RNA DOUBLE HELICES GENERATED FROM CRYSTAL STRUCTURES OF DOUBLE HELICAL DINUCLEOSIDE PHOSPHATES

John M. Rosenberg[+], Nadrian C. Seeman, Roberta O. Day[*] and Alexander Rich

Department of Biology
Massachusetts Institute of Technology
Cambridge, MA 02139, U. S. A.

Received February 25,1976

SUMMARY: The dinucleoside phosphates ApU and GpC form right-handed anti-parallel double helical fragments within their crystal lattices. Using a least squares procedure, we have generated the extended double helices which these fragments represent. ApU corresponds to a double helix with 11. 9 residues per turn and a pitch of 28. 1Å. The GpC double helix has 10. 4 residues per turn and a pitch of 26. 9A.

INTRODUCTION: It is well established that nucleic acids are capable of forming right-handed anti-parallel double helical structures held together by complementary hydrogen bonding between the bases. Specific pairing between adenine and thymine (or uracil) and between guanine and cytosine was postulated for nucleic acid double helices by Watson and Crick (1). We have (2-5) visualized this hydrogen bonding by determining the crystal structures of adenylyl-3',5'-uridine (ApU) and guanylyl-3',5'-cytidine (GpC) at atomic resolution. Besides its role in genetic material, the nucleic acid double helix has been shown to play a crucial part in stabilizing the structures of other cellular components. The determination of the crystal structure of yeast $tRNA^{phe}$ (6-8) indicates that large parts of the tRNA molecule are stabilized in the double helical conformation. There are also indications

[+]Present address: Department of Chemistry, California Institute of Technology, Pasadena, California 91109

[*]Present address: Department of Chemistry, University of Nebraska, Lincoln, Nebraska 65808

that messenger RNA, segments of ribosomal RNA and other single stranded cellular RNA's contain self-complementary sequences which are likely to assume double helical structures (see ref. 9 for a compilation). The structural parameters of RNA double helices are therefore of interest. Since both ApU and GpC form double helical segments in the crystal, we have been able to derive these quantities from the intrinsic helical symmetry of the molecules and here we compare these results with those parameters obtained from fiber studies.

METHODS: The helix parameters of a helical polymer are directly obtainable from its fiber diffraction pattern (10). However, a different procedure must be devised to extract these parameters from the crystal structure of a fragment, such as a dinucleoside phosphate, or the double helical stems of transfer RNA. Given a unit with intrinsic helical symmetry, it is possible to generate a much larger structure containing the same unit by repeating the symmetry operation sequentially, using the familiar concatenation methods of molecular model building. For example, this procedure was used to generate the 4_1 RNA helix (11) from the UpA2 molecule (12, 13). Unfortunately, this method tends to suffer seriously from an accumulation of errors. In order to circumvent such difficulties, a least squares procedure has been devised to extract this information from the coordinates of the dinucleoside phosphates*. A translation and a rotation are needed to generate a helix. Given two molecular units which are arranged in helical symmetry, a minimum of three vectors between subsequent equivalent atoms are necessary to define the helix. These may arise from at least three equivalent atom pairs in each of the residues. Imagine that vectors pass from atoms of one nucleotide to the same atom of the next nucleotide. If the tails of these vectors between the equivalent atoms are translated to the same origin of a coordinate system, their heads define a plane which is normal to the helix axis. The distance of the plane from the origin defines the helical repeat distance, i. e. the translational distance along the helix axis. When more than three vectors are available, as with ApU and GpC, the best least squares plane is generated.

Once the translational component of the helical operator is determined, the rotational component may be obtained by projecting the equivalent atoms into the least squares plane. The center of the helix is located by the least squares intersection of the perpendicular bisectors of the lines joining sequential equivalent projected atoms; the average angular separation of sequential equivalent projected atoms with respect to this point is the desired rotational parameter. The double helical symmetry described below was imposed by simultaneously using atoms from both strands. The fragment of the molecule which is treated as being geometrically equivalent between sequential residues along the chain is called the "helix probe."

*A FORTRAN computer program for determining the best least squares helix from a fragment is available upon request from the authors.

RESULTS: In the course of our calculations with the ApU and GpC structures, we discovered that the direction of the helix and the helical parameters may be very sensitive to the choice of helix probe. Small departures from helical symmetry in the atoms used to define the probe can result in serious stereochemical clashes when generating complete double helices from such procedures. This is a fundamental problem when dealing with a fragment as small as a dimer, but should diminish when dealing with longer oligomers.

Two different helix probes were used in this work. The smaller probe consisted of the ribose C1' and the atoms covalently bonded to it (glycosidic N, O1' and C2'). The larger, which gave almost identical results in the case of ApU, consisted of these atoms plus C3' and C4'. Ribose ring substituents other than the glycosidic nitrogen atoms were never included because of the small differences in conformation observed from ring to ring. The large probe did not result in meaningful stereochemistry for GpC, although the small one gave good results.

In generating a helix from a given translation and rotation, there is likely to be a discrepancy between the new calculated positions for atoms and the values observed in the crystal structure. We calculate this discrepancy in terms of a "helix sigma," which is defined as the average deviation of the coordinates generated by the helical operator when compared with their observed values. The value is 0.17Å for ApU and 0.22Å for GpC. Since we are considering atoms separated by as much as 11 or 12Å across the double helix, this corresponds to 1.5%-2% deviations from true double helicity. Unit weights were applied to all atomic positional parameters throughout these calculations.

The qualitative picture derived from these helices generated by single crystal analysis is very similar to that obtained from fiber studies

Table 1. Helical Parameters for Nucleic Acid Double Helices.

	ApU	GpC	RNA-A	RNA-A'	DNA-A	DNA-B
Translation (Å)	2.36	2.59	2.81	3.0	2.59	3.38
Rotation (°)	30.2	34.7	32.7	30.0	32.7	36.0
Pitch (Å)	28.1	26.9	30.9	36.0	28.5	33.8
Residue/turn	11.9	10.4	11.0	12.0	11.0	10.0
Residue density (residue/Å rise)	0.42	0.39	0.36	0.33	0.39	0.29
Base tilt (°)	23.	28.	13.	14.	19.	0.
Reference	This work	This work	14	14	15	15

despite some small differences in quantitative details (14, 15). The helical parameters are listed in Table 1, along with those of some nucleic acid double helices derived from fiber diffraction studies. Both the translational and rotational parameters for the double helices are positive, thus the helices are right-handed. We have generated full double helices from these coordinates, and side views are shown stereoscopically in figures 1 and 2.

The ApU and GpC double helices are qualitatively similar to each other; however, some differences may be noted in the figures. Because there are fewer residues per unit displacement in the GpC helix, it is somewhat narrower than the ApU helix. The base tilt is also somewhat more pronounced resulting in a major groove opening between the phosphates which is considerably reduced. The differences between the two double helices which we have derived may be due to the different lattice constraints in the two different crystals. The slightly different sizes of the A-U and the G-C base pairs may also contribute to the difference (4).

The minor grooves of the helices are extremely shallow, while the

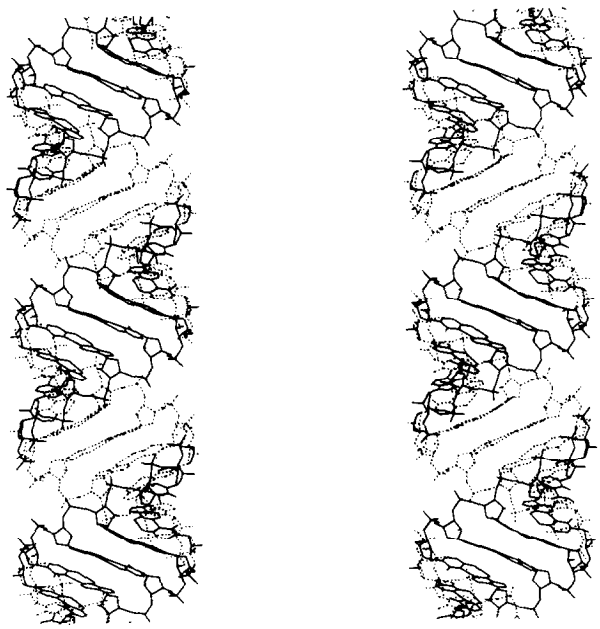

Figure 1. Stereoscopic representation of the RNA double helix derived from the ApU structure. In this stereoscopic drawing those components of the double helix which are on the side of the helix axis away from the reader have been dotted while solid lines are used to represent the side of the double helix towards the reader.

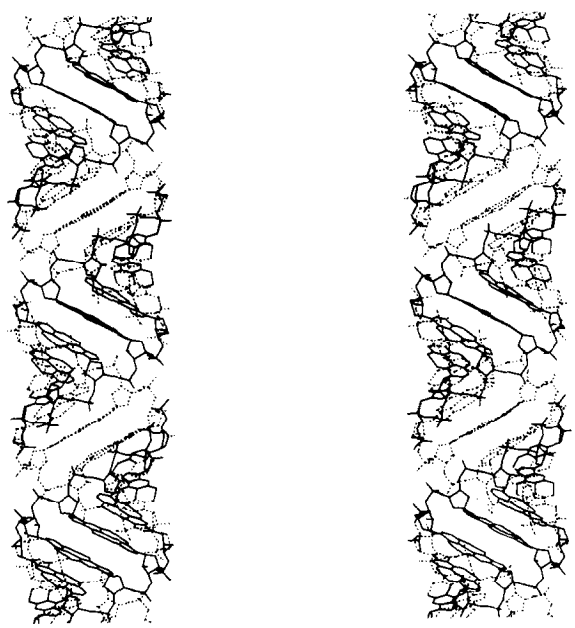

Figure 2. Stereoscopic representation of the RNA double helix generated by the GpC structure. The same conventions are used as in Figure 1. It should be noted that this helix is slightly thinner than that obtained from the ApU structure. This is associated with a somewhat greater tilting of the bases.

Table 2. These are the cylindrical polar coordinates in angstroms of a residue of the double helices. The Z axis corresponds to the helix axis, and the X axis is coincident with the dyad axis which passes through the base pair. The residue on the opposite strand may be generated by reversing the sign of both ϕ and Z. Successive residues along each strand may be generated by adding 30.2° to ϕ and 2.36Å to Z for ApU, and 34.7° to ϕ and 2.59Å to Z for GpC.

	ApU				GpC		
	r(Å)	ϕ(°)	Z(Å)		r(Å)	ϕ(°)	Z(Å)
C1'	9.97	29.30	-1.92	C1'	8.81	32.26	-2.52
O1'	10.43	36.71	-2.15	O1'	9.34	40.64	-2.70
C2'	10.33	24.96	-3.20	C2'	8.99	27.89	-3.94
O2'	11.71	23.23	-3.11	O2'	10.36	25.63	-4.00
C3'	10.18	31.15	-4.24	C3'	8.85	35.55	-4.87
O3'	10.77	29.85	-5.47	O3'	9.49	34.37	-6.10
C4'	10.89	37.25	-3.50	C4'	9.76	41.79	-4.06
C5'	11.10	44.23	-3.98	C5'	9.73	50.51	-4.44
O5'	9.82	47.99	-4.05	O5'	8.44	54.37	-4.42
P	9.96	57.00	-4.39	P	8.78	64.95	-4.76
O1P	10.95	57.97	-5.44	O1P	9.78	65.41	-5.84
O2P	8.62	61.04	-4.49	O2P	7.53	69.60	-4.85
C6U	7.64	36.10	-2.10	C6C	6.59	41.68	-2.39
C5U	6.31	36.70	-1.94	C5C	5.37	44.15	-2.04
C4U	5.75	26.15	-1.25	C4C	4.80	31.78	-1.31
O4U	4.55	22.28	-1.07	N4C	3.48	30.59	-0.88
N3U	6.77	18.88	-0.75	N3C	5.69	21.88	-0.89
C2U	8.10	21.29	-0.93	C2C	6.99	23.93	-1.32
N1U	8.46	29.20	-1.61	N1C	7.35	32.74	-2.14
O2U	8.98	16.54	-0.49	O2C	7.93	17.88	-1.08
N9A	8.58	28.73	-1.61	N9G	7.42	32.13	-2.14
C8A	7.74	36.26	-1.64	C8G	6.59	42.09	-2.26
N7A	6.48	34.27	-1.27	N7G	5.36	40.55	-1.75
C5A	6.66	22.77	-0.97	C5G	5.53	26.49	-1.25
C4A	7.99	20.81	-1.15	C4G	6.85	23.53	-1.53
N3A	8.76	13.28	-0.96	N3G	7.61	15.23	-1.31
C2A	8.12	5.82	-0.51	C2G	7.07	6.43	-0.70
N1A	6.83	3.48	-0.29	N2G	8.03	-0.36	-0.42
C6A	5.95	12.32	-0.50	N1G	5.74	4.42	-0.38
N6A	4.61	10.77	-0.28	C6G	4.81	14.76	-0.65
Na^+	10.64	-15.10	-1.18	O6G	3.60	13.17	-0.30

major grooves are very deep. Thus, potential sites for recognition in the shallow minor groove are perhaps more accessible to globular proteins than is the deeper major groove. Although the major groove of a long RNA double helix appears less accessible to globular macromolecules, a

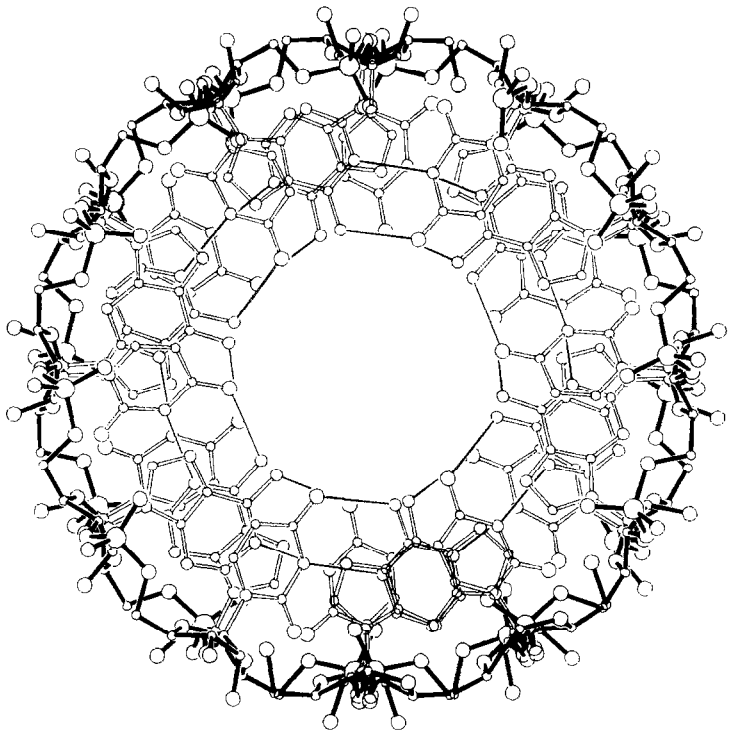

Figure 3. A view of the double helix generated from the ApU structure viewed down the helix axis. It should be noted that there is a hole in the center of the helix of approximately 9Å diameter. In this drawing the ribose phosphate backbone is drawn with solid bonds while the bonds in the bases are unshaded. Hydrogen bonds are indicated by single thin solid lines.

partial turn, such as found in the tRNA stems (6-8) exposes these sites to a greater extent. We have included in table 2 the helical coordinates of the sodium ion bound in the minor groove of ApU (2, 4). It is on the extreme radial periphery of the minor groove and is readily accessible from the surrounding solution. The presence of this sodium ion changes the local steric and electrostatic characteristics of the minor groove in the ApU region. O2' is the atom furthest from the helical axis, thereby making it readily accessible to proteins which are specific for RNA binding and it can also interfere with DNA specific binding proteins.

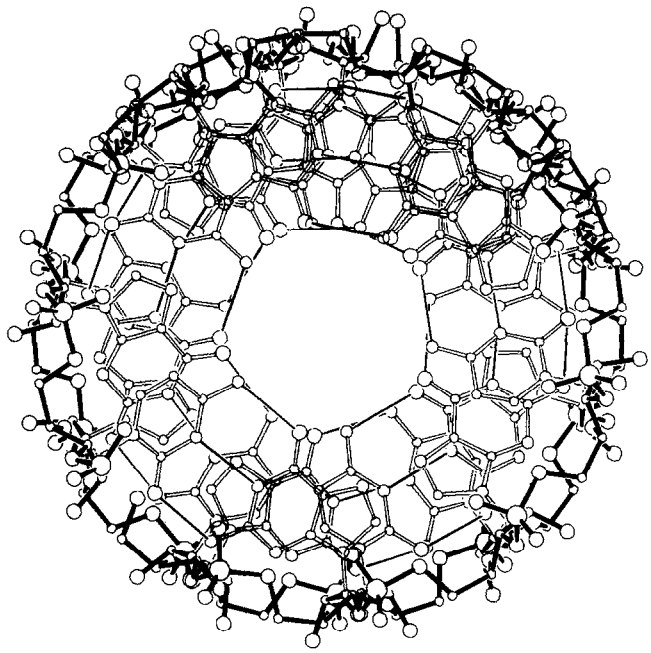

Figure 4. A view of the double helix generated by the GpC structure as viewed down the helix axis. The diameter of the hole in the center of this helix is 7Å, slightly smaller than that found in the ApU helix. The same conventions are used as in Figure 3.

It is instructive to emphasize that the RNA double helix and the A-form DNA double helix are qualitatively quite different from the familiar B-form DNA double helix (See table 1). In the B-DNA double helix, the bases are parallel to each other, are perpendicular to the helix axis and are intersected by that axis. Thus, the B-form DNA double helix is analogous to a thick ribbon which has been twisted about its long axis. In the RNA double helix, as can be seen in the figures 1 and 2, the bases are tilted with respect to each other and to the helix axis; furthermore, they are not intersected by the helix axis. Thus, when viewed down the helix axis, there are large holes at the center, as shown in figures 3 and 4. The RNA double helix is therefore analogous to a thick ribbon wrapped around a pole where the pole

represents the empty space in the center of figures 3 and 4. The major groove side of the bases corresponds to the surface which is in contact with the pole, while the minor groove is on the outside.

ACKNOWLEDGEMENTS: This research was supported by grants from the National Institutes of Health, the National Science Foundation, the National Aeronautics and Space Administration and the American Cancer Society. J. M. R. was a trainee of the N. I. H.; N. C. S. and R. O. D. are postdoctoral fellows of the N. I. H.

REFERENCES

1. Watson, J. D., and Crick, F. H. C. (1953) Nature, 171, 737-738.
2. Rosenberg, J. M., Seeman, N. C., Kim, J. J. P., Suddath, F. L., Nicholas, H. B., and Rich, A. (1973) Nature, 243, 150-154.
3. Day, R. O., Seeman, N. C., Rosenberg, J. M., and Rich, A. (1973) Proc. Nat. Acad. Sci. U. S. A., 70, 849-853.
4. Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P., and Rich, A. (1976) J. Mol. Biol., in press.
5. Rosenberg, J. M., Seeman, N. C., Day, R. O., and Rich, A. (1976) J. Mol. Biol., in press.
6. Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J., and Rich, A. (1973) Science, 179, 285-288.
7. Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C., and Rich, A. (1974) Science, 185, 435-440.
8. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C., and Klug, A. (1974) Nature, 250, 546-551.
9. Barrell, B. G., and Clark, B. F. C. (1974) Handbook of Nucleic Acid Sequences, Joynson-Bruvvers Ltd., Oxford, England.
10. Cochran, W., Crick, F. H. C., and Vand, V. (1952) Acta. Cryst., 5, 581-586.
11. Kim, S. H., Berman, H. M., Seeman, N. C., and Newton, M. D. (1973) Acta. Cryst., B29, 703-709.
12. Seeman, N. C., Sussman, J. L., Berman, H. M., and Kim, S. H. (1971) Nature New Biology, 233, 90-92.
13. Sussman, J. L., Seeman, N. C., Kim, S. H., and Berman, H. M. (1972) J. Mol. Biol., 66, 403-422.
14. Arnott, S., Hukins, D. W. L., and Dover, S. D. (1972) Biochem. Biophys. Res. Comm., 48, 1392-1399.
15. Arnott, S., and Hukins, D. W. L. (1972) Biochem. Biophys. Res. Comm., 47, 1504-1509.