California, USA
- Dickerson, R.E. (1998). *Nucl. Acids. Res.*, **36**, 1906-1926
- Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.M., Fratini, A.V. & Kopka, M.L. (1982). *Science*, **216**, 475-485
- Harper, N.A., Brannigan, J.A., Buck, M., Lewis, R.J., Moore, M.H., Schneider, B. (1998) *Acta Cryst.* **D54**, 1273-1284
- Karthe, P. & Gautham, N. (1998). *Acta Cryst.*, **D54**, 501-509
- Kumar, P.S. & Gautham, N. (1999) *Current Science*, **77**, 1076-1078
- Lalitha, V. & Yathindra, N. (1995). *Current Science*, **68**, 68-76
- Sadasivan, C. & Gautham, N. (1995). *J. Mol. Biol.* **248**, 918-930.
- Sadasivan, C., Karthe, P., & Gautham, N. (1994). *Acta Cryst.* **D50**, 192-196
- Wells, R.D. & Harvey, S.C. (eds) (1988). 'Unusual DNA Structures' Springer Verlag. New York, USA.
- Youderian, P., Bouvier, S. & Susskind, M.M. (1982). *Cell*, **30**, 843-853

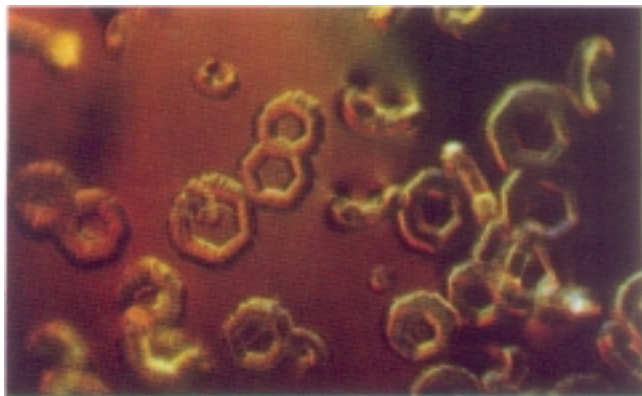


Fig.1: Ring crystals

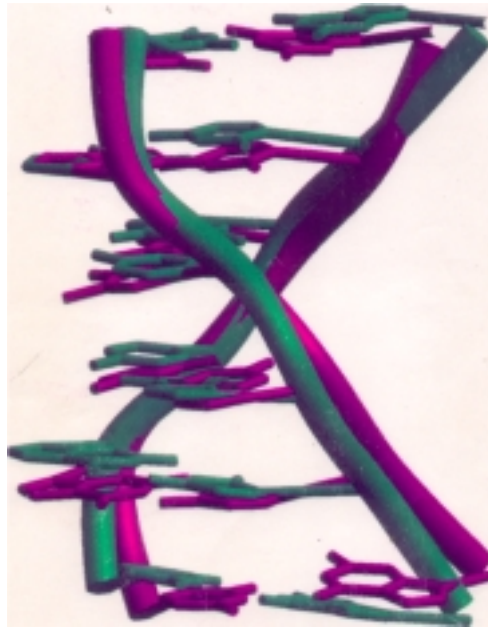


Fig.2: Least squares superposition of the structure of d(CGTGCG).d(CGCACG) on d(CGCGCG)2