Crystal structure of d(CGAAGC); association of two parallel-stranded duplexes through anti-parallel-stranded duplex formation

Satoru Shimizu, Tomonori Kobuna, Tomoko Sunami and Akio Takénaka
Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama 226-8501, Japan

At first, a parallel stranded right-handed double helix formation was found in the half of the sequence d(GCGAAAGCT), the remaining AGCT being formed a normal anti-parallel duplex. Even without the anti-parallel part, it was found that only the CGAA strongly prefers to form a parallel duplex with homo base-pairs, C:C^+, G:G, A:A and A:A. The hemi-protonation is essential for C:C^+ pair formation. To ensure its biological significance, it is necessary to survey sequences that form a longer parallel duplex. To examine such a possibility, crystal structure of d(CGAAGC) has been determined.

DNA hexamer d(CGAAGC) was synthesized on a DNA synthesizer. Crystallization conditions were surveyed at 4˚C by the hanging drop vapor diffusion method. X-ray diffraction data were collected at 100K with synchrotron radiation (λ=1.00Å) at PF (Tsukuba). Initial phase were estimated by the molecular replacement method using the d(CGAA) structure as a probe. The crystal structures were constructed and modified on electron density maps. The atomic parameters were refined with the program CNS.

The d(CGAA) parts of the two hexamers form a parallel duplex, as expected, but the remaining G and C residues of one strand are stacked on the end of the parallel duplex, while those of the other strand are not stacked extending to the outside. The stacked residues form an anti-parallel duplex with the extended residues of the adjacent parallel duplex to form a dimer between the two parallel duplexes related by a crystallographic two-fold symmetry.

Between the two strands of the parallel duplex, the C_1:$^+C_1^p$ pair (P indicates the counter stand.) is formed through the three hydrogen bonds between N4 and $O2^p$, between N3 and $N3^p$, and between O2 and $N4^p$. For this pairing, N3 of either C residue must be hemi-protonated. The G2:$G2^p$ base pair is formed through the two hydrogen bonds between N2 and $N3^p$ and between N3 and $N2^p$ in the minor grooves. The third A:A^p pair occurs between the major groove sites of adenine bases through the two N6-H…N7 hydrogen bonds. At the fourth residue, the A:A^p pairing occurs between the Watson-Crick sites with the two N6-H…N1 hydrogen bonds. In every pairing, the two bases are located at the trans positions to each other. These structural features are similar to those found in the d(GCGAAGCT) and d(CGAA) crystals.

At the fifth and the sixth residues, C6*:$G5^p$ and G5*:$C6^p$ (The asterisk indicates adjacent strand.) pairings organize the cytosine and the guanine flipped out from adjacent duplex. These pairings occur with three hydrogen bonds in Watson-Crick sites. All the ribose rings adopt C2'-endo conformations.

The C1'…C1' atomic distance of the paired nucleotide is the shortest at G2:$G2^p$ and the longest at the A4:A4p. If d(CGAAGC) were formed parallel duplex with homo base pair at all residues, it would be expected that G5:$G5^p$ have the same distance as G2:$G2^p$. So this conformation is unreasonable because the C1'…C1' atomic distance of G5:$G5^p$ pair is excessively shorter than that of A4:$A4^p$. In order to avoid that distortion may arise between A4 and G5, the sequences inserted one or two A, d(CGAAAGC) and d(CGAAAAGC), are designed and crystallized. The both crystals of d(CGAAAGC) and d(CGAAAAGC) are obtained. Now these analyses are going to be performed.

REFERENCES

1. Sunami, T., Kobuna, T., Kondo, J., Hirao, I., Watabe, K., Miura, K. and Takenaka, A. (2002) *Nucleic Acids Res*. **30**.5253-5260.
2. Kobuna, T., Sunami, T., Kondo, J and Takenaka, A. (2002) *Nucleic Acids Res Suppl*. **2**, 179-80.