Crystal structure of intercalated four-stranded d(C3T) at 1.4 Å resolution
(C-C+ base pairs/hydration/telomeres/base-stacking/parallel-stranded duplex)

CHULHEE KANG*, IMRE BERGER*†, CURTIS LOCKSHIN*, ROBERT RATLIFF†, ROBERT MOYZIS‡, AND ALEXANDER RICH*

*Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; †Department of Biophysical Chemistry, Hannover Medical School, 30623 Hannover, Germany; and ‡Center for Human Genome Studies, Los Alamos National Laboratories, Los Alamos, NM 87545


ABSTRACT The crystal structure of d(C3T), solved at 1.4 Å resolution, reveals that the molecule forms a four-stranded intercalated complex. It consists of two parallel-stranded duplexes, each of which is held together by cytosine-protonated cytosine base pairs. The two duplexes are intercalated with each other and have opposite strand orientation. The molecule has a flat, lath-like appearance, and the covalently bonded cytosines have a slow right-handed twist of 17.1°. However, there is considerable asymmetry. On one of the flat sides, the phosphate groups are rotated away from the center of the molecule. They are held in this orientation by bridging water molecules that bind the NH of cytosine and a phosphate group of an opposite chain. There is also considerable microheterogeneity in the structure. The cytosine hemiprotonation occurs even at pH 7 where stable crystals form.

For some time it has been known that nucleic acids containing stretches of cytidine residues can form parallel strands held together by cytosine-protonated cytosine base pairs (C-C+) (1-4). In an NMR analysis of d(TC3), Guérón and his associates (5) proposed an unusual structure in which two such parallel-stranded duplexes, held together by C-C+ base pairs, intercalate with each other in opposite polarity to form a four-stranded molecule. The evidence for the structure was based on strong interactions between the H1' protons on different strands, suggesting that the backbones were close together. Recent crystal structures of d(TAACC) (C.K., I.B., C.L., R.R., R.M., and A.R.), (unpublished data) and of d(C4) (6) were in agreement with the general conclusions deduced by the NMR studies. In telomeres at the end of chromosomes, stretches of cytosine residues are found, usually linked to thymine residues. Here, we present a crystallographic analysis of the structure of d(C3T) at 1.4 Å resolution. This enables us to visualize many aspects of the four-stranded intercalative structure in greater detail. From this structure we see that the molecule is capable of considerable conformational variability. 8

MATERIALS AND METHODS

The DNA segments were synthesized on an Applied Biosystems DNA synthesizer. The sodium salt of the nucleotide was dissolved at a concentration of 2 mM in a solution containing 20 mM MgCl2, 0.12 mM spermine, 80 mM SrCl2, 40 mM strontium cacodylate buffer (pH varied between 6 and 7), and 10% 2-methyl-2,4-pentanediol (MPD). The solution was set up as a hanging drop over 30% MPD. Crystals grew rather slowly. In ~3 weeks rectangular, brick-shaped crystals were observed that finally grew to a size of 0.5 x 0.5 x 1.0 mm.

X-ray diffraction analysis revealed that the molecule had crystallized in space group C2221 with a = 28.281, b = 44.341, and c = 50.472 Å. Three-dimensional x-ray diffraction data were collected on a Rigaku R-axis IIC imaging plate to a resolution of 1.4 Å at 4°C.

The cell parameters along with the symmetry of this orthorhombic space group suggested the presence of four strands of d(CCCCT) in the asymmetric unit. Careful examination of the diffraction data revealed a distinctly high intensity of the (0 14 0) reflection, suggesting a continuous stacking along the b axis with an average spacing of 3.16 Å (44.341/14). Based on this information, we built two strands of d(C3) with bases spaced 6.4 Å apart and arranged the two strands as a parallel duplex involving C-C+ base pairs. An identical duplex was rotated 180° perpendicular to the helical axis, translated by 3.2 Å parallel to the helical axis, and intercalated with opposite polarity into the first duplex according to a model based on NMR experiments (courtesy of M. Guérón). We built several starting models with varying helical rotation and backbone conformation, since the NMR data could not provide strong evidence either for the helical twist or for the geometry of the sugar phosphate backbone.

Rotation searches of all the models indicated a molecule with the helical axis approximately parallel to the crystallographic b axis, as we expected. Subsequent translation searches of some of the models yielded positions for the molecule that were free of close van der Waals contacts. Rigid body refinement followed by intensive simulated annealing using the program xplor (7) improved the density around the phosphate groups of the backbone and allowed us to build another model with more accurate helical twist based on the phosphate positions. We used this improved molecule as a starting point and molecular replacement was performed again. From the very beginning, we could identify two thymine residues stacked on top of six C-C+ layers and added those to the model. At this stage, simulated annealed omit maps (6) clearly showed the location of the remaining two thymine residues stacked perpendicular to the helix axis. Positional refinement followed by refinement of the temperature factors led to a final R factor of 17.7% for 5013 reflections above the 2σ level (based on Fo) between 10 and 1.4 Å. The free R factor (8) value based on a random 10% subset of reflections is 22.5%. The rsm deviation is 0.029 Å for bonds and 4.0° for angles based on the nucleic acid dictionary of xplor (7). The completeness of the entire data between 10 and 1.4 Å is 96% with 74% of the reflections above the 2σ level between 1.6 and 1.4 Å resolution. Fifty-six water molecules were identified, plus three ions believed to be Na+ based on their coordination geometry.

The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 191D and 191D-SF).
FIG. 1. Stereo views of d(C3T). (A) The molecule is composed of two sets of parallel-stranded duplexes held together by C·C* hydrogen bonds. The two duplexes intercalate and have opposite orientations. The molecule with white bonds has the 5' end at the top, while the molecule with green bonds has the 5' end at the bottom. Oxygen atoms are red, nitrogen atoms are blue, and phosphorus is yellow. The molecule is viewed largely from the wide groove. (B) Side view of the molecule showing the manner in which the two sugar phosphate chains at the narrow end of the molecule come close together in their antiparallel orientation. It should be noted that the sugar phosphate chains on the right side of the molecule are both fairly straight, whereas the chains on the left side of the molecule have the phosphates rotated sharply away from the center of the molecule. The thymine residues at the front of the molecule at both top and bottom are involved in lattice interactions. (C) A 2Fo - Fo electron density net is plotted in blue at the 1σ level. One set of chains is pink, while the other set is white. It can be seen that the electron density envelope covers the molecule thoroughly. Holes are found in the electron density map in the center of the sugar rings that can be seen in this view.
RESULTS

The molecules crystallized in a lattice in which the direction of the polynucleotide chains is parallel to the b axis. The asymmetric unit consists of four chains organized as two intercalated, parallel-stranded duplexes pointing in opposite directions. The molecule is shown in stereo photographs in Fig. 1. The two duplexes in Fig. 1 are depicted in different colors. The white duplex has its 5' end at the top of the diagram, while the green duplex has its 5' end at the bottom. It can be seen that the thymine residues at the 3' ends of the molecule adopt two different conformations. One thymine residue remains stacked parallel to the plane of the C-C+ base pairs, while the other is tilted away from the plane of the base pairs where it is involved in lattice interactions. The overall molecule is flat with a slow right-handed twist. Fig. 1A is viewed largely from the wide groove of the molecule, while Fig. 1B views the molecule from the narrow groove. Fig. 1C is a stereoview of the molecule from the wide groove with a 2Fo - Fc electron density net around the molecule. It should be noted that holes in the center of the sugar residues are readily observed at this resolution.

Examination of the molecule from the side (Fig. 1B) reveals that the two chains from different duplexes on the flat sides have a somewhat different conformation. The white chain on the right closest to the reader has a more or less flat and extended conformation. The green chain on the left closest to the reader has a conformation in which the phosphate groups are sharply bent away to the left, thereby increasing the distance between the phosphates of the two chains. The distance between the phosphate groups at the narrow end of the molecule varies from 6 to almost 9 Å with an average phosphate-phosphate (P-P) distance of 7.02 Å and a standard deviation of ±0.6 Å. Even though by its organization the molecule is capable of assuming three perpendicular twofold axes of symmetry, a striking asymmetry is evident. In Fig. 1B, both of the strands on the left-hand side of the molecule have the phosphate groups rotated out, away from the adjacent chain, compared to the fairly straight form of the chains seen on the right-hand side.

This asymmetry is seen most clearly in Fig. 2, which is a view down the molecule. In this figure, 4 C-C+ base pairs are shown. The two strands at the bottom of the molecule are relatively flat, while the strands on the upper part of the molecule have a curved or cup-shaped conformation due to the fact that the phosphate groups are rotated away from the flattened bottom strands. The P-P distances across the two wide grooves differ substantially. Across the bottom wide groove in Fig. 2, the P-P distances are 16.63 ± 0.17 Å. Across the top of Fig. 2, the P-P distances are 12.75 ± 0.22 Å. This asymmetry was also found in the crystal structure of d(TAAGCC), however, it was not reported in the NMR studies (5, 9), nor was it as pronounced in the x-ray structure of d(Ca) (6).

The base stacking interval averages 3.1 Å between adjacent cytosine residues along the axis. This figure is in agreement with the meridional spacing of 3.1 Å observed in x-ray fiber diffraction studies of hemiprotonated polycytidylic acid (1). Unlike the familiar 3.4 Å base stacking found in nucleic acid duplexes, the pyrimidine rings never stack on top of other pyrimidine rings in this structure. Instead, the exocyclic amino and carboxyl groups are stacked upon each other. This permits the bases to be closer than the 3.4 Å spacings found in double-stranded nucleic acids. This spacing is in agreement with the observed 3.1 Å spacing between uracil residues in N-methyl uracil in which only the exocyclic residues are stacked upon each other (10).

The helical twist between coaxially adjacent cytosine residues is 17.1 ± 3.4°. This figure is in close agreement with the 16° estimated in the NMR structure of d(TC) (5, 9) and is somewhat larger than the 12.4° twist observed in d(Ca). This suggests that the molecule has some conformational flexibility not only in the orientation of the phosphate groups, as pointed out above, but also in the helical arrangement of the successive cytidine residues. Of the 12 cytidine sugar rings found in the structure, 6 have the C3' endo pucker and two have a C2' endo pucker. Four have an envelope conformation in which both C2' and C3' are on the endo side of the ring. It is interesting that the four residues with the envelope conformation are found in those chains that have a straight sugar phosphate backbone, as on the right side of Fig. 1B. The difference in the conformation of the sugar phosphate chains is readily observed in the torsion angles. Thus, the torsion angle ω around the phosphate group adopts two different sets of values, one set associated with the sugar phosphate chains that are straight and another set associated with those in which the phosphate groups are rotated away from the center of the molecule.

Fig. 3A shows the electron density map of a C-C+ base pair, together with associated waters of hydration. The hydrogen bond lengths for that pair are listed in the figure. If we take the average of the six C-C+ pairs, we find hydrogen bond lengths of NH-O of 2.77 ± 0.08 Å, while the central NH-N hydrogen bond is 2.74 ± 0.08 Å. The amino group of each cytosine residue is also hydrogen bonded to a water molecule. The average NH-O hydrogen bond length is 3.00 ± 0.13 Å.

The water molecules bonded to NH groups play a special role in stabilizing the structure. As mentioned above, two strands on one side of the molecule have phosphate groups sharply bent away, as shown in Fig. 2. They are stabilized in this conformation by the water bound to NH groups on one side of the molecule. Thus, five cytosine NH groups are linked to the opposite strand by single bridging water molecules bound to phosphate oxygen atoms. The sixth cytosine is linked by two bridging water molecules to the phosphate group of the opposite chain.

The thymine residues play an essential role in building the lattice. A twofold screw axis is found perpendicular to the broad groove at the end of the molecule; this results in forming a thymine-thymine base pair holding molecules together along the helix axis. Fig. 3B shows an electron density map of the thymine-thymine base pairs held together by two hydrogen bonds. The other two thymine residues are not stacked in the molecule but have an orientation virtually parallel to the helical axis and perpendicular to the wide groove. These thymine residues form interactions that build up the lattice. Fig. 4 is a stereoview showing four molecules in the half unit cell as viewed with the b axis vertical and the

Fig. 2. A skeletal end view of the molecule in which four C-C+ base pairs are shown. The two chains at the top form a cup-like array, which is associated with the rotation of the phosphate groups sharply upward from the horizontal plane. The bottom of the molecule is more or less flat.
FIG. 3. Hydrogen bonding with associated $2F_o - F_c$ electron density maps. (A) Bonding between cytosine residues of the central C-C+ base pair showing three hydrogen bonding distances in angstroms. At this resolution, the electron density map shows a hole in the center of the pyrimidine rings. The amino groups of cytosine also form a hydrogen bond to water molecules, shown as circular peaks. On one side of the molecule, that water is bonded to a phosphate oxygen (not shown). (B) Electron density net of the thymine-thymine base pair. The two thymines are from symmetry-related adjacent molecules. They are stacked upon the C-C+ base pairs and hold adjacent molecules together along the helical axis.

c axis horizontal. The four thymine residues oriented parallel to the helix axis are seen in boldface almost in end view stacked upon each other in such a way as to build up the lattice. In the entire unit cell, the upper and the lower molecules along the b axis are stabilized by thymine-thymine pairing interactions, while the molecules to the right and to the left are stabilized by stacked thymidines. The thymine residue is stacked upon the O1' of the sugar of a symmetry-related thymidine building the lattice. This is a common type of stabilizing interaction (11, 12). At the current stage of refinement, there are 59 solvent molecules in the asymmetric unit. Each of the 12 cytosine bases is hydrogen bonded to one

FIG. 4. Stereo diagram of the lattice showing the packing of molecules in the unit cell. The b axis is vertical and the c axis is horizontal as indicated. Four four-stranded molecules are shown, half of the content of the unit cell. They are held together in the center by the stacking of thymine residues upon each other. Four thymine residues (boldface) from four different molecules form a horizontal stacking array. The remaining four four-stranded molecules in the unit cell are held by thymine-thymine base pairs (see Fig. 3B) and stacking interactions.
water molecule, and most of the other water molecules are clustered around the sugar phosphate backbone. Three (or possibly four) sodium ions have been identified in the lattice surrounded by distorted octahedra of water molecules and phosphate oxygens in their coordination sphere. Complete details will be published elsewhere.

**DISCUSSION**

Perhaps the most important component stabilizing this four-stranded intercalative cytosine structure is the hemiprotonation of the cytosines giving rise to three hydrogen bonds. This was first observed in the single-crystal analysis of cytosine-5-acetic acid (13). Hemiprotonation was also used in interpreting the structure of polyribocytidylic acid (1), and solution studies of that polymer (2, 4) as well as polydeoxyctydyllic acid (3) clearly illustrated the importance of hemiprotonation. It is interesting that crystals of d(Cy3T) were grown by using cacodylate buffers that were fixed at pH 6.0 as well as pH 7.0. X-ray diffraction data were collected from crystals grown at both of these pH values, and the data were identical. We must assume that the pK for cytosine in this environment was raised at least to pH 7 in order to stabilize the structure with three hydrogen bonds. A number of studies have been made of the extent to which the cytosine pK can be shifted by the formation of a stable structure of this type. The fact that this four-stranded cytosine structure appears to be stable at neutral pH enhances the possibility of its being found in a biological system.

Even though the structure might have considerable symmetry in its organization with two paired strands pointing in opposite directions and intercalated together, the actual structure appears to have a high level of asymmetry. This is associated with a twisting rotation of the phosphate groups at the narrow end of the molecule as well as variations in local twist geometry. This helps to reduce the electrostatic repulsion between the two closely packed chains that would destabilize the structure. It is curious that this rotation occurs largely in chains on one side of the molecule and not in the other. The stabilization of this phosphate rotation is associated with an intricate network of bridging water molecules that might almost be considered part of the structure.

Unlike the familiar double helices, fragments of the four-stranded, intercalated cytidine molecule have the shape of flat laths rather than cylinders. The laths are twisted slightly, but the basic shape is flat, and this is reflected in the manner in which the lattice is assembled. Many structural studies have been carried out on nucleic acid duplexes, so that we now have a fair understanding of the structural variations that molecule can adopt. The DNA double helix is far from a regular duplex but has numerous changes, both in helical twists and stacking interactions, that produce microheterogeneity. There is also considerable microheterogeneity that is found in the four-stranded cytidylic intercalated structures. It is clear that the molecule presents quite different surfaces at both its wide grooves as well as the narrow grooves. Furthermore, the narrow grooves themselves appear to be variable in ways that we do not fully understand. It will take further structural studies to uncover the range of variations in this four-stranded intercalated DNA.

We thank Dr. M. Guéron for providing us with details of the intercalated d(TC)2 structure and Dr. A. Herbert for helpful discussions. This research was supported by grants at Massachusetts Institute of Technology from the National Institutes of Health, the National Science Foundation, the Office of Naval Research, the American Cancer Society, and the National Aeronautics and Space Administration, and a US Department of Energy grant (R.M.).