A Structural Analysis of Three Arabinosylpyrimidines

Thesis by

James Stanley Sherfinski

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1974

(Submitted August 10, 1973)

To Lalo

Acknowledgements

An acknowledgement seems a meager way to thank one who has done for me what my advisor Dr. Richard E. Marsh has. I thank him specially, but there is more than thanks hidden in those words. My educational experience at Caltech has also been highlighted by discussions with many people, who for my fear of omitting someone, I will collectively but affectionately refer to as "the coffee crew." To the National Institutes of Health and Caltech, thanks for the Financial support and a chance to help teach the Caltech undergrads. I also acknowledge Dr. Harry Wood of the Cancer Chemotherapy National Service Center who supplied several of the compounds I have studied. Finally, a special note of thanks goes to my wife Sally, who happened along at just the right time.

A Structural Analysis of Three Arabinosylpyrimidines

Abstract

The crystal structures of three arabinosylpyrimidines are presented. Crystals of 1-β-D-arabinofuranosylcytosine hydrochloride are monoclinic, space group $P2_1$, with cell dimensions $\underline{a}$ = 6.540(2) Å, $\underline{b}$ = 15.260(2) Å, $\underline{c}$ = 6.758(3) Å, and β = 117.68(3)°; Z = 2. The original structure solution was by Patterson-Fourier techniques; a second method of solution is presented which takes advantage of the chloride ion being an anomalous scatterer in a centrosymmetric array. Full-matrix least-squares refinement yielded a final R index of 0.022 for 2490 data; the estimated positional standard deviation of a carbon atom is 0.002 Å. The cation is in the gauche-trans and C(2')-endo conformations; the glycosidic torsion angle ϕ_{CN} is 26.7°. Using the anomalous scattering of Cl-, O, and N in CuKα radiation, the absolute configuration of the sugar ring was determined.

The compound 1-β-D-arabinofuranosyluracil crystallizes in the orthorhombic space group $P2_12_12_1$ with cell dimensions $\underline{a}$ = 6.888(2) Å, $\underline{b}$ = 20.973(5) Å, $\underline{c}$ = 6.826(3) Å, and Z = 4. Solution was again by Patterson methods, and the structure was refined by full-matrix least squares to a final R index of 0.030 for 2034 counter-collected data (λ(CuKα) = 1.5418Å). The estimated positional standard deviation of a carbon atom is 0.002 Å. The absolute configuration of this molecule was also established, this time using only O and N as anomalous

scatterers. The sugar is in the C(2')-*endo*, *gauche*-*gauche* conformation which is stabilized by an intramolecular O(2')...O(5') hydrogen bond. The glycosidic torsion angle is 35.0°.

The nucleotide 1-β-D-arabinofuranosylcytosine-5'-monophosphate crystallizes from water in space group $P2_12_12_1$ as the trihydrate; there are four formula units per unit cell of dimensions *a* = 17.531(3) Å, *b* = 18.252(4) Å, *c* = 4.816(1)Å. Initial solution attempts centered on interpreting a projection Patterson map calculated with Beevers-Lipson strips and film data. Ultimately the structure solution was accomplished utilizing information from both a three-dimensional Patterson map and a direct methods solution in which no enantiomorph discrimination had been done. Least-squares refinement was halted at R = 0.046 for 1859 diffractometer data; the estimated standard deviation in the position of a carbon atom is 0.004 Å. The absolute configuration of the arabinose was determined utilyzing the anomalous scattering properties of P, O, and N in CuKα radiation. The sugar is in the C(3')-*endo*, *gauche*-*gauche* conformation; the glycosidic torsion angle is 30.7°.

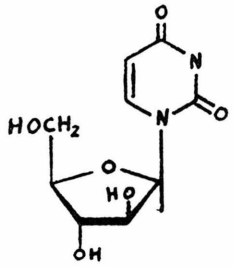
For the three nucleotides and nucleosides studied, the glycosidic torsion angle is restricted to a very narrow range of values, whereas three quite different sugar conformations are exhibited. The results indicate that

arabinofuranosylpyrimidines can with little effort conform to the same geometrical descriptions calculated for ribo- and deoxyribofuranosylpyrimidines in double-helical nucleic acid structures.

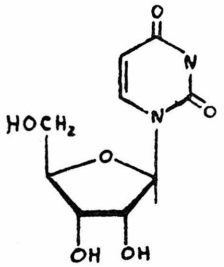
Table of Contents

INTRODUCTION

Over twenty years ago investigators isolated from the sponge Cryptotethya crypta the two unusual nucleosides arabinofuranosylthymine and arabinofuranosyluracil (1,2,3). The sugar was identified by standard carbohydrate techniques to be arabinofuranose, and such nucleosides have since been referred to as "spongosides."



arabinofuranosyl-
uracil



ribofuranosyl-
uracil

Figure 1.

The intriguing difference between the arabinose and ribose is the epimerization at carbon atom C(2'); in ribose, O(2') and O(3') are cis to each other, while in arabinose they are trans.

As more and more became known about nucleic acid double helices and their construction, the exact role of these "spongosides" in C. crypta was questioned. An analysis of nucleic acids isolated from the sponge yielded only ribose and 2'-deoxyribose components and no arabinonucleotides or nucleosides(4). Thus it appeared as though C. crypta's genetic material was similar to all others thus far studied.

Analysis of cell types in the sponge showed that very nearly all of the arabinosylnucleotide content was located in a particular type of cell comprising only 2% of the wet weight of the sponge. Furthermore, these cells had either very short DNA and RNA* strands, or no DNA at all, and had a low capacity for energy production and biosynthesis(5).

In the next few years it was discovered that the related arabinosylcytosine was an effective growth inhibitor of leukemic cells and other types of cancers(6-11). Chu and Fisher, in their early investigations, showed that ARAC inhibits the reduction of CDP to dCDP (6). However, Moore and Cohen found that ARACDP was no more effective in inhibiting reduction of CDP than was dCDP (7). Other researchers also have found that ARAC was more potent than ARACMP, but have explained the anomaly by the hypothesis that ARACMP and ARACDP must be dephosphorylated before passage into a cell or nucleus is possible (8). Clearly the inhibition of CDP reductase is not the complete answer, as thymidine also inhibits the enzyme, but shows no anti-cancer activity (12).

---

*Some of the abbreviations used here are DNA: deoxyribonucleic acid; RNA: ribonucleic acid; ARAC: arabinofuranosylcytosine; ARAU: arabinofuranosyluracil; ARAC.HCl: arabinofuranosylcytosine hydrochloride; ARACMP: arabinofuranosylcytosine-5'-monophosphate; CDP: ribofuranosylcytosine-5'-diphosphate; dCDP: deoxyribofuranosylcytosine-5'-diphosphate; ARACDP: arabinofuranosylcytosine-5'-diphosphate.

A second plausible explanation of the activity of ARAC is the hypothesis that the drug is incorporated directly into synthesizing DNA. This type of activity is supported by observations that ARAC can be phosphorylated by deoxycytidine kinase (6,13-16). Furthermore, it has been demonstrated that ARAC potency is highest if the drug is administered while the cells under study are in the S phase of reproduction, or that time when DNA replication is taking place (17,18). Momparler has shown that ARAC is indeed incorporated into replicating DNA, and that this nucleoside is located primarily at the 3'-terminus. Since polymer synthesis is 5' to 3', polymerization must cease upon incorporation of the arabinoside. Karon and coworkers have noticed a correlation between ARAC-induced cell death and an increase in chromatid breakage (20,21). This breakage may also indicate that ARAC terminates DNA synthesis.

An interesting correlary to the activity of ARAC is that the related nucleoside arabinofuranosyluracil has no anti-tumor properties whatsoever (10,13). ARAC seems to be DNA-specific in its activity, while uridine is incorporated only into RNA and not DNA. A reason for ARAU's inactivity may then be that DNA polymerase does not recognize the compound as a legitimate substrate because of the uracil base, while RNA polymerase will not recognize ARAU because of the different configuration of the sugar at C(2'). ARAC will deaminate <u>in vivo</u> to be excreted as ARAU in, for

instance, mammalian urine (10,13). Understandably perhaps, most of the pertinent literature relates to the potent ARAC rather than ARAU.

The novel construction of the 1-β-D-arabinofuranosyl nucleosides, coupled with the need to understand more fully the detailed geometry of these sometimes potent, sometimes not, compounds, prompted the instigation of this project. At the time this work was begun, only one structural report of an arabinosyl nucleoside had been reported(22). Since that time seven more, including the three structures presented here, have been described (23-27). Discussion of the other four will be left to the final chapter.

EXPERIMENTAL

ARAC.HCl

Crystals of ARAC.HCl were supplied by the Cancer Chemotherapy National Service Center. A small sample was dissolved in a minimum amount of distilled water at room temperature, and this solution was allowed to evaporate almost to dryness. Several needle-shape crystals elongated along c were isolated, although the majority of the sample precipitated as a clear, colorless, multicrystalline cake. One approximately triangular prism, 0.37 x 0.12 x 0.15 mm in size, was mounted on a glass fiber. Photographic investigations indicated the cell to be monoclinic, space group $P2_1$ or $P2_1/m$ (systematic absences: 0k0, k = 2n+1). The former was chosen because of the presumed optical purity of the compound and the measured density (by flotation in a benzene/carbon tetrachloride solution), which indicated only two molecules in the unit cell. This same crystal was then used for data collection.

ARAU

Crystals of ARAU were also obtained from the Cancer Chemotherapy National Service Center. Slow evaporation of a dilute aqueous solution resulted in the formation of a colorless multicrystalline cake. A small fragment was chipped away and mounted on a glass fiber for photographic investigation. Initial evidence indicated a monoclinic cell with approximate dimensions 6.7 x 20.8 x 9.6 Å, $\beta = 134^{\circ}$. Reduction of this cell proved to be fruitful, as the crystal system is orthorhombic (Table 2) with the spindle axis in this case parallel to the $[10\bar{1}]$ direction. Another crystal was isolated and photographed; this crystal was mounted with the spindle axis parallel to [010], and the photographs confirmed the choice of crystal systems and indicated the space group unambiguously to be $P2_12_12_1$. As the density measurement (by flotation in a carbon tetrachoride/ dibromomethane solution) had indicated four molecules in the unit cell, there was now only one molecule in the asymmetric unit, as opposed to two for the original monoclinic cell. Data were collected using the original, irregular chunk (0.10 x 0.18 x 0.16 mm) under conditions described below.

ARACMP

One hundred milligrams of ARACMP hydrate were purchased from Terramarine Bioresearch, LaJolla, California. From an aqueous solution lath-shaped crystals were easily obtained, radiating in a pincushion-type arrangement from a number of nucleation sites. Weissenberg and oscillation photographs indicated an orthorhombic cell, space group either $P2_12_12_1$ or $P2_12_12$ (systematic absences along the 00ℓ row were not accessible to this film investigation). A segment (0.14 x 0.13 x 0.05 mm) was cut from one isolated lath and used for all subsequent diffractometer data collection. A similarily shaped crystal was used for all film work, which included a set of hk0 Weissenberg photographs used for intensity measurements (see STRUCTURAL SOLUTIONS: ARACMP.$3H_2O$). Density analysis (by flotation in a carbon tetrachloride/dibromomethane solution) indicated three water molecules of crystallization; their presence was confirmed by the successful refinement of the structure.

DATA COLLECTION

Data used in the final refinements of all three crystal structures were measured under basically similar conditions. The instrument used was a General Electric XRD-5 quarter-circle diffractometer equipped with a scintillation counter and automated by Datex. A copper target X-ray tube was used with the radiation monochromatized by nickel filters and secondarily by a pulse height analyzer. Cell dimensions (Tables 1,2,and 3) and orientation matrices for each crystal were obtained separately from a least-squares fit to the observed 2θ and χ and ϕ observations for a number of reflections (see Table 4). Intensity data were collected using the θ-2θ scan technique, scanning at 1°/min. Background corrections were made. All net intensities were corrected for Lorentz-polarization effects and placed on an approximately absolute scale by means of a Wilson plot(28). Observational variances $\sigma^2(F_o^2)$ included counting statistics for the scan and background plus a term $(0.02S)^2$, where S is the scan count, to allow for instrumental variations(29). Additional information pertinent to each data collection is given in Table 4.

Table 1. ARAC.HCl Crystal Data.

$C_9H_{14}N_3O_5^+Cl^-$	F.W. = 279.7
Monoclinic	Space Group $P2_1$
$\underline{a}$ = 6.540(2) Å	Z = 2
$\underline{b}$ = 15.260(2) Å	F(000) = 292
$\underline{c}$ = 6.758(3) Å	V = 597 Å^3
β = 117.68(3)°	D_x = 1.551 g.cm^{-3}
λ(CuKα) = 1.5418 Å	D_m = 1.56(1)
μ = 30.5 cm^{-1}	

Table 2. Crystal Data for ARAU.

$C_9H_{12}N_2O_6$	F.W. = 244.2
Orthorhombic	Space Group $P2_12_12_1$
$\underline{a}$ = 6.888(2) Å	Z = 4
$\underline{b}$ = 20.973(5) Å	F(000) = 512
$\underline{c}$ = 6.826(3) Å	V = 986 Å^3
λ(CuKα) = 1.5418 Å	D_x = 1.645 g.cm^{-3}
μ = 12.2 cm^{-1}	D_m = 1.62(1)

Table 3. Crystal Data for ARACMP.$3H_2O$.

$C_9H_{14}N_3O_8P.3H_2O$	F.W. = 377.2
Orthorhombic	Space Group $P2_12_12_1$
$\underline{a}$ = 17.531(3) Å	Z = 4
$\underline{b}$ = 18.252(4) Å	F(000) = 824
$\underline{c}$ = 4.816(1) Å	V = 1541 $Å^3$
λ(CuKα) = 1.5418 Å	D_x = 1.614 $g.cm^{-3}$
μ = 22.1 cm^{-1}	D_m = 1.62(1)

Table 4. Data Collection Summary

Other details are mentioned in the text.

	ARAC.HCl	ARAU	ARACMP.$3H_2O$
Scan range	$R=1.91^{o}+0.00822(2\theta)$	$R=1.91^{o}+0.00822(2\theta)$	$R=1.91^{o}+0.00822(2\theta)$
Backgrounds	20 secs. at each extremus	40 secs. at each extremus	40 secs. at each extremus
Maximum $\sin^2\theta/\lambda^2$	0.4010 A^{o-2}	0.3925 A^{o-2}	0.3925 A^{o-2}
Total # of data	2544 (2 quadrants)	2034 (2 octants)	1859 (1 octant)
Total # of non-zero weight data	2490	2034	1859
Total # of data, $F_o^2>0$	2490 ($F_o > 3\sigma(F_o)$)	2001	1796
Crystal decay	less than 1% in F (judged by 3 check refs.)	about 1% in F (judged by 4 check refs.)	about 5% in F (judged by 4 check refs.)
Decay correction	none	none	linear
# of refs. to determine cell dims.	10	13	13

STRUCTURAL SOLUTIONS

ARAC.HCl

The initial structure solution was by Patterson-Fourier techniques. The positions of the chloride ion and two other atoms 4-5 Å distant were assigned from a three-dimensional Patterson map calculated using one quadrant of data, and four structure-factor Fourier-map cycles revealed the remaining non-hydrogen atoms. The R index was 0.21(30).

In retrospect, we thought that this particular solution may have been fortuitous, in that one atom, in this case Cl-, positioned in space group $P2_1$ presents a centrosymmetric array of scatterers, and any phase information derived from this one-atom structure will be centric. By the inclusion of the two other atoms in the first structure factor calculation we were able to break this symmetry problem, but not easily.

An alternative method of solution, utilizing the fact that chlorine exhibits a detectable anomalous scattering component ($\Delta f''$) for CuKα radiation, was chosen. Knowing the position of the chloride ion (by Patterson interpretation), and having a complete set of intensities and the Friedel-related intensities (hkl and $\overline{hkl}$), starting phases can be calculated which could be subjected to tangent formula expansion and refinement(31). The results would hopefully be an interpretable E-map. The phase calculation method (here

called BRR) is that used by Moncrief and Lipscomb(32), based on the pioneering work of Bijvoet, Raman, and Ramachandran (33, 34). Use of the tangent formula in conjuction with BRR phasing had been suggested by Hazell(35).

Writing the structure factor in its familiar form we have

$$\underline{F} = \underline{A} + i\underline{B}. \qquad (1)$$

Since an anomalous scatterer in a centrosymmetric array contributes to the structure factor in the form $i\Delta f''$, we can factor the imaginary component of the structure factor to get

$$\underline{F} = \underline{A} + i\underline{B}' + i\underline{F}_H'' \qquad (2)$$

where $\underline{F}_H''$ is the imaginary contribution to the structure factor by the heavy atom ($\underline{F}_H''$ can be calculated since the position of the heavy atom is known). For the Friedel-related reflection one does the same thing:

$$\overline{\underline{F}} = \overline{\underline{A}} + i\overline{\underline{B}} \qquad (3)$$

$$= \overline{\underline{A}} + i\overline{\underline{B}}' + i\overline{\underline{F}}_H'' \qquad (4)$$

$$= \underline{A} - i\underline{B}' + i\underline{F}_H''. \qquad (5)$$

For a pictorial representation of these vectors, see Fig. 2.

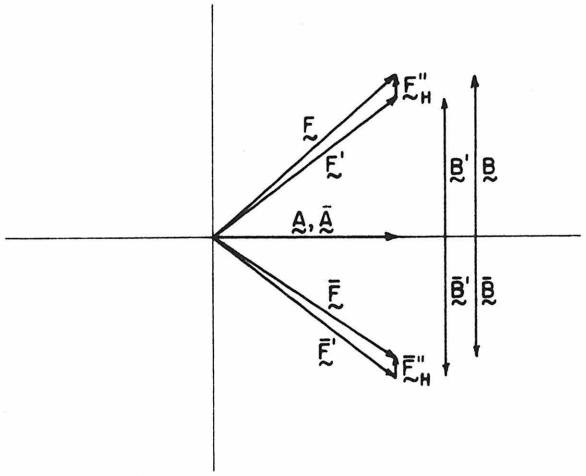
Now we can expand the expressions for $|\underline{F}|^2$ and $|\overline{\underline{F}}|^2$ and obtain a relation between $\underline{B}'$, $\underline{F}_H''$, $\underline{F}$, and $\overline{\underline{F}}$. Since $|\underline{F}|^2 = \underline{F}\cdot\underline{F}^*$,

$$|\underline{F}|^2 = (\underline{A} + i\underline{B}' + i\underline{F}_H'')\cdot(\underline{A} - i\underline{B}' - i\underline{F}_H'')$$

$$= |\underline{A}|^2 + |\underline{B}'|^2 + |\underline{F}_H''|^2 + 2(B')(F_H''). \qquad (6)$$

where B' and F_H'' are signed magnitudes of the respective

Fig. 2. A diagram showing the vector relationships as used in the BRR phasing derivation.



vector quantities. Likewise,

$$|\bar{\underline{F}}|^2 = (\underline{A} - i\underline{B}' + i\underline{F}_H'')\cdot(\underline{A} + i\underline{B}' - i\underline{F}_H'')$$

$$= |\underline{A}|^2 + |\underline{B}'|^2 + |\underline{F}_H''|^2 - 2(B')(F_H''). \quad (7)$$

Taking the difference between $|\underline{F}|^2$ and $|\bar{\underline{F}}|^2$ we get

$$|\underline{F}|^2 - |\bar{\underline{F}}|^2 = 4(B')(F_H''). \quad (8)$$

Solving for B' we get

$$B' = (|\underline{F}|^2 - |\underline{F}|^2)/(4F_H''). \quad (9)$$

Then simply $i\underline{B}' = \underline{i}B'$, and

$$i\underline{B} = i\underline{B}' + i\underline{F}_H''. \quad (10)$$

Since $\underline{F}$ is the vector sum of $\underline{A}$ and $\underline{B}$,

$$|\underline{A}| = (|\underline{F}|^2 - |\underline{B}|^2)^{\frac{1}{2}}. \quad (11)$$

The only ambiguity remaining in the phase angle $\tan^{-1}(B/A)$ is the sign of A. That problem is resolved by taking the sign of A to be the sign of F_{calc} using only the contribution from the centrosymmetric heavy atom.

From the ARAC.HCl data set were derived normalized structure factors (31), and those reflections with E's greater than 1.7 (a total of 60) were considered for further phase determination. For use in the calculation of B' from (eq. 9) and thus $\underline{B}$ and $\underline{A}$, $|\underline{F}|$ was taken to be equal to $|F_{obs}(hkl)|$, $|\bar{\underline{F}}| = |F_{obs}(h\bar{k}l)|$ (36), and $\underline{F}_H''$ was the calculated contribution to F using the $\Delta f''$ form factor for chlorine. The dispersion correction for Cl- was taken from International Tables for X-ray Crystallography(37). Phases calculated as described above were accepted if $||\underline{F}| - |\bar{\underline{F}}|| \geq 0.3$ e$^-$ (the standard deviation of most observed

F's was about 0.3 e^-), and if B' as calculated from (eq. 9) was less than $|F_{obs}(hkl)|$. A total of 31 phases were determined in this manner and were used as input to the tangent formula(31) for refinement and expansion. The result was 208 phased reflections with E's greater than 1.35 which were used to calculate a Fourier map. All atoms were easily located from this map, and the resulting structure factor calculation yielded an R index of 0.27. The 17 carbon, nitrogen and oxygen positions were among the 22 largest peaks on the map, excluding the chloride peak. Five extraneous peaks, all approximately the same height as the five weakest-appearing atoms, were easily eliminated from consideration by chemical reasoning. Three of these five were mirror images of actual atom sites, but were less than half the size of their corresponding true peaks. The original Fourier map, phased by the chloride ion and two other atoms, did indicate the positions of 15 of the 17 carbon, nitrogen, and oxygen atoms, but only if one considered the 40 largest peaks on the map.

The above results indicate that the BRR phasing method, especially when used in conjunction with the tangent formula, can be a valuable structure solution technique. The average phase error for the 31 BRR-phased reflections was about 19°, with four errors greater than 45° and one of about 110°. After tangent refinement, the average phase error for this same set of reflections was 22°, but only one phase was in error by more than 45°(about 75°).

Due to the small anomalous effects observed in the intensities, only 7 of the 31 BRR phases were more than 45° from either 0° or 180°. In some cases this might cause the tangent refinement to converge to a centric solution. This problem could possibly be avoided by also using BRR phases for those reflections where $|\underline{B}|$ was calculated to be greater than the value of F_{obs}. Assuming the large calculated $\underline{B}$ value to imply that the corresponding phase was near $\pm 90°$, a starting phase could be assigned on the basis of the calculated sign of $\underline{B}$. For the three such reflections in this study, the errors for the initial BRR phases would have been 30°, 1°, and 45°.

In comparing the two methods used in the structure solution of ARAC.HCl, the BRR phasing method used with the tangent formula produced results superior in quality compared to the traditional structure factor-Fourier method. The BRR E-map was of significantly higher quality (with respect to peak shape and resolution) than the initial Fourier map phased by the chloride and two light atoms. Another important advantage of the BRR technique is that the model is of the correct hand. The success of the method is, however, highly dependent on the quality of the data used. For a relatively weak anomalous scatterer such as Cl-, small differences between Friedel-related intensities must be reproducibly measurable. This sensitivity to small

differences also implies that absorption effects must either be small or correctable. These facts should be taken into account when considering the BRR-tangent formula method as a possible solution technique.

ARAU

The observed large magnitude of the 2 0 0 reflection suggested possible base stacking normal to the a axis, especially considering the fact that $|a|/2 = 3.44$ Å is an optimal base stacking distance(38). The Patterson function calculated at $u = 0$ indicated just that, as the familiar pattern of concentric hexagons was immediately apparent. From packing considerations, only two possible ways to orient the pyrimidine ring could be considered, making the solution problem simpler. Interpretation of the three Harker sections limited the position of the pyrimidine ring in the unit cell to only several possibilities, and the correct choice was promptly made.

Phasing on this partial structural information, since $x = 0$ for the nine atoms thus far located, yielded a centric set of phases, and images of the two enantiomorphic sugars were, to a certain degree, present on the subsequent Fourier map. Extrapolating from the ARAC.HCl structure and other known nucleoside conformations, a trial structure was developed but which had a number of defects. The R index at this point was 0.40. Phases calculated from this partial structure were then used as input to the tangent formula(31) for expansion and refinement. Two hundred twelve phases were calculated in this fashion, and the subsequent E map yielded the entire structure. The R index was 0.20.

ARACMP.$3H_2O$

The very short c axis of the crystals suggested the possibility of a solution in projection which was too great to ignore. Since no diffractometer would have been available for several weeks, hk0 intensity data were collected using a Weissenberg camera and a three-film packet in the cassette. The films were exposed for approximately 48 hours and were processed simultaneously in the normal fashion. Each film was indexed, and intensities for the diffraction maxima were assigned by comparison with a calibrated film strip. This strip had a single reflection exposed such that 24 images covered a range of intensities of 6 to 320 in a geometric progression. Average interfilm scale factors were calculated and applied to scale up those intensities measured on the second and third films. Results were averaged together to yield a list of 212 hk0 intensities. For those reflections too weak to be observed, the value assigned to the intensity was one-third that of the weakest standard spot. For those reflections whose intensities were too strong to be measured even on the third film, a value 10% larger than that of the strongest standard spot was assigned to those reflections as measured on the third film. These intensities were then scaled up by the appropriate scale factors to give final intensities. Polarization and Lorentz corrections were applied to obtain values for $|F^2|$.

Using Beevers-Lipson strips (39, 40, 41), which allow for the summation of Fourier components by hand, a uv projection Patterson was calculated (Fig. 4). Figure 5 shows the same Patterson map calculated by computer using diffractometer data collected at a later time. Figure 3 has on the same scale the four sets of phosphorus-light atom vectors. That the pseudo-twofold symmetry exhibited by these vectors is more clearly displayed by the Patterson shown in Fig. 5 than Fig. 4 may rationalize to a certain extent the lack of success in deriving a trial structure from the hand-calculated Patterson. Several phosphorus positions could satisfy that Patterson, and what proved to be the correct solution was discounted since it made use of a Patterson peak at $(u,v) = (\frac{1}{2},\frac{1}{2})$, which we thought to be a superposition of smaller Patterson vectors. At some time later a structure factor calculation was performed using the final atomic parameters, and the results were compared to the structure factors as measured from the films. The R index was about 0.21; all the very strong reflections had had their intensities seriously underestimated, and the differences in the two Pattersons (Figs. 4 and 5) perhaps reflect these poor measurements.

The structure solution of ARACMP was accomplished using a set of data collected utilizing graphite-monochromatized molybdenum radiation. These data had been collected in the usual way, but were rejected midway through refinement for

Fig. 3. A uv projection view showing the phosphorus-light-atom vectors for the structure of ARACMP.

Fig. 4. A uv projection Patterson map calculated using film data and Beevers-Lipson strips.

Fig. 5. A uv projection Patterson map calculated by computer using diffractometer-collected data.

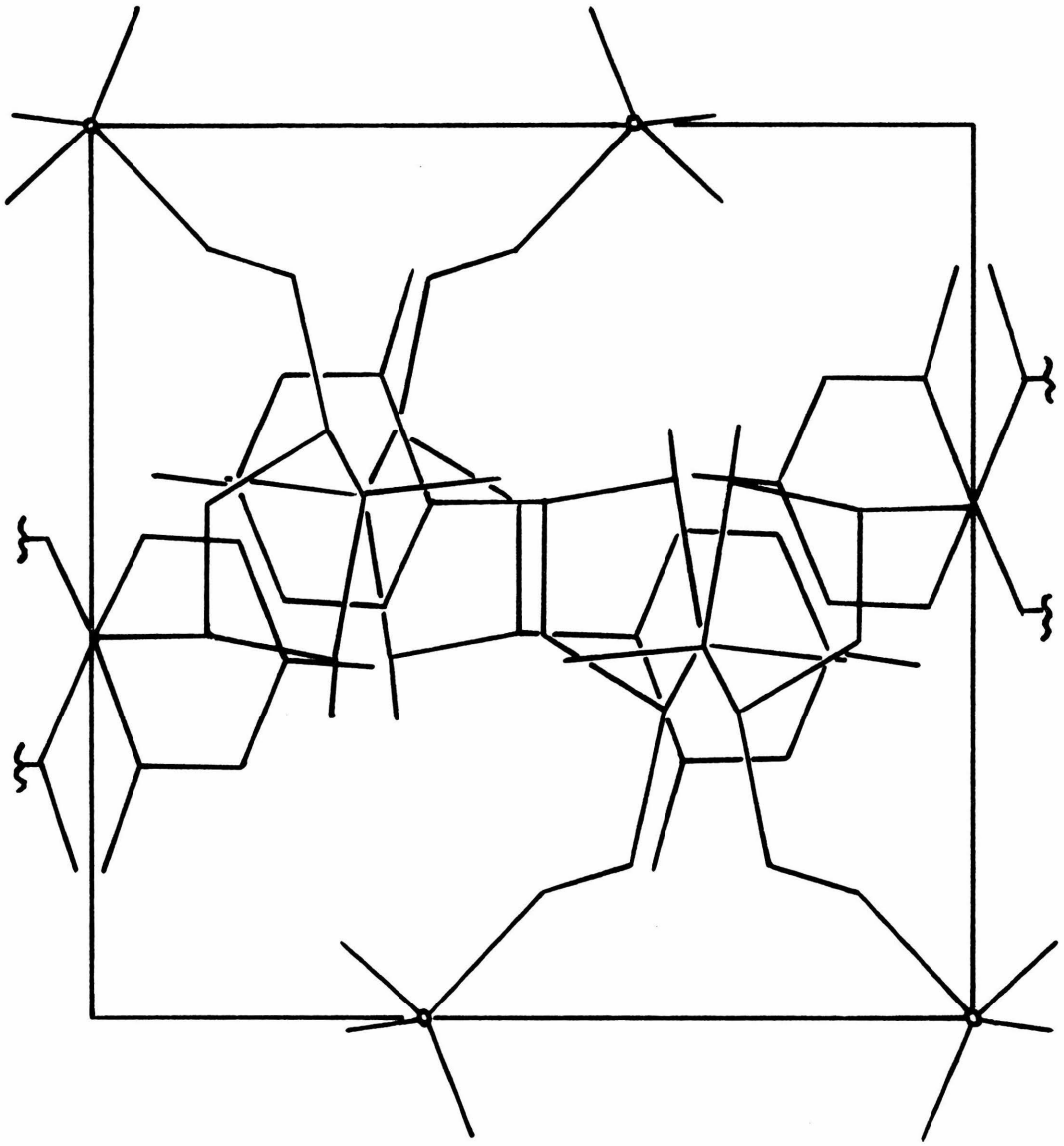


Fig. 3

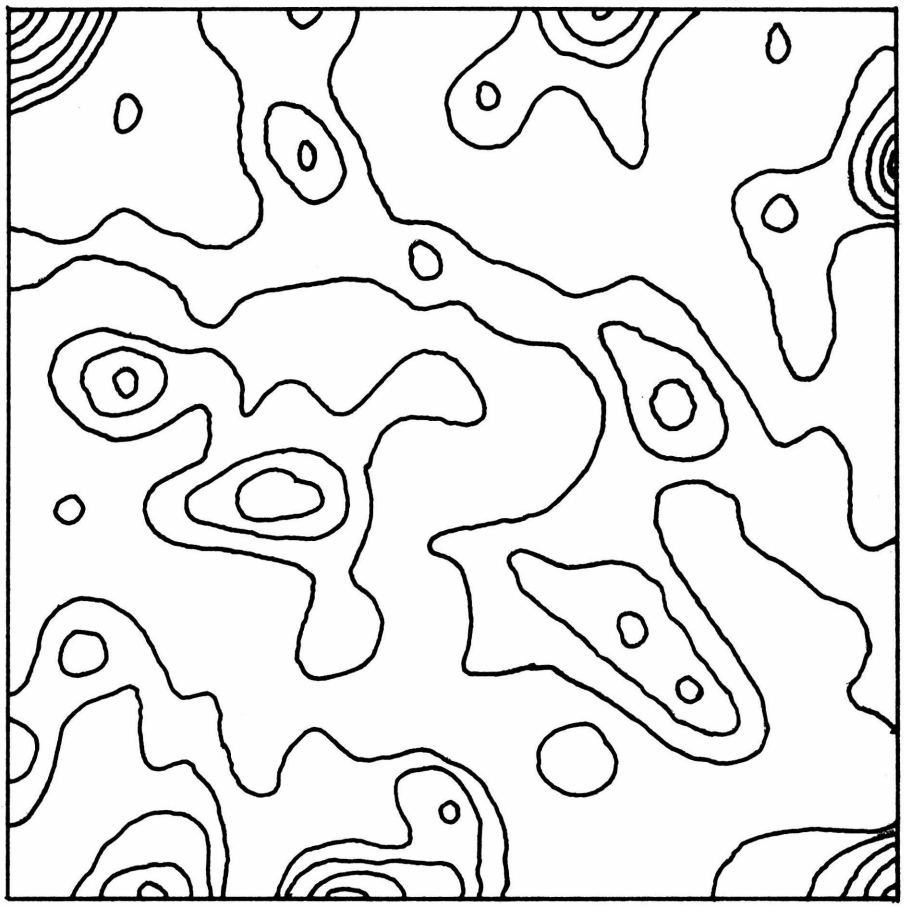


Fig. 4

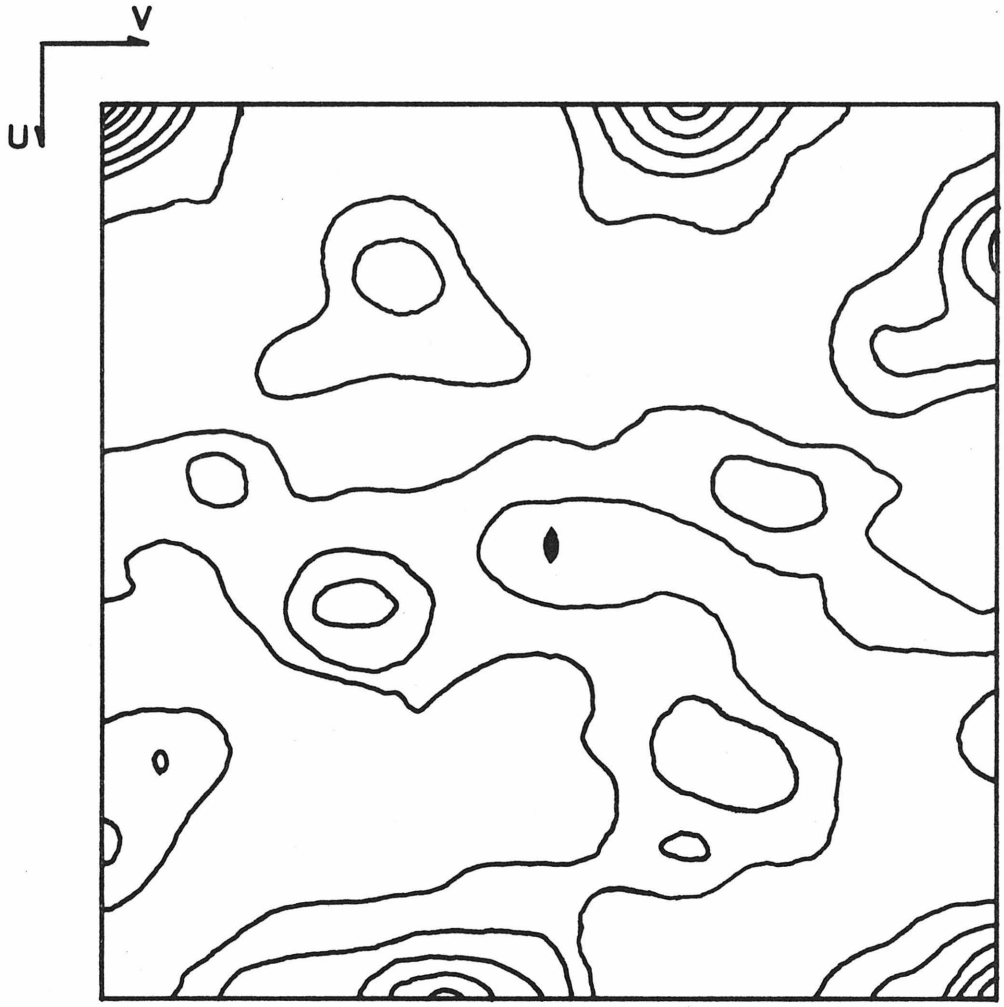


Fig. 5

lack of precision in the weaker data. The position of the phosphorus atom was determined from a three-dimensional Patterson map to be approximately (-0.25, 0.33, 0.88). Phasing on the phosphorus atom alone, due to its special position, resulted in a Fourier map showing a mirror plane at $x = \frac{1}{4}$, and the resolution was too poor to yield any new structural information.

The next step was again direct methods. Phases for four reflections were assigned to specify the origin and enantiomorph(42), the 14.0.0 reflection was assigned a phase value of π by Σ_1 considerations(31), and two symbols were also used to initiate symbolic addition(43,44),(see Table 5). Seventy-seven phased E's resulted and were expanded to a set of 208 phases by use of the tangent formula(31). It had been apparent from the start that the reflection chosen to determine the enantiomorph (the 072) could not resolve the ambiguity in the $\underline{x}$ coordinates, and thus did not accomplish its purpose. Evidence for this error was the centric set of phases as given by the tangent formula refinement. Nonetheless an E map was calculated using these phases. Two mirror-related images of the PO_4 group were easily discernible, as was most of the cytosine ring and a large fragment of the arabinose; the resolution of this map was much improved over the original Fourier. A problem still existed in correlating an arabinosyl image with the proper phosphate image, however, but by comparing a Fourier map phased by the

Table 5. Initial Phasing Data for ARACMP.$3H_2O$.

hkl	E	Ø
5 3 0	2.93	$-\pi/2$
0 3 1	2.62	$\pi/2$
7 0 2	2.25	0
0 7 2	2.86	$\pi/2$
14.0.0	3.37	π
7.16.2	2.16	A = π
7.14.1	2.79	B = 0

PO_4 group from the centrosymmetric E map with that E map, the structure was directly determined. The four oxygen atoms of the phosphate group were sufficient to break the pseudosymmetry introduced by the phosphorus atom alone, and the resulting Fourier contained one molecular image which was slightly stronger than the other. Structure factor-Fourier calculations and tangent refinement successfully yielded starting coordinates for all non-hydrogen atoms. The R index was 0.23.

REFINEMENT

Refinement of all three structures was by least-squares minimization of the quantity $\sum \underline{w}(F_o^2-k^2F_c^2)^2$ where $1/k$ is the scale factor for F_o and weights $\underline{w}$ were taken equal to $\sigma^{-2}(F_o^2)$(see DATA COLLECTION). Form factors used for C, N, and O were taken from International Tables for X-Ray Crystallography(37); those for Cl- and P were from Cromer and Waber(45). Hydrogen form factors were those of Stewart, Davidson and Simpson(46).

ARAC.HCl

Initial refinement was based on one quadrant of data. After four cycles of isotropic refinement (R = 0.081), a difference map indicated the positions of all the hydrogen atoms except those of the hydroxyl groups, appearing as peaks ranging from 0.2 to 0.5 e.$\mathring{A}^{-3}$. With the located hydrogen atoms contributing to F_c, anisotropic thermal parameters were introduced for the heavy atoms and refined for two cycles. A subsequent difference map clearly indicated the positions of the remaining hydrogen atoms. Further refinement, with a parameter list that included coordinates of all 32 atoms, anisotropic temperature parameters for the 18 heavy atoms, isotropic temperature factors for the 14 hydrogen atoms, a scale factor and an extinction parameter(47, 48), converged at an R index of 0.025.

The atomic form factors of O, N, and Cl- were then corrected for the real and imaginary effects of anomalous dispersion, using values for Cl from International Tables for X-Ray Crystallography(37) and Hope's corrections for N and O (49). Refinements of both enantiomorphs were continued until parameter shifts were less than 0.3σ for the heavy atoms and less than 1σ for the hydrogen atoms. Results of the two refinements are summarized in Table 6. Although no absorption corrections were applied to the data, we feel the differences in statistical sums are sufficient to establish

Table 6. ARAC.HCl Enantiomorph Determination.

	(L)	(D)
R	0.032	0.023
R_w	0.006	0.003
GOF	2.59	2.04

(for M = 1293 data and S = 219 parameters)

the enantiomorph. Comparison of transmission factors for representative hkl and $h\bar{k}l$ reflections showed relative differences of no more than 4-5%. Moreover, absorption effects on intensities are systematic throughout reciprocal space, while Bijvoet differences are random. Therefore we conclude that the sugar in this crystal is D-arabinose.

Coordinates and temperature parameters for the two refinements agreed within two standard deviations except for the y coordinates of the lighter atoms, which differed by an approximately constant amount of 0.045 Å. A polar dispersion error of this magnitude was predicted by Cruickshank and McDonald (50) as the result of refining an incorrect enantiomorph with the chlorine scattering factor corrected for anomalous scattering in CuKα radiation.

Final least-squares refinement of the D enantiomer was based on the full data set (two quadrants). Two matrices were collected, one containing the coordinates and anisotropic temperature factor coefficients of the heavy atoms, scale factor, and extinction parameter(47, 48), and the second containing the coordinates and isotropic temperature parameters of the hydrogen atoms. After two cycles of refinement, the largest parameter shifts associated with the two matrices were 0.4σ and 1.1σ, the R index was 0.022, the weighted R was 0.003, the goodness-

of-fit was 1.83. The final parameters are listed in Tables 7 and 8; observed and calculated structure factors are given in Table 9.

Table 7. Final ARAC.HCl Non-Hydrogen Atom Parameters and their Estimated Standard Deviations. All parameters have been multiplied by 10^4. The anisotropic temperature factor is of the form $T = \exp[-2\pi^2(h^2a^{*2}U_{11}+\ldots+2kℓb^*c^*U_{23})]$. The final value of the extinction parameter g is $(7.6\pm0.3)\times 10^{-5}e^{-2}$.

	x	y	z	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
N(1)	6095(2)	8190(1)	2506(2)	271(5)	277(5)	300(5)	-22(4)	157(4)	6(4)
C(2)	7338(2)	8505(1)	1474(2)	289(6)	328(6)	333(7)	-19(5)	166(6)	28(5)
O(2)	6807(2)	8365(1)	-468(2)	421(6)	718(8)	322(6)	-170(6)	206(5)	-44(5)
N(3)	9273(2)	8991(1)	2843(2)	306(5)	337(6)	362(6)	-60(4)	177(5)	20(5)
C(4)	10005(3)	9147(1)	5035(2)	310(6)	251(5)	376(7)	-10(5)	149(5)	-4(5)
N(4)	11909(3)	9600(1)	6169(3)	406(7)	398(7)	481(8)	-136(6)	153(6)	-41(6)
C(5)	8654(3)	8817(1)	6004(3)	414(7)	343(6)	336(7)	-34(6)	189(6)	-65(6)
C(6)	6745(3)	8352(1)	4709(3)	370(7)	320(7)	355(7)	-19(5)	218(6)	-12(5)
C(1')	4085(2)	7635(1)	1128(2)	238(6)	302(6)	314(6)	-14(5)	140(5)	27(5)
O(1')	2526(2)	7672(1)	2021(2)	315(5)	281(5)	539(6)	-7(4)	284(5)	5(4)
C(2')	4759(2)	6665(1)	1185(2)	300(6)	301(6)	383(7)	-21(5)	187(5)	-38(5)
O(2')	6445(2)	6415(1)	3333(2)	285(5)	334(5)	520(7)	39(4)	149(4)	67(4)
C(3')	2542(3)	6211(1)	804(2)	330(7)	324(6)	351(7)	-62(5)	170(6)	-25(6)

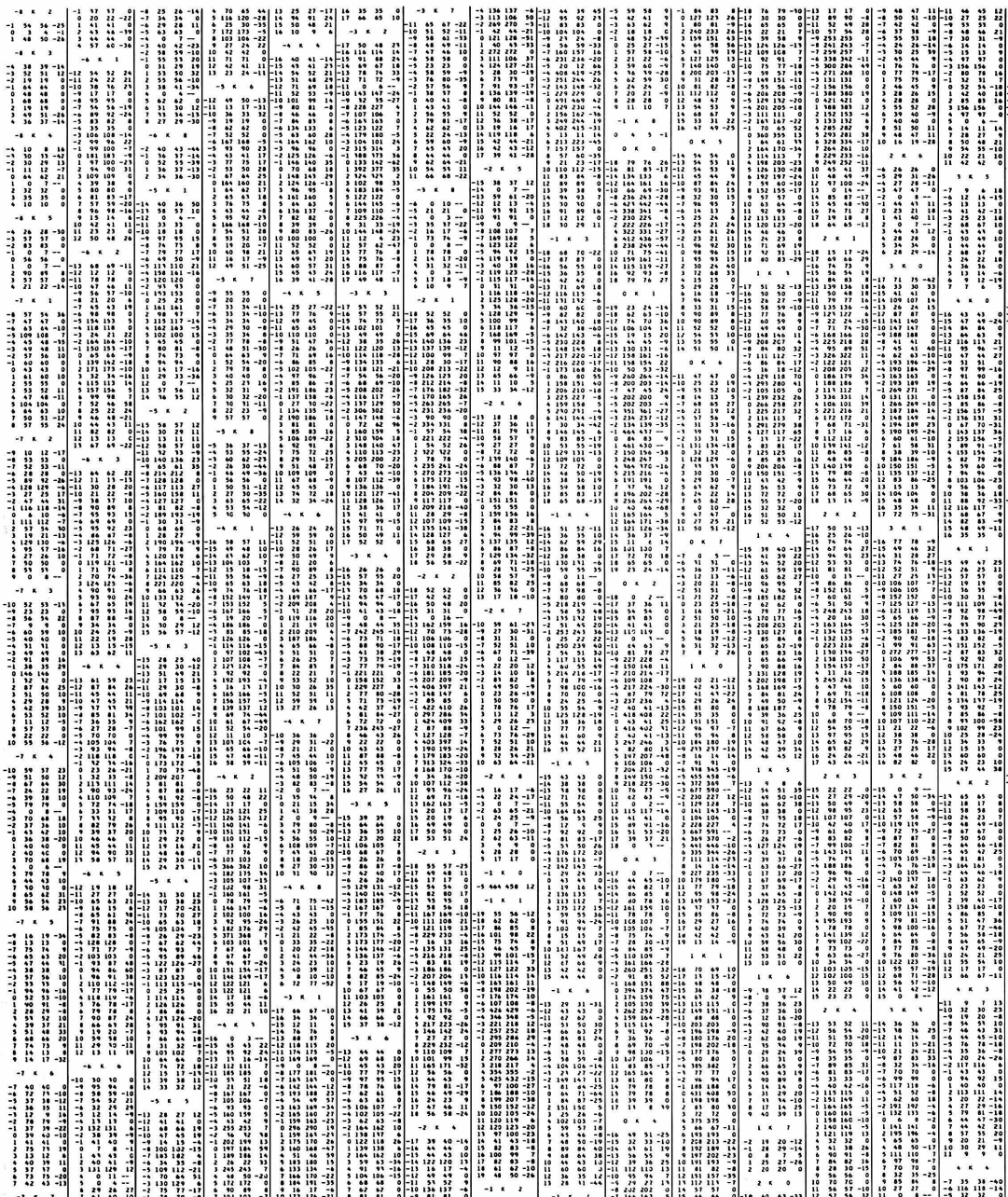
Table 7 (continued)

O(3')	1033(2)	6178(1)	-1519(2)	445(6)	604(7)	357(6)	-239(5)	186(5)	-109(5)
C(4')	1590(2)	6812(1)	1993(2)	236(5)	304(7)	343(7)	-24(5)	141(5)	30(5)
C(5')	2157(3)	6530(1)	4327(3)	370(7)	424(8)	390(8)	44(6)	213(6)	77(6)
O(5')	875(2)	7075(1)	5056(2)	410(6)	557(7)	406(6)	47(5)	242(5)	16(5)
Cl^-	-2653(1)	5000(0)	-1341(1)	367(2)	349(1)	492(2)	-16(2)	236(1)	-25(2)

Table 8. Final ARAC.HCl Hydrogen Atom Parameters and Their Estimated Standard Deviations. All coordinate values have been multiplied by 10^3.

	x	y	z	B(Å^2)
HN(3)	1011(4)	920(2)	221(4)	4.4(0.5)
HN(4)	1233(4)	977(2)	756(4)	5.9(0.6)
H'N(4)	1256(4)	982(2)	549(4)	4.4(0.5)
HC(5)	918(4)	893(2)	748(4)	5.1(0.5)
HC(6)	565(3)	811(1)	524(3)	3.3(0.4)
HC(1')	342(3)	787(1)	-24(3)	2.8(0.3)
HC(2')	518(3)	650(1)	-3(3)	2.6(0.3)
HO(2')	768(3)	671(1)	375(3)	4.0(0.4)
HC(3')	290(3)	561(1)	135(3)	3.6(0.4)
HO(3')	-16(4)	591(1)	-165(3)	3.7(0.4)
HC(4')	-6(3)	682(1)	101(3)	3.8(0.4)
HC(5')	160(4)	593(1)	439(3)	3.9(0.4)
H'C(5')	368(3)	653(1)	529(3)	3.2(0.4)
HO(5')	130(5)	687(2)	640(4)	6.4(0.6)

Table 9. Observed and Calulated Structure Factors for ARAC.HCl. Each set of four columns contains, from left to right, the k index, $10|F_o|$, $10|F_c|$, and $10|F_o^2-F_c^2|/\sigma(F_o^2)$. A reflection with (--) in the fourth column was given zero weight in the least-squares refinement.



ARAU

The initial ARAU model was refined by least squares for two cycles using one octant of data. The R index dropped to 0.096, and a difference map revealed the positions of all hydrogen atoms. Allowing the hydrogen atoms to contribute to F_c but not refining them, anisotropic thermal parameters for the heavy atoms were introduced and varied for two cycles along with non-hydrogen atom coordinates, scale factor and extinction parameter(47, 48). The R index was 0.031. Two more least-squares calculations were performed in which all parameters were allowed to shift, including coordinates and isotropic temperature factors for the hydrogen atoms. Parameters were factored into two matrices, one containing the heavy-atom coordinates and the second the anisotropic thermal parameters, hydrogen atom parameters, and the scale and extinction factors. These cycles reduced the R to 0.026.

At this time the form factors for N and O were corrected for anomalous dispersion effects in CuKα radiation according to Hope(49). Two structure factor calculations were done, one for the D and one for the L enantiomorph. Results of these two calculations are in Table 10. Differences in these sums are much less pronounced than for ARAC.HCl for at least two reasons: (1) the anomalous dispersion corrections for N and O are much less than for Cl-, such that the effect on the observed structure factors is

Table 10. ARAU Enantiomorph Determination

	(L)	(D)
R	0.0331	0.0329
R_w	0.00457	0.00454
GOF	2.20	2.08

(for M = 2034 data and S = 203 parameters)

smaller; (2) the two enantiomorphic sets of parameters had not been refined using the corrected form factors, but only compared through structure factor calculations. Had they been refined independently, the differences in the sum values could very well have been greater.

Refinement continued on the D-enantiomorph parameters in the same fashion as before, but using a data set expanded to include the two octants. Three cycles were performed; in the third, no shift was greater than $\frac{1}{2}\sigma$. The final R index was 0.030, the weighted R 0.00403, and the goodness-of-fit 1.96. Final parameters are listed in Tables 11 and 12. Observed and calculated structure factors are in Table 13.

Table 11. Final ARAU Non-Hydrogen Atom Parameters and their Estimated Standard Deviations. The y-coordinates have been multiplied by 10^5, all other parameters have been multiplied by 10^4. The anisotropic temperature factor is $T = \exp[-2\pi^2(h^2a^{*2}U_{11}+\ldots+2k\ell b^*c^*U_{23})]$. The final value of the extinction parameter g is $(6.5\pm0.5)\times 10^{-6}e^{-2}$.

	x	y	z	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
N(1)	-110(2)	20491(5)	1623(2)	396(7)	221(4)	225(5)	11(5)	-47(7)	-0(4)
C(2)	-25(2)	23573(6)	-135(2)	342(7)	227(7)	239(6)	33(6)	-37(8)	-1(5)
O(2)	-100(2)	20842(4)	-1717(1)	661(8)	290(4)	247(5)	20(5)	-44(7)	-38(4)
N(3)	126(2)	30084(5)	-22(2)	408(7)	225(4)	240(6)	34(5)	-25(7)	32(5)
C(4)	362(2)	33718(6)	1655(2)	282(7)	250(4)	314(7)	34(5)	-2(8)	-50(6)
O(4)	536(2)	39497(4)	1518(2)	490(6)	227(4)	430(6)	8(4)	4(7)	-38(5)
C(5)	348(2)	30061(6)	3434(2)	419(8)	314(7)	246(7)	32(7)	-32(8)	-66(6)
C(6)	104(2)	23761(6)	3368(2)	422(8)	325(7)	218(6)	29(7)	-26(9)	-9(6)
C(1')	-324(2)	13506(5)	1605(2)	326(7)	232(4)	255(6)	-21(6)	-24(7)	8(6)
O(1')	-1396(2)	11812(4)	3285(2)	313(5)	290(4)	386(6)	41(4)	83(6)	64(4)
C(2')	1593(2)	9861(6)	1732(2)	316(7)	238(7)	233(7)	-24(6)	22(7)	8(6)
O(2')	2962(2)	12901(4)	2943(1)	311(5)	361(4)	296(6)	-74(4)	-9(5)	26(5)

Table 11 (continued)

C(3')	879(2)	3543(6)	2566(2)	330(7)	225(7)	229(6)	-3(5)	-7(7)	5(5)
O(3')	192(2)	-99(4)	956(1)	552(7)	234(4)	299(5)	-43(5)	-6(6)	-38(4)
C(4')	-779(2)	5501(6)	3930(2)	325(7)	223(4)	303(8)	-35(6)	18(7)	20(6)
C(5')	-293(2)	5802(7)	6076(2)	423(9)	359(7)	274(7)	-25(7)	56(8)	14(7)
O(5')	1317(2)	9942(5)	6455(2)	432(6)	439(7)	319(6)	-10(6)	12(6)	72(5)

Table 12. Final ARAU Hydrogen Atom Parameters. Coordinate values have been multiplied by 10^3.

Atom	x	y	z	B(A^2)
HN(3)	-0(2)	320(1)	-103(2)	2.8(0.4)
HC(5)	52(2)	321(1)	464(2)	2.2(0.3)
HC(6)	4(2)	212(1)	451(3)	2.8(0.4)
HC(1')	-104(2)	123(1)	43(2)	1.9(0.3)
HC(2')	214(3)	91(1)	46(3)	3.2(0.4)
HO(2')	277(3)	121(1)	394(4)	6.0(0.6)
HC(3')	193(3)	12(1)	331(3)	2.8(0.4)
HO(3')	1(3)	-37(1)	131(3)	4.4(0.5)
HC(4')	-188(3)	26(1)	376(2)	2.4(0.3)
HC(5')	4(2)	14(1)	640(2)	2.7(0.3)
H'C(5')	-145(3)	72(1)	684(3)	3.6(0.4)
HO(5')	97(4)	132(1)	702(3)	5.3(0.5)

Table 13. Observed and Calculated Stucture Factors for ARAU. Each set of four columns contains, from left to right, the k index, $10|F_o|$, $10|F_c|$, and $10|F_o^2-F_c^2|/\sigma(F_o^2)$.

ARACMP.$3H_2O$

Initial least-squares refinement of the ARACMP trial model made use of a data set collected with molybdenum radiation. Isotropic refinement of all non-hydrogen atoms converged at an R index of 0.13 using all data collected out to $\sin^2\theta/\lambda^2 = 0.35$. A difference map calculated at this juncture revealed the positions of all hydrogen atoms except for that one participating in the O(10)...O(10) hydrogen bond. Two half-hydrogen atoms were used to represent the hydrogen bonding scheme here, and anisotropic refinement commenced. Two cycles in which the coordinates and six-parameter temperature factors for each heavy atom were varied along with a scale factor reduced the R index to 0.097. No hydrogen atom parameters were refined. During this stage of refinement it became increasingly clear that the quality of the weaker and high-angle data was sufficiently poor to warrant recollecting the data set using copper radiation. Experimental details for the collecting of this second set of data have been given (see EXPERIMENTAL-ARACMP).

The initial structure-factor calculation using the molybdenum radiation-refined parameters and the CuKα data set gave an R index of 0.067. Three cycles of refinement reduced the R index to 0.047. Coordinates and isotropic temperature factors for all hydrogen atoms except those associated with the water molecule O(10) were now being

refined with the heavy atom temperature factors and a scale factor; a second matrix contained the coordinates for the heavy atoms. Difference maps calculated after each cycle of refinement failed to indicated anything but a disordered hydrogen-bonding scheme along the screw axis relating O(10) to O(10), and so this model was retained through subsequent refinements.

Once again enantiomorph determination was attempted. Form factors for N and O were corrected as described previously (49), as was the form factor for P (51). Structure factor calculations for both enantiomorphic sets of parameters were performed, and the results are shown in Table 14. The differences in statistical sums favor heavily the D- structure over the L-. As one might expect, the strength of this indication is greater than that for the ARAU determination (Table 10), but not nearly as strong as for ARAC.HCl(Table 6). Although $\Delta f''$ for P in CuKα radiation is nearly that of Cl-, the fact that the "neuter parameters" were used in the ARACMP determination and refined D- and L- parameters were used in the ARAC.HCl comparison would likely make the latter a more powerful test. And although the methods of enantiomorph determination for both the ARAU and ARACMP cases were identical, the much larger $\Delta f''$ of P with respect to the $\Delta f''$s of N and O would make the ARACMP test more positive than the ARAU test. The entire preceding discussion assumes of course that in all cases the measured intensities are of equal

Table 14. ARACMP. $3H_2O$ Enantiomorph Determination

	(L)	(D)
R	0.062	0.057
R_w	0.01203	0.00993
GOF	2.47	2.25

(for M = 1859 data and S = 289 parameters)

significance statistically. By Hamilton's R-ratio test(52), all three enantiomorph determinations are significant at the $\alpha = 0.005$ level.

Continued refinement converged after 5 cycles with an R index of 0.046, a weighted R of 0.00665, and a goodness-of-fit of 1.69. All shifts were less than 0.6σ, save three. These three were for coordinate values associated with O(2'), HO(8) and O(8), which were oscillating about their means by approximately $\pm\sigma$. Hydrogen atoms associated with O(10) were dealt with in the following way: H(10) was placed as accurately as possible from previous difference maps; two half-hydrogens, H(10a) and H(10b), were placed 1/3 the distance along vectors from O(10) to O(10) atoms at $(\frac{1}{2}-x, 1-y, \frac{1}{2}+z)$ and $(\frac{1}{2}-x, 1-y, z-\frac{1}{2})$. None of these hydrogen atoms were refined in the last several cycles. Final parameters are listed in Tables 15 and 16. Observed and calculated structure factors are in Table 17.

Table 15. Final ARACMP.$3H_2O$ Non-Hydrogen Atom Parameters and their Estimated Standard Deviations. The x and y coordinates have been multiplied by 10^5; all other parameters have been multiplied by 10^4. The form of the anisotropic temperature factor is $T = \exp[-2\pi^2(h^2a^{*2}U_{11}+\ldots+klb^*c^*U_{23})]$. The final value of the extinction parameter g is $(0.6\pm0.4)\times 10^{-6}e^{-2}$.

	x	y	z	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
N(1)	3215(16)	34607(16)	7937(6)	269(14)	270(14)	323(18)	58(14)	-33(13)	-48(15)
C(2)	10348(21)	31695(20)	7396(9)	304(19)	311(19)	389(23)	54(18)	-32(20)	-26(19)
O(2)	16136(14)	33744(17)	8491(7)	311(14)	574(19)	704(24)	71(14)	-148(16)	-262(21)
N(3)	10357(17)	26174(17)	5404(7)	241(16)	351(19)	408(23)	66(14)	12(16)	-41(16)
C(4)	4133(20)	23550(17)	4093(8)	319(17)	213(17)	309(21)	21(15)	30(19)	20(17)
N(4)	4885(17)	18621(17)	2137(8)	299(17)	361(19)	481(23)	32(15)	13(19)	-111(18)
C(5)	-3005(20)	26549(20)	4832(10)	252(19)	327(20)	486(27)	1(15)	-19(19)	-76(21)
C(6)	-3145(20)	31970(19)	6711(9)	279(19)	292(19)	466(28)	41(16)	30(19)	-50(20)
P	-24472(5)	34556(5)	8810(2)	202(5)	270(5)	186(5)	-6(4)	12(4)	-16(5)
O(5')	-18626(14)	41016(14)	8144(5)	333(14)	343(14)	280(15)	-17(12)	30(11)	-51(12)
O(6)	-24964(15)	30119(13)	6209(6)	402(16)	366(14)	290(13)	-61(13)	10(16)	-40(13)
O(7)	-31694(14)	37620(15)	9887(6)	280(14)	501(17)	426(17)	-2(12)	-1(13)	-120(16)

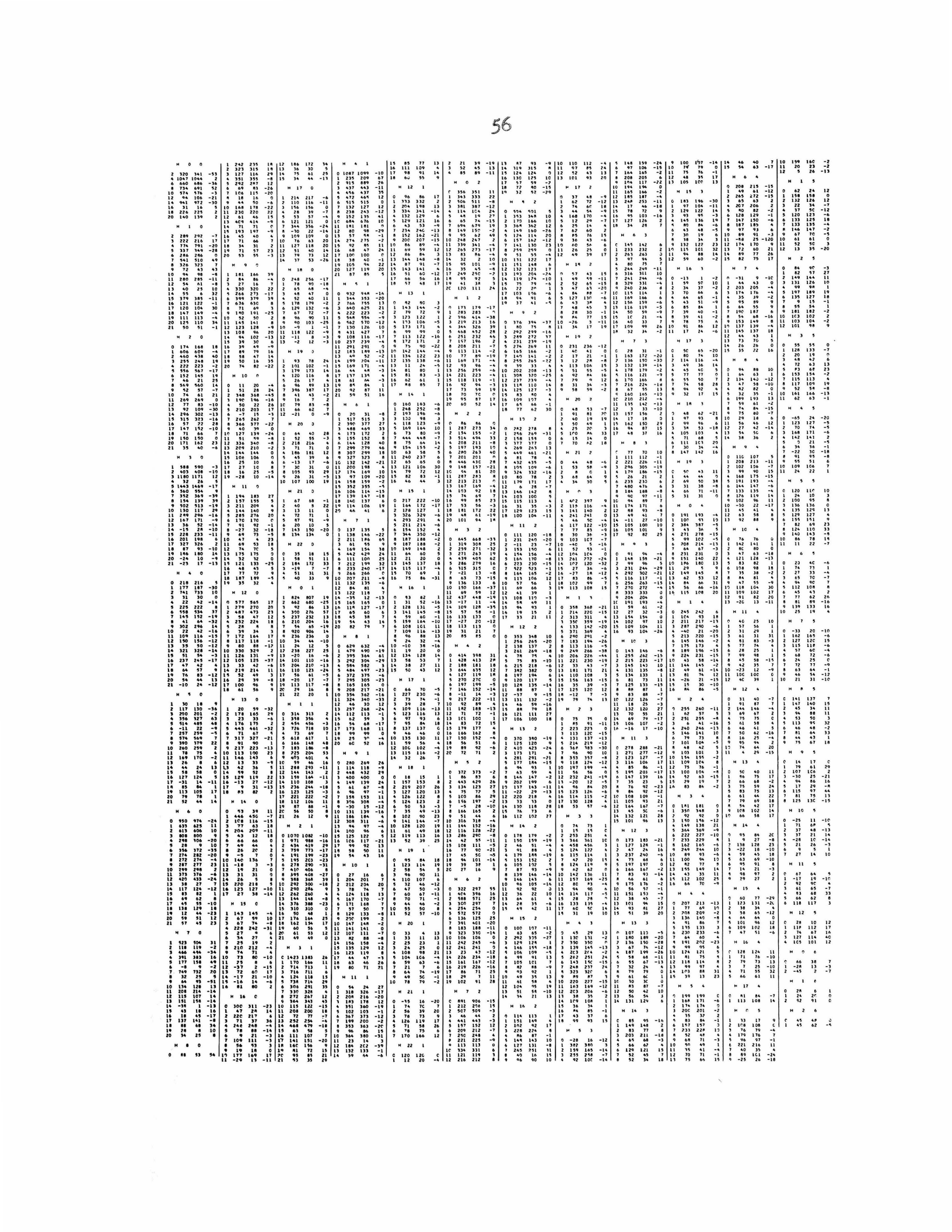
Table 15 (continued)

O(8)	-20247(15)	29873(14)	11055(6)	393(17)	454(17)	321(16)	34(13)	67(16)	61(16)
C(1')	2848(20)	41043(19)	9821(8)	297(19)	309(19)	254(20)	35(15)	-47(16)	-16(18)
O(1')	-4260(14)	41129(12)	11156(6)	371(12)	282(12)	277(13)	43(11)	26(14)	39(13)
C(2')	3560(21)	48357(19)	8247(8)	311(19)	305(19)	281(23)	-15(16)	-54(17)	-2(17)
O(2')	7674(15)	47845(15)	5736(6)	343(16)	457(17)	367(18)	-10(13)	64(14)	20(16)
C(3')	-4784(21)	50304(19)	7763(8)	336(20)	277(19)	233(20)	-17(16)	-37(18)	28(16)
O(3')	-5522(14)	57879(14)	7327(7)	375(16)	305(14)	579(20)	-12(13)	-100(16)	151(15)
C(4')	-8460(21)	47706(19)	10418(7)	350(20)	240(19)	237(22)	41(15)	18(17)	-29(17)
C(5')	-16831(21)	46096(20)	10365(8)	353(20)	324(22)	327(25)	20(17)	49(18)	-64(19)
O(9)	31673(16)	34259(16)	4029(7)	525(19)	424(19)	631(22)	-2(17)	48(20)	-17(22)
O(10)	23258(18)	46555(18)	5033(8)	472(19)	616(20)	836(27)	19(16)	76(20)	24(22)
O(11)	39610(16)	37423(15)	9120(7)	464(19)	498(19)	500(24)	-54(15)	37(20)	-65(20)

Table 16. Final ARACMP.$3H_2O$ Hydrogen Atom Parameters and Their Estimated Standard Deviations. Coordinate values have been multiplied by 10^3.

Atom	x	y	z	B(Å^2)
HN(3)	147(2)	246(2)	490(11)	3.4(1.0)
HN(4)	95(3)	168(3)	206(11)	5.4(1.3)
H'N(4)	8(2)	165(2)	155(9)	3.3(1.0)
HC(5)	-74(2)	249(2)	399(10)	2.8(0.9)
HC(6)	-74(2)	343(2)	709(9)	3.9(1.0)
HC(1')	67(2)	405(2)	1110(9)	2.4(0.8)
HC(2')	60(2)	521(2)	953(8)	1.5(0.7)
HO(2')	94(4)	499(4)	585(16)	10.7(2.3)
HC(3')	-72(2)	475(2)	645(8)	0.9(0.7)
HO(3')	-99(2)	586(3)	683(11)	4.3(1.1)
HC(4')	-75(2)	512(2)	1165(8)	2.2(0.9)
HC(5')	-185(2)	435(2)	1212(8)	2.5(0.9)
H'C(5')	-192(2)	505(2)	998(10)	3.9(1.1)
H(8)	-225(3)	295(3)	1184(12)	4.4(1.5)
H(9)	299(3)	305(3)	383(14)	6.3(1.5)
H'(9)	281(4)	376(4)	440(18)	11.1(2.3)
H(11)	367(3)	353(3)	1015(13)	5.2(1.5)
H'(11)	376(3)	379(3)	779(14)	6.2(1.7)
H(10)	191(0)	451(0)	523(0)	12.0(0)
H(10a)	247(0)	488(0)	673(0)	12.0(0)
H(10b)	247(0)	488(0)	339(0)	12.0(0)

Table 17. Observed and Calculated Structure Factors for ARACMP.$3H_2O$. Each set of four columns contains, from left to right, the h index, $10|F_o|$, $10|F_c|$, and $10|F_o^2-F_c^2|/\sigma(F_o^2)$.



STRUCTURE DESCRIPTIONS

ARAC.HCl

Distances and angles for the cation are shown in Fig. 6 and Table 18. One anomaly in the bond distances is the C(1')-C(2') distance of 1.540 Å. This distance compares with 1.508 Å found in 2'-deoxycytidine hydrochloride(53) and 1.510 Å and 1.513 Å for the monoclinic and orthorhombic forms of cytidylic acid (54,55). This difference is large enough to be significant and could be the result of the close contact of 2.754 Å between O(2') and N(1). That this contact is uncomfortably short is verified by the angle O(2')-C(2')-C(1') being 4° larger than the average value for the C(2')-*endo* ribose ring(56,57). (The term C(2')-*endo* implies that for a hexofuranose, ring atom C(2') deviates from the best four-atom plane *toward* the same side of the plane as C(5'). See Fig. 7). Distances and angles in the cytosine moeity agree fairly well with those compiled by Voet and Rich (58).

The arabinofuranosyl residue is in the *anti*, *gauche*-*trans*, and C(2')-*endo* conformations (59,60,56). *Anti* describes the molecular conformation about the C(1')-N(1) glycosidic bond, and *gauche*-*trans* describes the conformation about the C(4')-C(5') bond. The various torsion angles describing these conformations are given in Table 28. Figure 7 shows a view of the ARAC cation as viewed down the C(1')-N(1) bond The glycosidic torsion

Fig. 6. Distances and angles involving the non-hydrogen atoms of ARAC.HCl. Estimated standard deviations are approximately 0.002 Å and 0.02°.

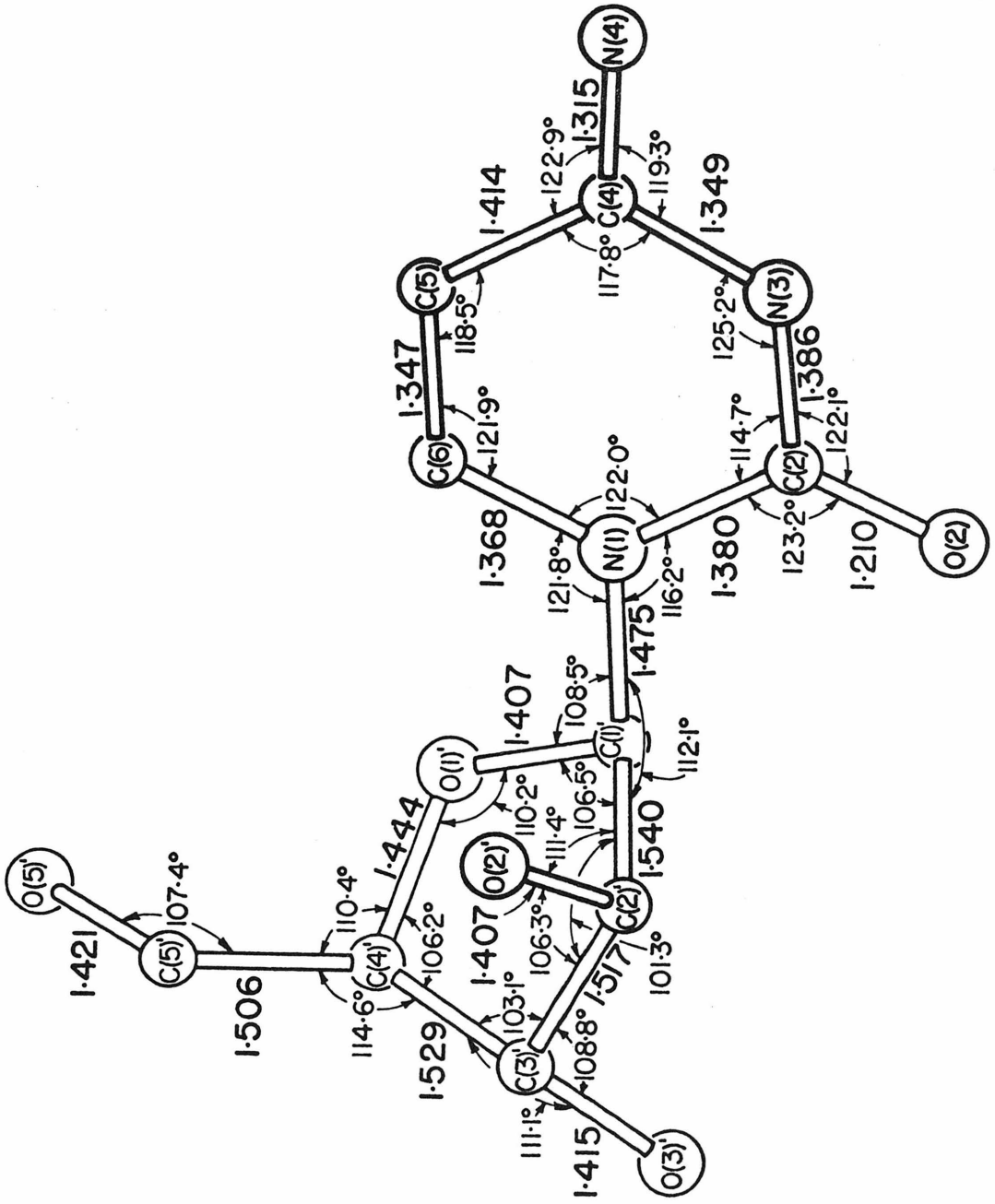
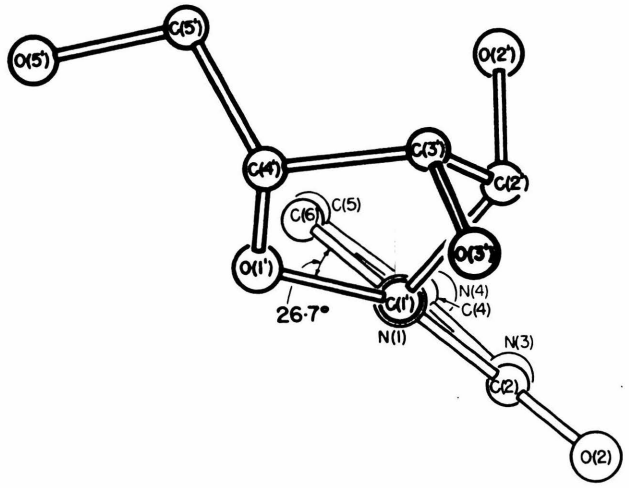


Table 18. Bond Distances and Angles Involving the Hydrogen Atoms in ARAC.HCl. Formal esd's are about 0.02-0.03 Å and 1.5-2.0°.

Distance		Angle			
HN(3) - N(3)	0.90 Å	HC(5) - C(5) - C(4)	116°	HN(3) - N(3) - C(2)	116°
HN(4) - N(4)	0.89	HC(5) - C(5) - C(6)	125	HN(3) - N(3) - C(4)	119
H'N(4) - N(4)	0.83	HC(6) - C(6) - C(5)	124	HN(4) - N(4) - C(4)	121
HC(5) - C(5)	0.90	HC(6) - C(6) - N(1)	115	H'N(4)- N(4) - C(4)	119
HC(6) - C(6)	1.01	HC(1') - C(1') - N(1)	106	H'N(4)- N(4) - HN(4)	118
HC(1') - C(1')	0.90	HC(1') - C(1') - C(2')	114		
HC(2') - C(2')	1.01	HC(1') - C(1') - O(1')	109		
HC(3') - C(3')	0.98	HC(2') - C(2') - C(1')	114		
HC(4') - C(4')	0.97	HC(2') - C(2') - C(3')	110		
HC(5') - C(5')	1.00	HC(2') - C(2') - O(2')	113		
H'C(5')- C(5')	0.90	HC(3') - C(3') - C(2')	109		
HO(2') - O(2')	0.85	HC(3') - C(3') - O(3')	108		
HO(3') - O(3')	0.85	HC(3') - C(3') - C(4')	117		
HO(5') - O(5')	0.88	HC(4') - C(4') - C(3')	104		
		HC(4') - C(4') - O(1')	109		
		HC(4') - C(4') - C(5')	112		
		HC(5') - C(5') - C(4')	113		
		HC(5') - C(5') - O(5')	104		
		H'C(5')- C(5') - HC(5')	105		
		H'C(5')- C(5') - O(5')	113		
		H'C(5')- C(5') - C(4')	114		
		HO(2') - O(2') - C(2')	112		
		HO(3') - O(3') - C(3')	105		
		HO(5') - O(5') - C(5')	101		

Fig. 7. The ARAC cation as viewed down the C(1')-N(1) bond. The effect of the displacement of C(1') from the pyrimidine plane can be seen in the slight tip of that plane from the line of sight.



angle has a value of 26.7°, restricted from varying very much from that value by the short contacts between O(2') and the cytosine ring - in particular with O(2) and HC(6). This restriction can be seen more clearly in the stereoview of the ARAC cation shown in Fig. 8. The more common gauche-gauche conformation(56) about the C(4')-C(5') bond has yielded to a C(2')-endo puckering of the arabinose ring. With O(5') lying over the sugar ring (gauche-gauche) and C(2')-endo, O(5') and O(2') would approach quite close together - in fact, close enough to form an intramolecular hydrogen bond (see below - ARAU).

A stereoscopic view of the structure is shown in Fig. 9. Of the three hydroxyl groups, two - O(2') and O(5')- donate hydrogen bonds to other hydroxyl groups; O(3') is hydrogen-bonded to a chloride ion. All three hydrogen atoms associated with nitrogen atoms form hydrogen bonds with chloride ions, resulting in an irregular fourfold coordination about the chloride ion (Fig. 10). Details of the hydrogen bonds are given in Table 19.

The six-atom pyrimidine nucleus is very nearly planar, with the extracyclic substituents deviating from that plane (Table 20). Of particular note is the large displacement of C(1') (see Fig. 7), which may be a manifestation of the interaction of O(2') and N(1) mentioned above.

Fig. 8. A stereographic illustration of the ARAC.HCl cation, as drawn by OR TEP(61). Thermal ellipsoids are drawn at the 50% probability level. Hydrogen atoms are represented as spheres of arbitrary radius.

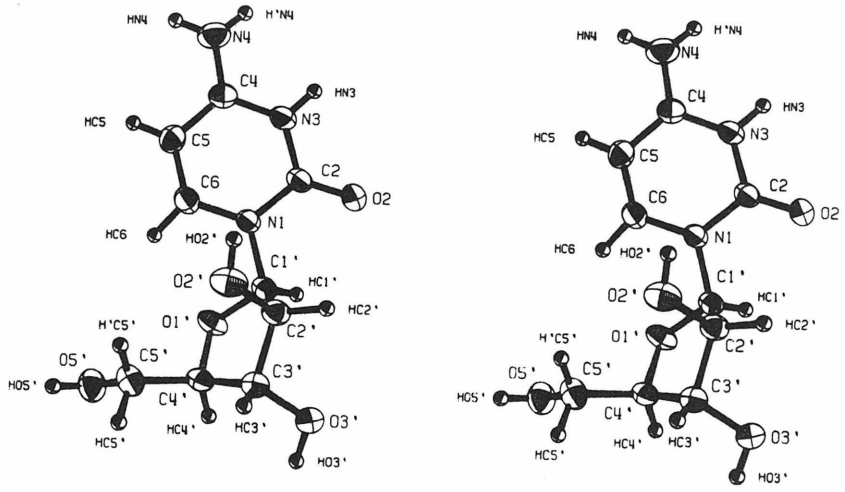


Fig. 9. A stereoscopic view(61) of the crystal structure of ARAC.HCl.

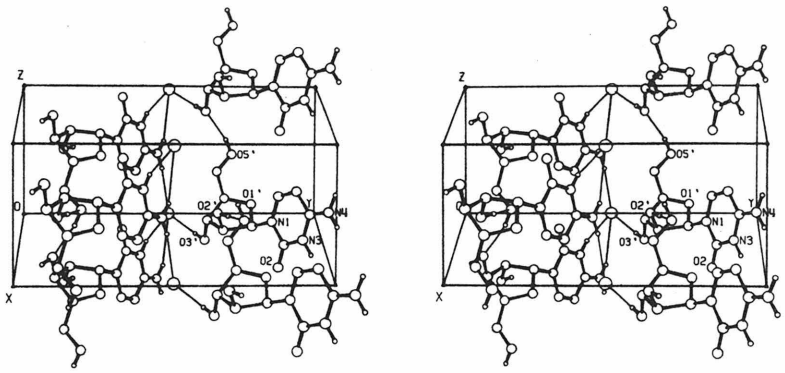


Fig. 10. A stereoscopic view(61) showing the environment of the chloride ion. Thermal ellipsoids are drawn at the 50% probability level.

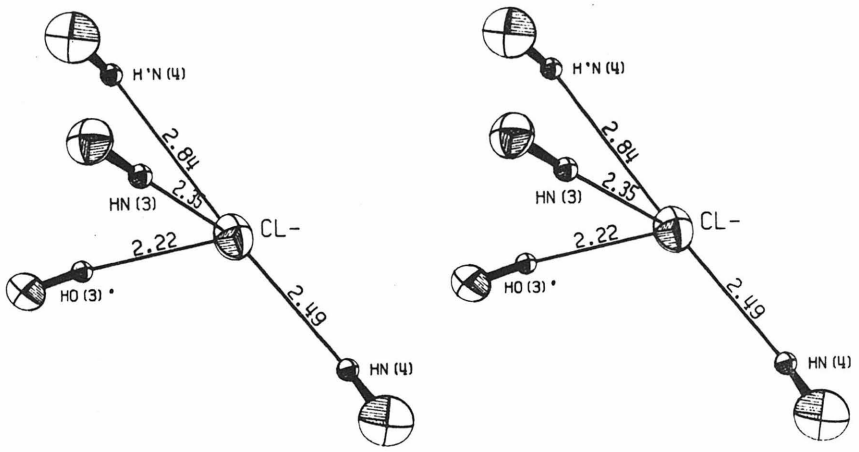


Table 19. Distances and Angles for the Hydrogen Bonds, D-H...A, as Found in the Structure ARAC.HCl.

D	H	A	Distance D-A	Distance H...A	Angle D-H...A
N(4)	HN(4)	Cl- (a)	3.354 Å	2.49 Å	165°
N(4)	H'N(4)	Cl- (b)	3.568	2.84	147
N(3)	HN(3)	Cl- (b)	3.221	2.35	165
O(3')	HO(3')	Cl-	3.054	2.25	166
O(5')	HO(5')	O(3')(c)	2.649	1.83	155
O(2')	HO(2')	O(5')(d)	2.762	1.93	165

(a) at 1-x, $\frac{1}{2}$+y, 1-z
(b) at 1-x, $\frac{1}{2}$+y, -z
(c) at x, y, 1+z
(d) at 1+x, y, z

Table 20. Deviations from the Best Plane of the Pyrimidine Ring in the Structure of ARAC.HCl. The least-squares plane was passed through the six ring atoms, N(1)-C(6), all weighted equally. Its direction cosines relative to a, b and c are -0.486, 0.839 and 0.007; the origin-to-plane distance is 8.566 Å.

Atom	Dev.(Å)	Atom	Dev.(Å)	Atom	Dev.(Å)
N(1)	-0.006	C(1')	-0.085	HC(5)	-0.019
C(2)	-0.003	O(2)	-0.022	HC(6)	0.049
N(3)	0.011	N(4)	-0.030	Cl(a)	0.272
C(4)	-0.010	HN(3)	-0.005	Cl(b)	0.223
C(5)	0.002	HN(4)	0.062	O(2)(c)	0.027
C(6)	0.006	H'N(4)	0.048		

(a) at 1-x, $\frac{1}{2}$+y, 1-z
(b) at 1-x, $\frac{1}{2}$+y, -z
(c) at x, y, 1+z

Thermal ellipsoids for the cytosine cation and the chloride ion are shown in Figs. 8 and 10. The largest mean-square displacements are exhibited by atoms O(3') and O(2) and are directed approximately parallel to the view vector in Fig. 8. The orientation of the chloride ion thermal ellipsoid seems to be primarily determined by its two shortest hydrogen bonds, involving HN(3) and HO(3').

ARAU

As was mentioned in the description of the ARAC.HCl cation, the combination of C(2')-endo and gauche-gauche conformations for an arabinofuranose would result in a close O(2')...O(5') contact. As can be seen in Fig. 11, this is just what has happened in the structure of ARAU. The O(2')...O(5') hydrogen bond distance of 2.72 Å suggests strongly the feasibility of a 2'-5' cyclic monophosphate similar to 2'-3' cyclic AMP. In fact there is some evidence that such a cyclic compound does exist(5). Two structural reports of the crystal structure of ARAC(23,24) show this nucleoside to be also in the C(2')-endo, gauche-gauche conformation. While one is a preliminary report only(23), the other does confirm the presence of an internal hydrogen bond(24). Film investigation of ARAC crystals obtained from a dimethylsulfoxide solution resulted in a significantly different unit cell than that already published(23,24), but no further work has been done with those crystals.

Distances and angles for ARAU are given in Fig. 12 and Table 21. In this nucleoside the interaction between sugar and pyrimidine ring is not quite as severe as that found for ARAC.HCl. The O(2')...N(1) distance has lengthened in this case to 2.797 Å, 0.043 Å longer than the same distance in ARAC.HCl, even though the shortening of the C(1')-C(2') bond distance by some 0.012 Å would have a shortening effect on that contact distance. The two exocyclic bond angles

Fig. 11. A view of the ARAU molecule looking down the C(1')-N(1) bond. The intramolecular bond between O(2') and O(5') is shown.

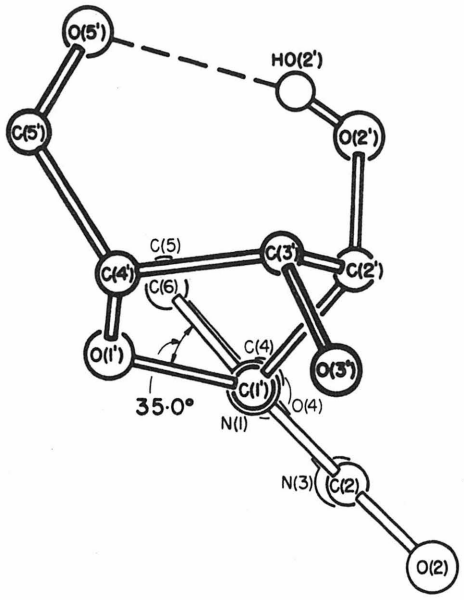


Fig. 12. Heavy-atom distances and angles for the ARAU molecule. Estimated standard deviations are 0.002 Å and 0.3°.

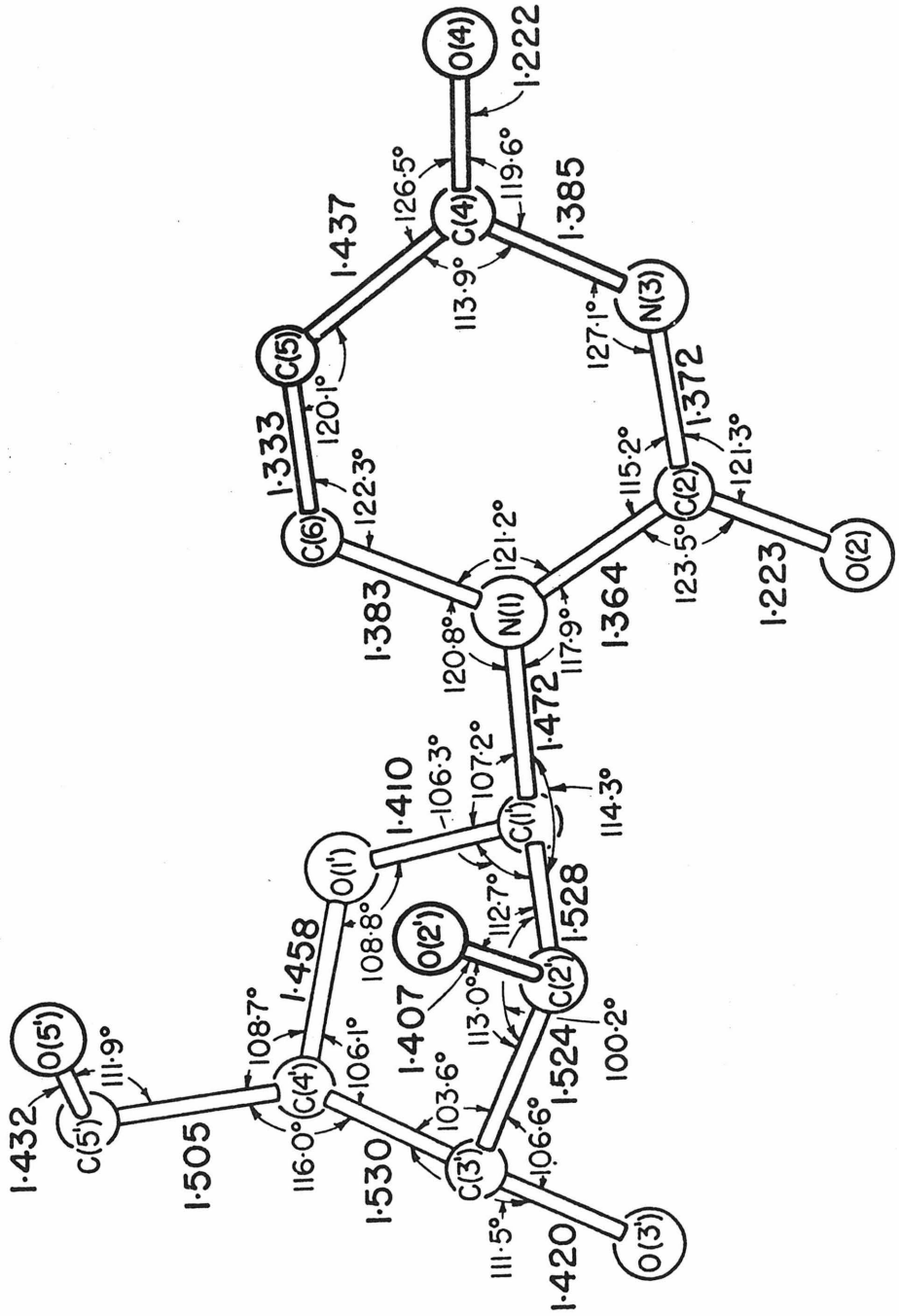


Table 21. Bond Distances and Angles Involving the Hydrogen Atoms in ARAU. Estimated standard deviations are about 0.03 Å and 2°.

Distance			Angle				
HN(3)	- N(3)	0.81 Å	HN(3)	- N(3)	- C(2)	116°
HC(5)	- C(5)	0.94	HN(3)	- N(3)	- C(4)	117
HC(6)	- C(6)	0.95	HC(5)	- C(5)	- C(4)	119
HC(1')	- C(1')	0.97	HC(5)	- C(5)	- C(4)	120
HC(2')	- C(2')	0.96	HC(6)	- C(6)	- C(5)	123
HC(3')	- C(3')	1.01				
HC(4')	- C(4')	0.97				
HC(5')	- C(5')	0.98				
H'C(5')	- C(5')	1.00				
HO(2')	- O(2')	0.72				
HO(3')	- O(3')	0.81				
HO(5')	- O(5')	0.82				

Angle (cont'd)				
HC(6)	- C(6)	- N(1)	115°
HC(1')	- C(1')	- N(1)	108
HC(1')	- C(1')	- C(2')	111
HC(1')	- C(1')	- O(1')	110
HC(2')	- C(2')	- C(1')	112
HC(2')	- C(2')	- O(2')	110
HC(2')	- C(2')	- C(3')	109
HC(3')	- C(3')	- C(2')	112
HC(3')	- C(3')	- O(3')	112
HC(3')	- C(3')	- C(4')	111
HC(4')	- C(4')	- C(3')	110
HC(4')	- C(4')	- O(1')	107
HC(4')	- C(4')	- C(5')	108
HC(5')	- C(5')	- C(4')	103
HC(5')	- C(5')	- O(5')	111
H'C(5')	- C(5')	- HC(5')	111
H'C(5')	- C(5')	- O(5')	110
H'C(5')	- C(5')	- C(4')	110
HO(2')	- O(2')	- C(2')	109
HO(3')	- O(3')	- C(3')	109
HO(5')	- O(5')	- C(5')	111

O(2')-C(2')-C(1') and C(2')-C(1')-N(1) have increased, however, by 1 and 2° - the first increase perhaps due in part to the O(2')...O(5') hydrogen bond. One possible manifestation of these dimension changes is that C(1') is now very nearly in the plane of the six-atom pyrimidine ring, as can be seen from Fig. 11 (see also Table 22). Distances and angles for the uracil fragment show no significant discrepancies from the average values reported by Voet and Rich (58)

The crystal packing of ARAU is shown in Fig. 13. The pseudo-translational relationships between symmetry-related uracil moeities is apparent here, and were responsible for the relative ease of the structure solution by Patterson techniques. All oxygen and nitrogen atoms save O(1') and N(1) participate in the four independent hydrogen bonds; O(3'), while capable of being both a donor and acceptor, only donates - to O(4). There are no outstandingly short hydrogen bonds; actually the O(2')...O(5') interaction is the shortest one observed in the structure. Distances and angles for the hydrogen bonds are given in Table 23. Base stacking also contributes to the crystal packing forces, as can be seen in Fig. 13. As has been noted (38), such stacking does not require complete overlap of the heterocycles, and in fact rarely does. The base stacking distance is 3.44 Å.

Table 22. Deviations from the Plane of the Pyrimidine Ring in the Structure of ARAU. The least-squares plane was passed through the six ring atoms, N(1)-C(6), all weighted equally. Its direction cosines relative to a, b, and c are -0.993, 0.110, 0.050; the origin-to-plane distance is 0.586Å.

Atom	Dev.	Atom	Dev.	Atom	Dev.
N(1)	0.017 Å	C(1')	0.001 Å	O(5') (a)	-1.377Å
C(2)	-0.029	O(2)	-0.094	O(2') (b)	1.564
N(3)	0.021	O(4)	0.010	O(3') (c)	0.833
C(4)	0.001	HN(3)	0.12		
C(5)	-0.014	HC(5)	-0.04		
C(6)	0.005	HC(6)	0.03		

(a) at x, y, z-1
(b) at $x-\frac{1}{2}$, $\frac{1}{2}-y$, -z
(c) at -x, $y+\frac{1}{2}$, $\frac{1}{2}-z$

Fig. 13. The crystal structure of ARAU as viewed down the a axis. The positive b axis is from left to right as the drawing is viewed.

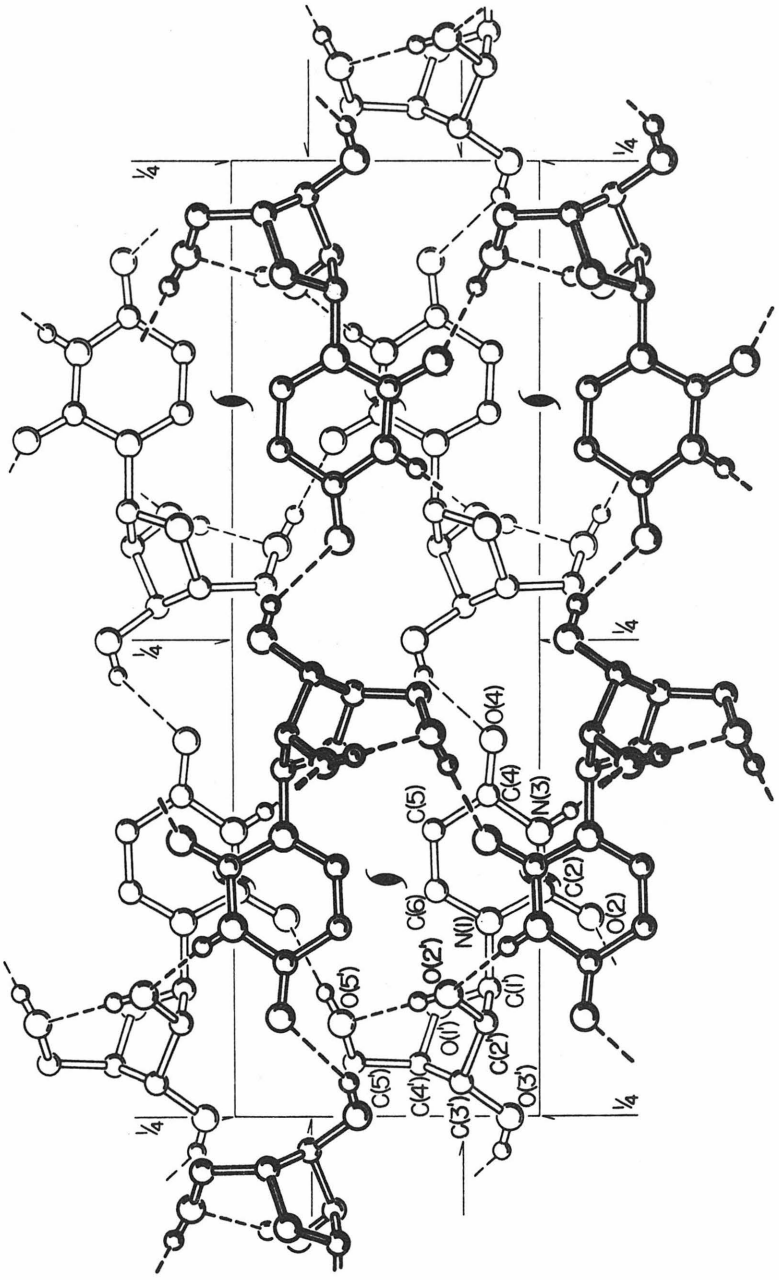
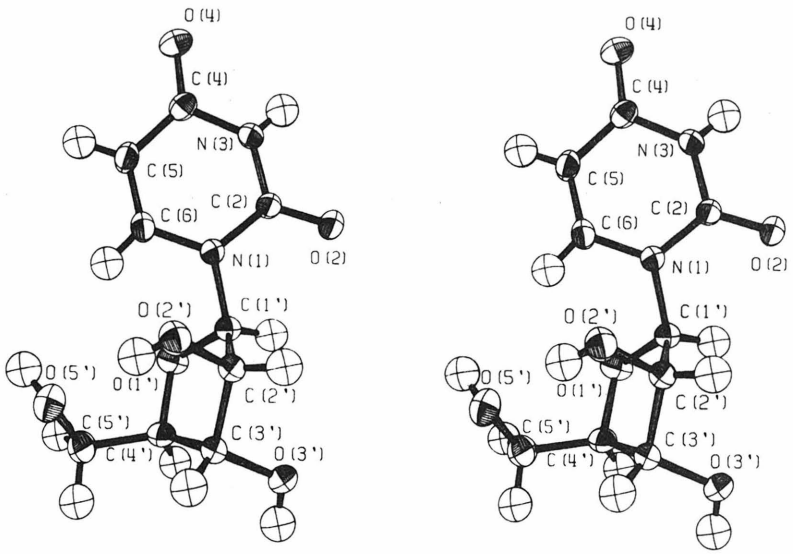


Table 23. Distances and Angles for the Hydrogen Bonds D-H...A as Found in the Crystal Structure of ARAU.

D	H	A		Distance D-A	Distance H...A	Angle D-H...A
O(5')	HO(5')	O(2)	(a)	2.782 Å	1.97 Å	175°
N(3)	HN(3)	O(2')	(b)	2.892	2.19	146
O(3')	HO(3')	O(4)	(c)	2.826	2.09	152
O(2')	HO(2')	O(5')		2.724	2.04	161

(a) at x, y, z+1
(b) at $x+\frac{1}{2}$, $\frac{1}{2}-y$, -z
(c) at -x, $y-\frac{1}{2}$, $\frac{1}{2}-z$

Fig. 14. A stereoscopic drawing of the ARAU molecule(61). Thermal ellipsoids are drawn at the 50% probability level. Hydrogen atoms are represented as spheres of arbitrary radius.



A stereoscopic view of the ARAU molecule is shown in Fig. 14. The largest thermal motions are exhibited by atoms O(2), O(4), and O(3'); exocyclic base substituents commonly vibrate like this. The size and orientation of the O(3') thermal ellipsoid can be explained by its participation in only one hydrogen bond, which in turn limits that atom's thermal motion in the direction of the hydrogen bond.

ARACMP.$3H_2O$

The sugar ring folding in ARACMP is C(3')-*endo* while the conformation about the C(4')-C(5') bond is *gauche*-*gauche*. A P-O(5')-C(5')-C(4') torsion angle of 137° extends the phosphate group away from the arabinose, however, in a fashion common to most nucleotide structures already studied(60,62). The glycosidic torsion angle is 30.7°, making the conformation about the C(1')-N(1) bond *anti*.

Distances and angles for ARACMP are given in Fig. 15 and Table 24. Dimensions for the cytosine fragment agree very well with those of ARAC.HCl, the averages of Voet and Rich(58), and the values as found in 1-methylcytosine hydrochloride(63). The arabinose moeity is again showing possibly significant deviations from expected bond lengths and angles. The only other reported precise determination of an arabinofuranosyl nucleoside crystal structure (besides those reported here) is that of arabinofuranosyl-4-thiouracil(27). In that structure the sugar is also C(3')-*endo* and *gauche*-*gauche*. With respect to those reported bond lengths, for ARACMP the bond C(2')-C(3') is longer by 0.016 Å and the bond C(3')-O(3') is shorter by 0.014 Å. Other values for molecular dimensions are essentially equal for the two structures.

Saenger has compiled average dimensions of C(2')-*endo* and C(3')-*endo* ribose units and has shown that they differ significantly(57). One might expect then a difference in

Fig. 15. Distances and angles involving the non-hydrogen atoms in ARACMP. Estimated standard deviations are 0.005 Å and 0.5°.

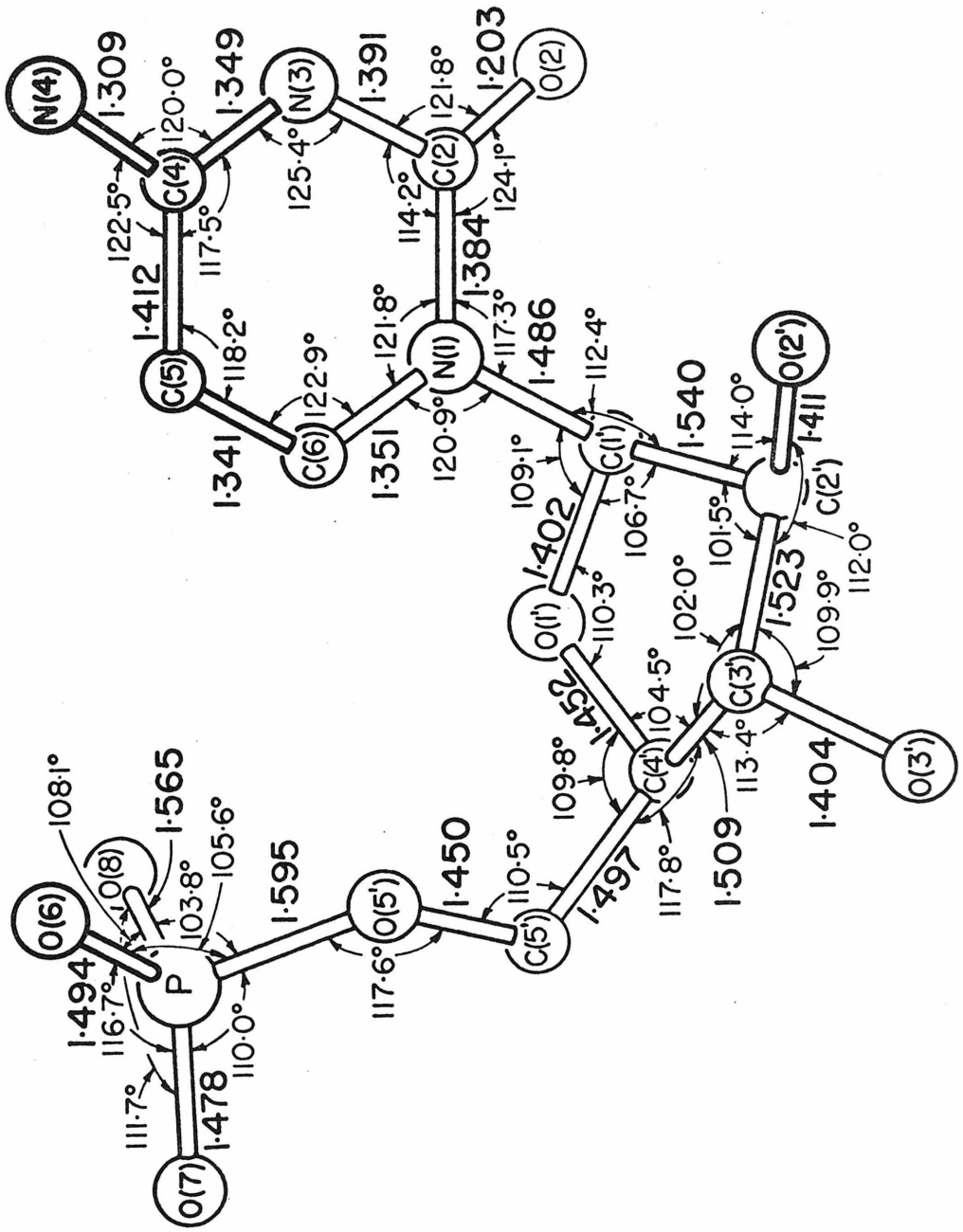


Table 24. Bond Distances and Angles Involving the Hydrogen Atoms in ARACMP.$3H_2O$. Formal esd's are about 0.05 Å and 3-4°.

Distance		Angle	
HN(3) - N(3)	0.85 Å	HN(3) - N(3) - C(2)	117°
HN(4) - N(4)	0.88	HN(3) - N(3) - C(4)	118
H'N(4) - N(4)	0.87	HN(4) - N(4) - C(4)	112
HC(5) - C(5)	0.93	H'N(4) - N(4) - C(4)	117
HC(6) - C(6)	0.88	H"N(4) - N(4) - HN(4)	126
HC(1') - C(1')	0.91		
HC(2') - C(2')	1.02		
HC(3') - C(3')	0.92		
HC(4') - C(4')	0.89		
HC(5') - C(5')	1.01		
H'C(5') - C(5')	0.93		
HO(2') - O(2')	0.48		
HO(3') - O(3')	0.82		
HO(8) - O(8)	0.55		
HO(9) - O(9)	0.76		
H'O(9) - O(9)	0.89		
H(10) - O(10)	0.79		
H(10a) - O(10)	0.95		
H(10b) - O(10)	0.93		
H(11) - O(11)	0.81		
H'(11) - O(11)	0.73		

Angle	
HC(5) - C(5) - C(4)	120°
HC(5) - C(5) - C(6)	122
HC(6) - C(6) - C(5)	121
HC(6) - C(6) - N(1)	116
HC(1') - C(1') - N(1)	107
HC(1') - C(1') - O(1')	110
HC(1') - C(1') - C(2')	112
HC(2') - C(2') - C(1')	109
HC(2') - C(2') - C(3')	110
HC(2') - C(2') - O(2')	110
HC(3') - C(3') - C(2')	115
HC(3') - C(3') - C(4')	102
HC(3') - C(3') - O(3')	114
HC(4') - C(4') - C(3')	105
HC(4') - C(4') - C(5')	110
HC(4') - C(4') - O(1')	110
HC(5') - C(5') - C(4')	111
HC(5') - C(5') - O(5')	105
HC(5') - C(5') - H'C(5')	116
H'C(5') - C(5') - C(4')	106
H'C(5') - C(5') - C(5')	108
HO(2') - O(2') - C(2')	100
HO(3') - O(3') - C(3')	106
HO(8) - O(8) - P	102
HO(9) - O(9) - H'O(9)	111
H(10) - O(10) - H(10a)	107
H(10) - O(10) - H(10b)	121
H(10a) - O(10) - H(10b)	117
H(11) - O(11) - H'(11)	107

C(2')-endo and C(3')-endo dimensions for the arabinosyl nucleosides also. And indeed there are. Most noticeable is the very short C(3')-C(4') bond distance of 1.509 Å as found in ARACMP (C(3')-endo), compared with 1.529 Å and 1.530 Å for ARAC.HCl and ARAU (C(2')-endo). The glycosidic bond C(1')-N(1) is 0.01 Å longer than the value found for both ARAC.HCl and ARAU. The C(1')-C(2') bond length of 1.540 Å is identical to that of ARAC.HCl, but is still quite long when compared to the C(3')-endo average value reported by Saenger(57).

The inter-ring contact of O(2')...N(1) has shortened with respect to the value found for ARAU, to 2.753 Å, which is identical to the distance in ARAC.HCl. Also, as was just mentioned, the C(1')-C(2') bond length is a long 1.540 Å compared to Saenger's C(3')-endo average of 1.530 Å (57) and with ARAU's value of 1.528 Å. Again we see, with the long C(1')-C(2') bond length and short O(2')...N(1) contact, a large deviation of C(1') from the cytosine ring (Table 25), this time over 0.1 Å. The effect of this displacement can be seen in Fig. 16.

The crystal structure of ARACMP.3H_2O is shown in Fig. 17. As was mentioned in REFINEMENT:ARACMP.3H_2O, one hydrogen atom of the water molecule O(10) was not unambiguously located, and a disordered model (H(10a), H(10b)) was used to represent this hydrogen atom in the final refinement. However, for clarity this O(10)...O(10)

Table 25. Deviations from the Plane of the Pyrimidine Ring in the Structure of the ARACMP.$3H_2O$. The least-squares plane was passed through the six ring atoms, N(1) - C(6), all weighted equally. Its direction cosines relative to a, b and c, are 0.111, 0.678 and -0.727; the origin-to-plane distance is 1.546 Å.

Atom	Dev.	Atom	Dev.	Atom	Dev.
N(1)	0.017 Å	C(1')	0.146 Å	HC(5)	-0.011 Å
C(2)	-0.015	O(2)	-0.033	HC(6)	0.067
N(3)	-0.000	N(4)	0.104	O(6) (a)	0.071
C(4)	0.014	HN(3)	0.059	O(7) (a)	0.301
C(5)	-0.013	HN(4)	-0.001	O(11)(b)	-0.500
C(6)	-0.003	H'N(4)	-0.034		

(a) at $x+\frac{1}{2}$, $\frac{1}{2}-y$, $1-z$

(b) at $x-\frac{1}{2}$, $\frac{1}{2}-y$, $1-z$

Fig. 16. A view of the ARACMP cation looking down the C(1')-N(1) bond vector. A comparison of this figure with Figs. 7 and 11 will dramatically show the difference between C(2')-endo and C(3')-endo conformations.

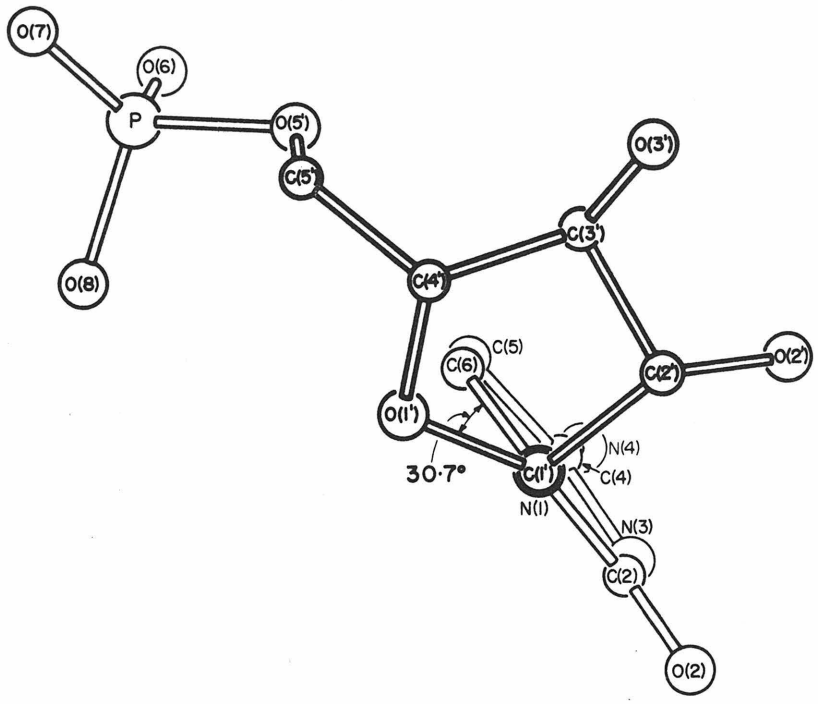
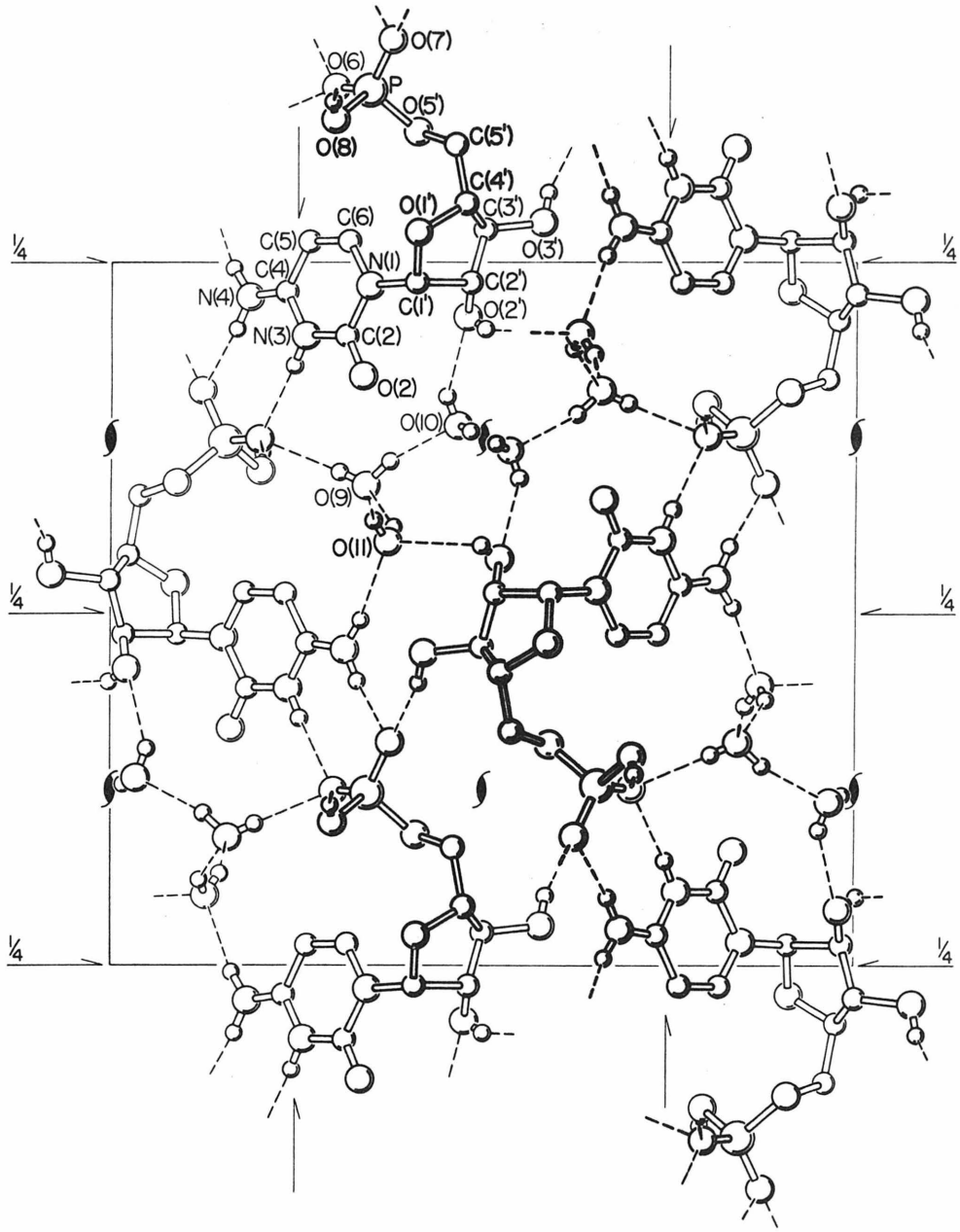


Figure 17. A drawing of the crystal structure of ARACMP.$3H_2O$. The view direction is parallel to -c; the b axis is from left to right and the a axis is from top to bottom.



hydrogen-bonded chain is depicted as being ordered in Fig. 17. The refinement of hydrogen atom parameters resulted in unreasonable parameters for the O(2')-H...O(11) hydrogen bond (see Table 26); however, we feel that the existence of such a bond is confirmed by the O...O distance of 2.84 Å and the fact that both hydrogens of O(11) refined quite nicely (Tables 16 and 21). The logical conclusion is that O(2') does not accept a hydrogen bond from O(11) and that the reverse is true. The three solvent molecules hydrogen bond together to form a stream of water flowing in the z-direction. In a manner similar to that found in the two structures of cytidylic acid (54,55), the exposed nitrogens of the cytosine ring hydrogen bond to O(6) and O(7) of a phosphate group of a symmetry-related molecule. Again, as in all structures of cytosine compounds so far reported, atoms O(2) and O(1') do not participate in true hydrogen bonds.

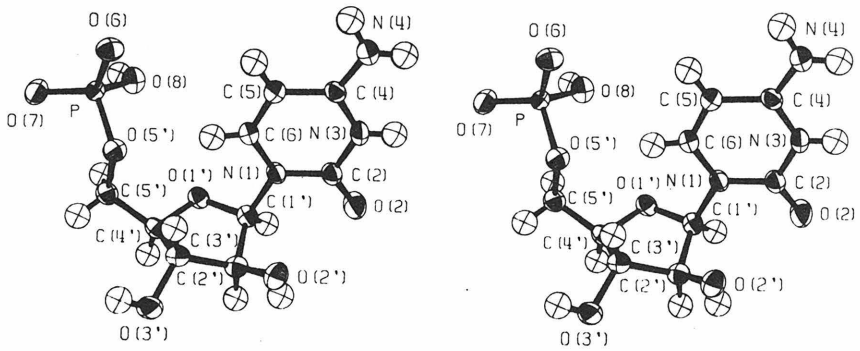
A stereoscopic representation of ARACMP and its calculated thermal ellipsoids is shown in Fig. 18. Other than the three water oxygen atoms, whose U_{ij}'s are listed in Table 15, the largest thermal motions are exhibited by O(2), O(7), and O(3'). The very large thermal parameters of O(10) perhaps suggest why the one hydrogen atom (H(10a), H(10b)) was not located.

Table 26. Distances and Angles for the Hydrogen Bonds D-H...A as found in the Structure of ARACMP.$3H_2O$.

D	H	A		Distance D-A	Distance H...A	Angle D-H...A
N(3)	HN(3)	O(6)	(e)	2.923 Å	2.08 Å	176°
N(4)	HN(4)	O(7)	(e)	2.790	1.98	154
N(4)	H'N(4)	O(11)	(a)	2.959	2.11	167
O(2')	HO(2')	O(11)	(b)	2.838	2.47	137
O(3')	HO(3')	O(7)	(c)	2.660	1.88	160
O(8)	H(8)	O(6)	(d)	2.617	2.15	144
O(9)	H(9)	O(6)	(e)	2.873	2.12	173
O(9)	H'(9)	O(10)		2.729	1.87	162
O(10)	H(10)	O(2')		2.763	2.07	147
O(10)	H(10a)	O(10)	(f)	2.785	1.83	175
O(10)	H(10b)	O(10)	(g)	2.785	1.86	175
O(11)	H(11)	O(9)	(d)	2.803	2.07	150
O(11)	H'(11)	O(9)		2.878	2.19	156

(a) at $x-\frac{1}{2}$, $\frac{1}{2}-y$, $1-z$
(b) at $\frac{1}{2}-x$, $1-y$, $z-\frac{1}{2}$
(c) at $-\frac{1}{2}-x$, $1-y$, $z-\frac{1}{2}$
(d) at x, y, $z+1$
(e) at $x+\frac{1}{2}$, $\frac{1}{2}-y$, $1-z$
(f) at $\frac{1}{2}-x$, $1-y$, $z+\frac{1}{2}$
(g) at $\frac{1}{2}-x$, $1-y$, $z+\frac{1}{2}$

Figure 18. A stereoscopic illustration of the ARACMP cation(61). Thermal ellipsoids are drawn at the 50% probability level; hydrogen atoms are drawn as spheres of arbitrary radius.



DISCUSSION

To date there have been structural reports of eight different arabinosyl nucleotides or nucleosides, seven of which contain a pyrimidine as the base(22-27; this thesis). Three of these seven have been described here (see Fig. 19). From the structural information presented by these studies, several conclusions can be drawn and conjectures raised.

The crystallographic evidence suggests quite strongly that at least for the arabinosylpryrimidines the glycosidic torsion angle ϕ_{cn} is severely limited in range. For the seven such structures studied, this parameter has an average value of 31 ± 3^{o} (see Table 27). Haschemeyer and Rich (64) have calculated allowable ranges of this parameter for ribosylpyrimidines, using short intramolecular contacts to define the limits, and found the average range to be much greater. If the sugar folding is C(3')-endo, they calculate a range of 25-105°; for C(2')-endo, the range is 25-80°. Crystal structure determinations of such compounds support the prediction of Haschemeyer and Rich, in that all pyrimidine nucleotides and nucleosides so far investigated, save three, exhibit glycosidic torsion angles which fall into the anti range ($\phi_{cn} = 30\pm90^{o}$; ref. 59). (The three exceptions are 4-thiouridine(65), 6-methyluridine(66), and cytidine-2':3'-cyclic monophosphate(67)).

Despite the apparent constancy of the glycosidic torsion angle in the arabinosylpyrimidines, the sugar

conformation seems to be quite flexible, perhaps indicating a low barrier to interconversion. Of the four possible conformers (C(2')-endo, gauche-gauche or gauche-trans; and C(3')-endo, gauche-gauche or gauche-trans),* three are represented by the structures presented in this thesis. the one conformation that is not shown here is elsewhere (26), by ARAA, an arabinosylpurine.

Although there is little evidence to favor a particular sugar ring puckering for the arabinosylpyrimidines, there is some indication that a gauche-gauche conformation is preferred regardless of the ring folding. A gauche-gauche and C(2')-endo conformation results in an intramolecular hydrogen bond, as seen in the structure of ARAU presented here; this is the same conformation observed in the structure of ARAC (23,24). And a gauche-gauche, C(3')-endo conformation may be stabilized by an energy-minimizing C(6)-H...O(5') interaction (60,68,69). This second conformation is observed here in the structure of ARACMP and also in the structure of arabinosyl-4-thiouracil (27), as well as in numerous ribosyl nucleotide structures.

There also appears to be a correlation between the N(1)...O(2') ring-ring interaction distance, the C(1')-C(2')

---

*No distinction is made here between gauche-trans and trans-gauche; neither seems to be affected by any short intramolecular contacts. Also, the exo conformations appear to be quite unpopular and were ignored.

bond length, and the displacement of atom C(1') from the plane of the pyrimidine ring. Using the results of the three structures reported here plus dimensions reported for arabinofuranosyl-4-thiouridine (ARA-S^4U; ref. 27), the following hypothesis can be suggested: as the ring-ring interaction becomes more intense (defined by a shortening of the N(1)...O(2') distance), the C(1')-C(2') bond distance increases and the displacement of atom C(1') from the plane of the pyrimidine ring also increases. For the four structures under discussion, the N(1)...O(2') distance is either 2.75 Å or 2.80 Å, the upper value being that observed for ARAU; and ARAU is the only one of the four molecules not to exhibit a significant displacement of atom C(1') from the plane of the pyrimidine ring. Variation of the C(1')-C(2') bond distance is much less and only borders on significance. In addition to the correlations discussed above, the length of the C(3')-C(4') bond distance appears to be related to the sugar ring puckering. The spread of values observed for this bond distance is quite large - the shorter lengths being associated with the C(3')-endo conformation and the longer with the C(2')-endo conformation.

In the past six months there have been published crystallographic studies of dinucleotides which mimic the polynucleotide conformations suspected of being used in nucleic acid double helices (70,71). These structures also verify the trend of nucleotides of fitting into one or two

overall conformations in the solid state: C(2')-endo, gauche-gauche, anti; and C(3')-endo, gauche-gauche, anti. Arabinosyl nucleotides can also conform to such descriptions, as can be seen from the results presented here and elsewhere (24-29).

A close look at helix models of DNA and RNA (68) shows that by placing an arabinosyl sugar into the helix in place of a ribose or deoxyribose, serious stereochemical interferences result between O(2') of that sugar and the 3'-nucleotide neighboring the arabinosyl residue. It is suggested here, more for the purposes of discussion than as fact, that metabolically inhibitive arabinofuranosyl nucleotides are so, not because of their inability to conform to helix torsions, since they can, but rather, because of the adverse interactions of the O(2') oxygen atom with the synthesizing polynucleotide. The correlations between ring-ring interactions, bond lengths, and out-of-plane displacements discussed above, although significant, are probably of minor importance relative to the configurational difference at C(2') of the arabinofuranose and the (deoxy)ribofuranose.

Table 27. Conformational Summary of Reported Arabinonucleotides and Nucleosides.

Structure	ϕ_{cn}	Sugar Folding	C(4') - C(5') Conformation
ARAC.HCl	26.7°	C(2') endo	g-t
ARAU	35.0	C(2') endo	g-g
ARACMP.$3H_2O$	30.7	C(3') endo	g-g
ARA-S^4U (27)	36.0	C(3') endo	g-g
ARA-Br^5U (22)	30.	C(1')exo-O(1')endo	g-g
ARAT (25)	30.	C(1')exo-O(1')endo	g-g
ARAC (23,24)	28.8	C(2') endo	g-g
ARAA (26)	57.8	C(3') endo	g-t

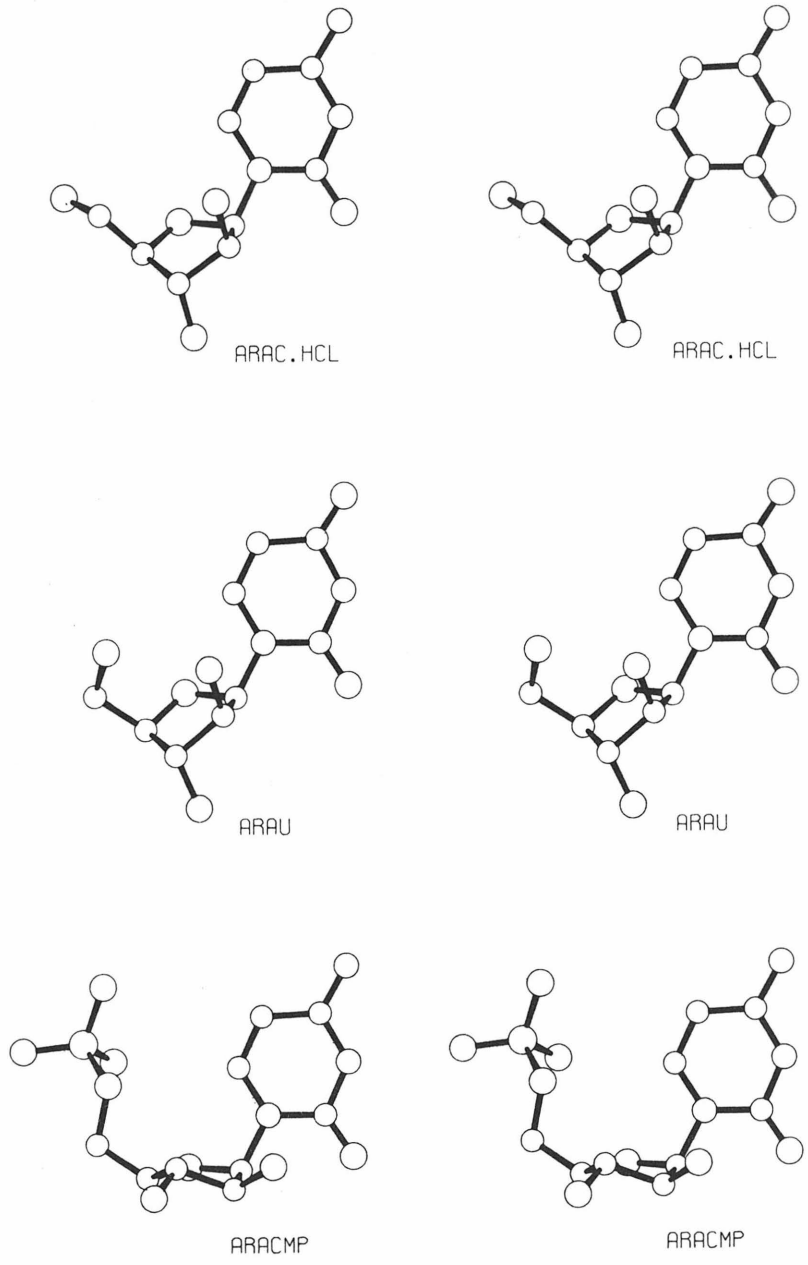
*Other abbreviations used in this table are ARA-S^4U: arabinofuranosyl-4-thiouracil; ARA-Br^5U: arabinofuranosyl-5-bromouracil; ARAT: arabinofuranosylthymine; ARAA: arabinofuranosyladenine. ARA-Br^5U and ARAT crystallize isostructurally.

Table 28. Torsion Angles in ARAC.HCl, ARACMP.$3H_2O$ and ARAU.

A positive angle corresponds to a right-handed screw (72). Estimated standard deviations are about $\frac{1}{2}^{o}$.

Torsion System	ARAC.HCl C(2')endo	ARAU C(2')endo	ARACMP.$3H_2O$ C(3')endo
C(6) - N(1) - C(1') - O(1')	26.7°	35.0°	30.7°
O(1') - C(1') - C(2') - C(3')	25.2	38.2	-26.0
C(1') - C(2') - C(3') - C(4')	-34.2	-34.7	37.0
C(2') - C(3') - C(4') - O(1')	32.8	20.7	-35.7
C(3') - C(4') - O(1') - C(1')	-4.6	3.4	20.2
C(4') - O(1') - C(1') - C(2')	-17.9	-26.5	4.0
O(5') - C(5') - C(4') - C(3')	-171.5	55.8	53.3
O(5') - C(5') - C(4') - O(1')	68.7	-63.6	-66.1
O(8) - P - O(5') - C(5')			-64.2
P - O(5') - C(5') - C(4')			137.2

Fig. 19. A stereographic view summarizing the three arabinofuranosyl conformations presented. The view in all three cases is normal to the plane of the pyrimidine ring.



ADDENDUM

As this work was being written, it was discovered that the structures of ARACMP and ARAU had also been solved by other research groups. The structure of ARACMP will be reported at the 1973 ACS Meeting in Chicago (M. Sundaralingam, 1973; private communication). The structure of ARAU has been reported in preliminary form (P. Tollin, H.R. Wilson, and D.W. Young., Nature New Biology, 242, 49(1973)). In addition, the structural description of arabinofuranosyladenine hydrochloride is also being given at the 1973 Chicago ACS meeting (M. Sundaralingam, 1973; private communication). No significant differences can be discerned between the results available for the ARAU and ARACMP structures mentioned above and those reported in this thesis. The conformation of ARAA.HCl is C(3')-endo, gauche-gauche, with a glycosidic torsion angle of 29.7°.

So it goes.

Footnotes

(1) W. Bergmann and R.J. Feeney, J. Amer. Chem. Soc., 72, 2809(1950).

(2) W. Bergmann and R.J. Feeney, J. Org. Chem., 16, 981(1951).

(3) W. Bergmann, J. Marine Res.(Sears Found.), 8, 137(1949).

(4) W. Bergmann, J.C. Watkins, and M.F. Stempien, J. Org. Chem., 22, 1308(1957).

(5) S.S. Cohen, "Progress in Nucleic Acid Research and Molecular Biology," Vol. 5, J.N. Davidson and W.E. Cohn, Ed., Academic Press, New York, N.Y., 1966, p 1.

(6) M.Y. Chu and G.A. Fisher, Biochem. Pharmacol., 11, 423(1962).

(7) E.C. Moore and S.S. Cohen, J. Biol. Chem., 242, 2116(1967).

(8) P. J. Ortiz, M.J. Manduka, and S.S. Cohen, Cancer Res., 32,1512(1972).

(9) D.B. Smith and E.H.Y. Chu, ibid.,32, 1651(1972).

(10) P.J. Burke, A.H. Owens, J. Colsky, B.I. Shnider, J.H. Edmonson, A. Schilling, H.S. Brodovsky, H.J. Wallace, and T.C. Hall, ibid., 30, 1512(1970).

(11) J.S. Evans, E.A. Musser, G.D. Mengel, K.R. Forsblad, and J.H. Hunter, Proc. Soc. Exp. Biol. Med., 106, 350(1961).

(12) E.D. Whittle, Biochim. Biophys. Acta, 14, 44(1966).

(13) W.A. Creasey, R.J. Papac, M.E. Markin, P. Calabresi, and A.D. Welch, Biochem. Pharmacol., 15, 1417(1966).

(14) M.Y. Chu and G.A. Fisher, ibid., 14, 333(1965).

(15) D. Kessel, T.C. Hall, and I. Wodinsky, Science, 156, 1240(1967).

(16) J.P. Durham and D.H. Ives, Fed. Proc., Fed. Amer. Soc. Exp. Biol., 27, 538(1968).

(17) S. Vadlamudi and A. Golden, Cancer Chemother. Rep., 55, 547(1972).

(18) H.E. Skipper, F.M. Schabel, and W.S. Wilcox, ibid., 51, 125(1967).

(19) R.L. Momparler, Biochem. Biophys. Res. Commun., 34, 465(1969).

(20) M. Karon and W.F. Benedict, Science, 178, 62(1972).

(21) M. Karon, W.F. Benedict, and N. Rucker, Cancer Res., 32, 2612(1972).

(22) P. Tougard, Biochem. Biophys. Res. Commun., 37, 961(1969).

(23) O.L.- Soubeyran and P. Tougard, C.R. Acad. Sci., Ser. C, 276, 403(1973).

(24) A.K. Chwang and M. Sundaralingam, Nature New Biology, 243, 78(1973).

(25) P. Tougard, C.R. Acad. Sci., Ser. C, 273, 878(1971).

(26) G.J. Bunick and D. Voet, Amer. Crystallogr. Ass. Meetings (Abstract) (University of Florida, Gainesville, January 14-18, 1973).

(27) W. Saenger, J. Amer. Chem. Soc., 94,621(1972).

(28) A.J.C. Wilson, Nature(London), 150, 151(1942).

(29) G.H. Stout and L.H. Jensen, "X-Ray Structure Determination," Macmillan, New York, N.Y., 1968, p 456.

(30) $R = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$; $R(\text{weighted}) = \Sigma w(F_o^2 - F_c^2)^2 / \Sigma w F_o^4$; goodness-of-fit = $(\Sigma w(F_o^2 - F_c^2)^2/(M-S))^{1/2}$ where M = number of non-zero weight data and S = number of parameters refined.

(31) J. Karle and H. Hauptman, Acta Crystallogr., 9, 635(1956).

(32) J.W. Moncrief and W.N. Lipscomb, ibid., 21, 322(1966).

(33) A.F. Peerdeman and J.F. Bijvoet, ibid., 9, 1012(1956).

(34) G.N. Ramachandran and S. Raman, Curr. Sci., 25, 348(1956).

(35) A.C. Hazell, Nature(London), 227, 269(1970).

(36) J.A. Ibers, Acta Crystallogr., 22, 604(1967).

(37) "International Tables for X-Ray Crystallography," Vol. III, Kynoch Press, Birmingham, 1962, p 202-214.

(38) C.E. Bugg, J.M. Thomas, M. Sundaralingam, and S.T. Rao, Biopolymers, 10, 175(1971).

(39) C.A. Beevers and H. Lipson, Phil. Mag., 17, 855(1934).

(40) C.A. Beevers and H. Lipson, Nature(London), 137, 825(1936).

(41) H. Lipson and C.A. Beevers, Proc. Phys. Soc.(London), 48, 772(1936).

(42) H. Hauptman and J. Karle, Acta Crystallogr., 9, 45(1956).

(43) W.H. Zachariasen, ibid., 5,68(1952).

(44) J. Karle and I.L. Karle, ibid., 21,849(1966).

(45) D.T. Cromer and J.T. Waber, ibid., 18, 104(1965).

(46) R.F. Stewart, E.R. Davidson, and W.T. Simpson, J.Chem. Phys., 42, 3175(1965).

(47) W.H. Zachariasen, Acta Crystallogr., 16, 1139(1963).

(48) A.C. Larson, ibid., 23, 664(1967).

(49) H. Hope, private communication, (1969). For O, $\Delta f' = 0.049$, $\Delta f'' = 0.032$; for N, $\Delta f' = 0.032$, $\Delta f'' = 0.019$.

(50) D.W.J. Cruickshank and W.S. McDonald, Acta Crystallogr., 23, 9(1967).

(51) D.T. Cromer, ibid., 18, 17(1965).

(52) W.C. Hamilton, ibid., 18, 502(1965).

(53) E. Subramanian and D.J. Hunt, ibid., Sect. B, 26, 303(1970).

(54) C.E. Bugg and R.E. Marsh, J. Mol. Biol., 25, 67(1967).

(55) M. Sundaralingam and L.H. Jensen, ibid., 13, 914(1965).

(56) M. Sundaralingam, Biopolymers, 7, 821(1969).

(57) W. Saenger and Eckstein, J. Amer. Chem. Soc., 92, 4712(1970).

(58) D. Voet and A. Rich, "Progress in Nucleic Acid Research and Molecular Biology," Vol. 10, J.N. Davidson and W.E. Cohn, Ed., Academic Press, New York, N.Y., 1970, p 183.

(59) J. Donohue and K.N. Trueblood, J. Mol. Biol., 2, 363(1960).

(60) E. Shefter and K.N. Trueblood, Acta Crystallogr., 18, 1067(1965).

(61) C.K. Johnson, "OR TEP, a Fortran Thermal Ellipsoid Plot Program for Crystal Structure Illustrations," U.S. Atomic Energy Commission Report ORNL-3794, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1965.

(62) M. Sundaralingam and L.H. Jensen, J. Mol. Biol., 13, 930(1965).

(63) B.L. Trus and R.E. Marsh, Acta Crystallogr., Sect. B, 28,1834(1972).

(64) A.E.V. Haschemeyer and A. Rich, J. Mol. Biol., 27, 369(1967).

(65) W. Saenger and K.H. Scheit, ibid., 50, 153(1970).

(66) D. Suck, W. Saenger, and H. Vorbruggen, Nature (London), 235,333(1972).

(67) C.L. Coulter and M.L. Greaves, Science, 169, 1097(1970).

(68) S.-H. Kim, H.M. Berman, N.C. Seeman, and M.D. Newton, Acta Crystallogr., Sect. B, 29, 703(1973).

(69) P. T'so, N. Kondo, M. Schweizer, and D. Halles, Biochemistry, 8, 997(1969).

(70) J.M. Rosenberg, N.C. Seeman, J.J.P. Kim, F.L. Suddath, H.B. Nicholas, and A. Rich, Nature(London), 243 150(1973).

(71) R.O. Day, N.C. Seeman, J.M. Rosenberg, and A. Rich, Proc. Nat. Acad. Sci. USA, 70, 849(1973).

(72) W. Klyne and V. Prelog, Experientia, 16, 521(1960).

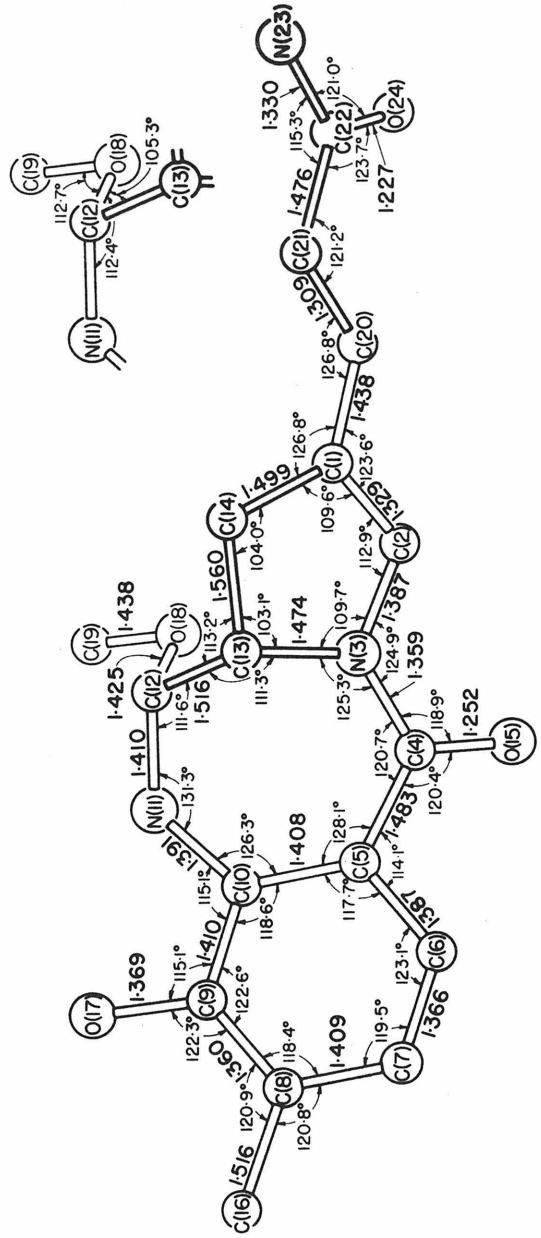
APPENDIX

In addition to the three crystal structures described in the thesis text, a number of others have also been investigated, either by me, or in conjunction with someone else. Short descriptions of the projects will be given in this appendix.

I. Anthramycin methyl ether and anthramycin have been shown to be effective in inhibiting the synthesis of nucleic acids(1). The methyl ether derivative hydrolyzes to yield the parent compound in aqueous solution(2). In conjunction with Prof. Larry Dahl of the University of Wisconsin and Dr. Robert M. Sweet of UCLA, we have determined the crystal structure of anthramycin methyl ether monohydrate.

The crystals of anthramycin methyl ether monohydrate, $C_{17}N_3O_4H_{19}.H_2O$, are orthorhombic, space group $P2_12_12_1$, with cell dimensions $\underline{a}$ = 13.505, $\underline{b}$ = 16.224, $\underline{c}$ = 7.975 Å. There are four formula units per unit cell. Solution was by the combined symbolic addition-tangent formula method (3,4), and the structure was refined by full-matrix least squares to a final R index of 0.086 for 1402 data. The crystal structure is held together by five independent hydrogen bonds, three of which include the water molecule as either donor or acceptor. The anthramycin molecule is in a strained conformation due to the fusion of the non-planar seven-membered ring with the planar benzene ring (Fig. 1).

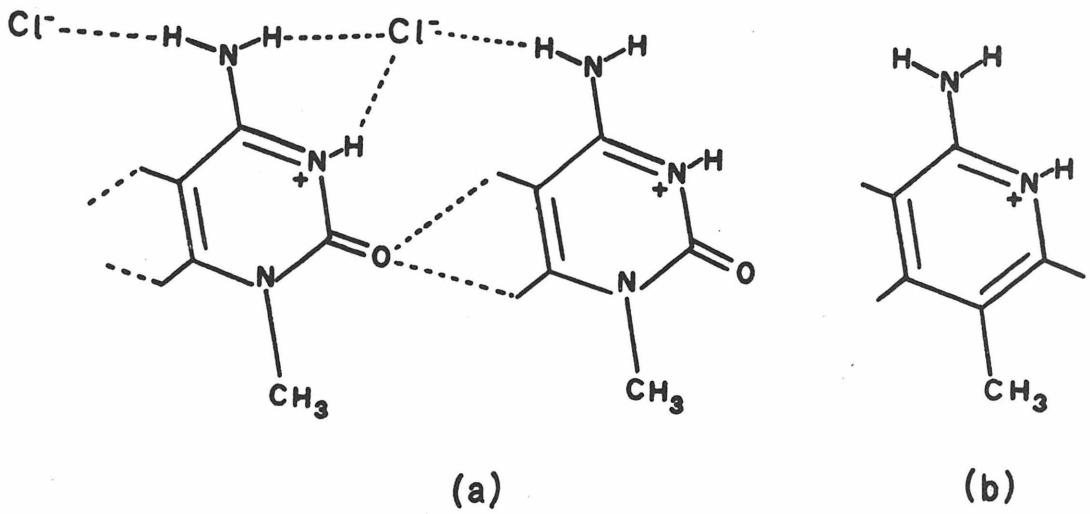
Fig. 1. A drawing of anthramycin methyl ether showing the heavy-atom distances and angles. Estimated standard deviations are 0.007-0.009 Å and 0.5°.



II. A feature common to the crystal structures of all hydrochloride salts of cytosine derivatives that have so been reported is the translational relationship of cytosine rings and chloride ions shown in Fig. 2a(5). While this arrangement must be primarily determined by the N-H...Cl- hydrogen bonds, the C-H...O interactions may also be of some importance(6,7). Characteristic values for the C...O and H...O distances are 3.2 Å and 2.6 Å; while this latter value is too large to suggest a true hydrogen bond, appreciable electrostatic attraction may contribute to the stability of the arrangement. To help in determining the importance of these interactions, we have investigated the crystal structure of 2-amino-5-methylpyridine hydrochloride (Fig. 2b). This compound can be considered a "deoxy" derivative of 1-methylcytosine hydrochloride, and our intention was to compare the packing arrangements of the two crystal structures and determine the effect of removing the opportunity for C-H...O interactions.

Approximately ½ gram of practical grade 2-amino-5-methylpyridine was recrystallized from toluene, then from dilute hydrochloric acid. Clear, needle-shaped crystals of the hydrochloride salt (2A5MP.HCl) were isolated and proved to be monoclinic, space group $P2_1/c$. Cell dimensions are $\underline{a}$ = 7.144(8), $\underline{b}$ = 12.248(4), $\underline{c}$ = 8.845(4) Å, β = 107.61°. 2A5MP was also recrystallized from hydrobromic acid, the resulting crystals being isostructural with those of the

Fig. 2. (a) The translational relationship between cytosine rings and chloride ions as found in a number of hydrochloride salts of cytosine derivatives; (b) the 2-amino-5-methyl-pyridinium cation.



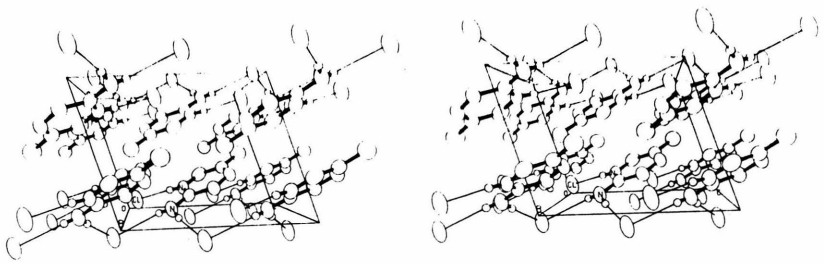
(a) (b)

hydrochloride salt. Cell dimensions for 2A5MP.HBr are $\underline{a}$ = 7.070(3), $\underline{b}$ = 12.684(5), $\underline{c}$ = 9.173(3) Å, β = 109.41(3)°.

Diffraction data for 2A5MP.HCl were collected using a locally assembled, Syntex-automated, E&A full-circle diffractometer(8). Molybdenum radiation monochromatized by a graphite crystal (λ(MoKα) = 0.7107 Å) was used, and reflections with $\sin^2\theta/\lambda^2$ less than 0.32 were measured by the θ-2θ scan technique. Solution of the structure was by traditional Patterson-Fourier methods, and full-matrix least-squares refinement of atomic parameters converged at a final R index of 0.074 for 1277 data. No further work was done with the 2A5MP.HBr crystals.

A stereoscopic drawing of the crystal structure of 2-amino-5-methylpyridine hydrochloride is shown in Fig. 3. Three N-H...Cl- hydrogen bonds are the primary structure-determining forces, with some molecular overlap also contributing. The two independent interplanar distances are about 3.37 and 3.44 Å. Of particular note is the lack of similarity between the packing in this case and the structure of 1-methylcytosine hydrochloride(9). A reasonable conclusion is that when the opportunity for C-H...O interactions is removed from the structure of 1-methylcytosine hydrochloride (by replacing that cation by the 2-amino-5-methylpyridinium cation), the structure must change to accommodate for the missing interactions.

Fig. 3. A stereoview of the 2-amino-5-methylpyridine hydrochloride structure looking down the b axis



III. As a part of a program of study concerning the non-bonding interactions of peri-substituted naphthalene structures, we have in collaboration with Prof. J.D. Robert's group undertaken the investigation of the crystal structures of several such compounds(10,11). Dr. Robert Wheland initially began a project of determining the crystal structure of 1,8-diiodonaphthalene. However, no amount of trying could yield a satisfactory trial solution that would fit the required six diiodonaphthalene molecules into the measured unit cell. To convince ourselves that the experimental data were or were not at fault, I reinvestigated the crystalline sample and grew new crystals from several different solvents. No crystals could be found which exhibited the same diffraction pattern as the crystal used for data collection. There was enough of a similarity in the patterns to suggest a possible twinning in the first case, however.

Using graphite-monochromatized molybdenum radiation and a Datex-automated G.E. XRD-5 quarter circle diffractometer, new intensity data were collected. A triclinic cell of dimensions a = 12.711(3), b = 8.049(2), c = 7.684(2) Å, α = 89.24(2)°, β = 93.49(3)°, γ = 107.70(2)° and a measured density of 2.55 g.cm^{-3} indicated 3 molecules in the unit cell. The volume of the new unit cell was one-half that of the original cell. Since 1,8-diiodo-naphthalene cannot possess a molecular center of symmetry,

we proceeded under the assumption that there were three molecules in triclinic space group P1. From a three-dimensional Patterson map, trial coordinates for six iodine atoms were determined. It soon became apparent, however, that the structure was not that of 1,8-diiodonaphthalene, but rather a 2:1 complex of 1,8-diiodonaphthalene and 1,5-diiodonaphthalene in space group $P\bar{1}$. The 1,5-substituted molecule lies on a center of symmetry, while the 1,8-substituted naphthalenes stack across centers of symmetry. In addition, the 1,5-diiodonaphthalene molecule is 25% disordered by a 180° rotation about the I-I vector (Fig. 4). Full-matrix least-squares refinement converged at an R index of 0.049 based on 3193 reflections. The 1,8-diiodonaphthalene molecule, due to the strong peri-interactions of the iodine atoms, shows significant skeletal distorsions as well as in- and out-of-plane displacements of the iodine atoms. The crystal structure is shown in a stereoview in Fig. 5.

Fig. 4 A stereoview of the 1,5-diiodonaphthalene molecule showing both orientations exhibited by the disorder.

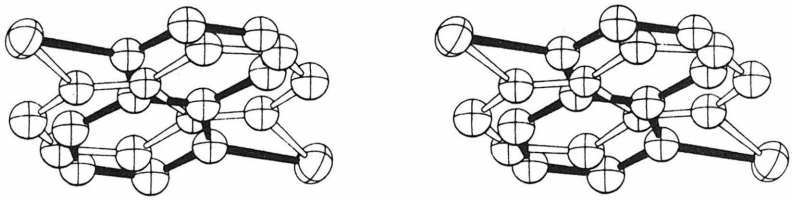
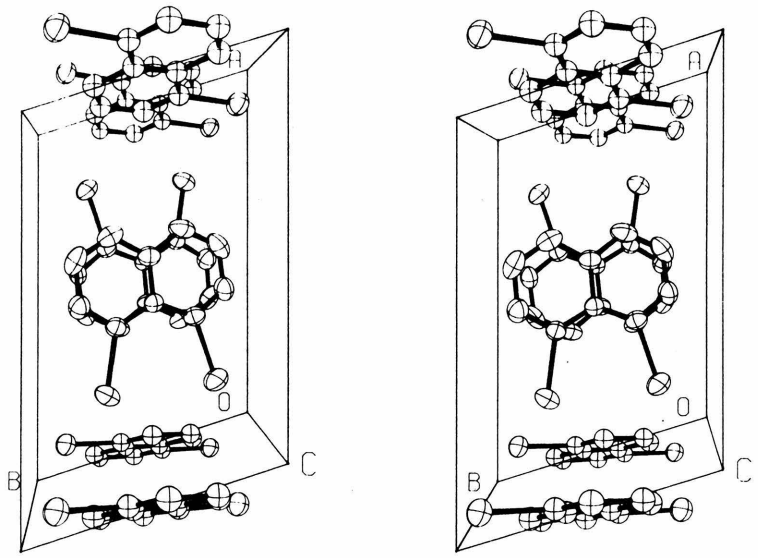


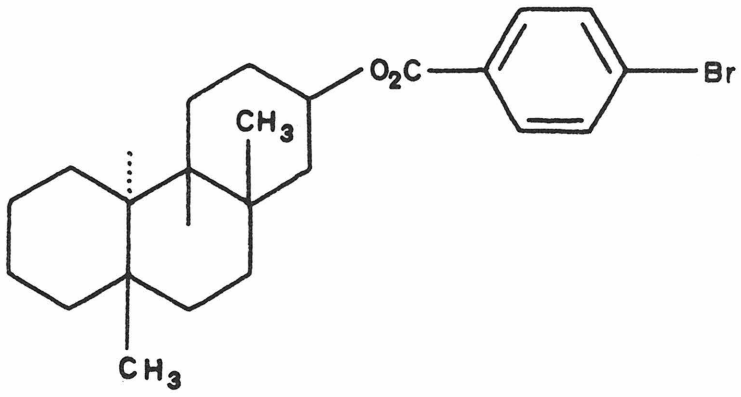
Fig. 5. A stereoview of the structure of the 2:1 complex of 1,8-diiodonaphthalene and 1,5-diiodonaphthalene. Only the predominant orientation of the 1,5-substituted naphthalene is shown here. The view of the disordered molecule in Fig. 4 is looking in nearly the same direction.



IV. In conjunction with the synthetic organic chemists at Caltech, and when time and funds permit, structural investigations of synthetic intermediates are often undertaken to confirm the suspected configurations of the compounds. An undergraduate, David Dummit, under my supervision, began one such project as a Chem 3 project but had to discontinue his work at the beginning of the summer (1972), having determined space group and obtained preliminary cell parameters. As the structure was still of some importance to Ron Trust, the graduate student who originally supplied the compound, I continued the work.

The proposed structure of the compound is shown in Fig. 6. Diffractometer data were collected using CuKα radiation. Solution was by Patterson-Fourier techniques, and isotropic refinement of the structure was halted at an R index of 0.14, even though it was clear that a much better fit to the data was possible. There was no obvious gain in continuing any further; since a difference map showed no maxima greater than about 0.7 eA^{-3}, we were satisfied that the structure was correct, and that the compound in question did indeed have the configuration shown in Fig. 6.

Fig. 6



References for Appendix

(1) S.B. Horwitz, "Progress in Molecular and Subcellular Biology," Vol. 2, F. Hahn, Ed., Springer-Verlag, Berlin, 1971.

(2) W. Leimgruber, V. Stefanovic, F. Schenker, A. Karr, and J. Berger, J. Amer. Chem. Soc., 87, 5791(1965).

(3) J. Karle and I.L. Karle, Acta Crystallogr., 21, 849(1966).

(4) J. Karle and H.H. Hauptman, ibid., 9, 635(1956).

(5) J.S. Sherfinski and R.E. Marsh, ibid., Sect. B, 29,192(1973).

(6) J. Donohue, "Structural Chemistry and Molecular Biology," N. Davidson and A. Rich, Ed., W.H. Freeman and Co., San Francisco, 1968, p 433.

(7) W.C. Hamilton and J.A. Ibers, "Hydrogen Bonding in Solids," W.A. Benjamin, Inc., New York, 1968, p 14-16.

(8) S. Samson and D.A. Hansen, Acta Crystallogr., Sect. B, 28,930(1972).

(9) B.L. Trus and R.E. Marsh, ibid., Sect. B, 28, 1834(1972).

(10) J.-B. Robert, J.S. Sherfinski, R.E. Marsh and J.D. Roberts, in preparation.

(11) H.M. Einspahr, J.-B. Robert, R.E. Marsh and J.D. Roberts, Acta Crystallogr., Sect. B, 29, 000(1973).

Propositions

I. I propose to do the crystal structure of 5-methylisocytosine hydrochloride as the third of a series of structures (others were 1-methylcytosine hydrochloride and 2-amino-5-methylpyridine hydrochloride) investigating the existence of C-H...O hydrogen bonds in the solid state.

II. I propose to carry out investigations as to the nature of the interaction of the antibiotic anthramycin with helical DNA. Specifically, I intend to determine if anthramycin is capable of forming a phosphate ester with DNA.

III. I propose to carry out a complete three-dimensional crystallographic analysis of a nucleic acid-protein complex containing aminoacylated tRNA, GTP, and the elongation factor T_u. It is as this complex that the aminoacylated tRNA is presented to a polypeptide-synthesizing ribosome.

IV. I propose to do a complete clinical and biochemical evaluation of a newly-synthesized group of nucleotide analogs to determine their chemotherapeutic and antibiotic values.

V. I propose to investigate the source of the very high number of low-energy pulses observed in diffractometer counting circuitry, and in addition propose a polynomial correction factor for coincidence losses.

Proposition I.

Hydrogen bonds involving nitrogen, oxygen, fluorine and chlorine as acceptor and donor atoms have been well established as primary contributors to the stability of many crystalline substances and solution systems. The high electronegativities of these atoms are believed to cause a slight residual negative charge density on the atom while inductively causing hydrogen atoms covalently bonded to them to take on a slight positive charge. To a first approximation then, the hydrogen bond can be considered an electrostatic interaction; it may be possible for there to be some covalent character in the stronger hydrogen bonds, but this contribution must be small due to the unavailability of suitable hydrogen atomic orbitals.

Accepted as the hydrogen bond is, there still remains the problem of defining when a hydrogen bond does exist. Hamilton and Ibers (1) give as a criterion for the existence of a hydrogen bond that both the donor-acceptor and hydrogen-acceptor distances be less than the corresponding sum of the van der Waals radii. Several problems arise with this definition. One is that it is only applicable to diffraction studies or experiments which can yield interatomic distances. Also, in the case of X-ray studies, the derived hydrogen-atom parameters are quite unreliable due to the poor approximation used for the charge density

of hydrogen and thus its scattering power. Another is the basic uncertainty in values of the published van der Waals radii. Pauling (2) notes with respect to his values that the numbers may be considerably in error, depending on the directionality of contact and the electronegativity of (covalent) nearest neighbors. In defining the limits of what actually constitutes a hydrogen bond, it might be worthwhile to investigate more extensively systems that by existing standards contain "marginal" hydrogen bonds.

One method for studying the weaker interactions is to look at systems which contain the questionable C-H...O hydrogen bond. Carbon, having a considerably lower electronegativity than fluorine, oxygen, nitrogen or chloride must be a much poorer donor atom than the others mentioned if any of the theories concerning hydrogen bonds are correct. Although there have been numerous discussions concerning C-H...O hydrogen bonds (see, for example refs. 1 and 3), in most cases the interactions under discussion seem to be the result of other stronger packing forces and are not primary contributors to the determination of the crystal structure.

Several factors cannot allow the rejection of a C-H...O attractive force, however. The H...O distances in a number of cases are 0.4 - 0.5 Å shorter than the accepted sum of van der Waals radii (2.6 Å; ref. 2). Also, the carbon atom involved is generally in some type of activating

environment, allowing the attached hydrogen atom to assume a net partial positive charge. Caffeine (4), which exhibits a C-H...O H...O contact distance of 2.07 Å, utilizes a ring carbon flanked by two nitrogen atoms. Uracil (5), having two short C-H...O contacts in the crystalline state (H...O distances of 2.18 Å and 2.23 Å), utilizes two adjacent carbon atoms as donors, one next to a nitrogen atom and the other next to a carbonyl carbon atom. There are also several examples of strong C-H...O interactions involving the carbon atoms of a protonated imidazole ring (6,7). In all the above cases it does seem possible that although the C-H...O interactions are not determining forces in crystal packing, they could very well be lowering the total energy of the crystal structure.

This entire question concerning the nature of the C-H...O interaction is not purely academic. Recent discussions on the different nucleotide conformations in DNA and RNA indicate that a C-H...O interaction may be of some importance in maintaining the favorable C(3')-endo, gauche-gauche, anti conformation suspected of being the predominant nucleotide conformer in nucleic acid polymers(8). There is evidence for this C-H...O contribution to stability not only from solid state investigations, but also from spectroscopic studies of 5'-nucleotides in solution (9).

A system of crystallographic structures has been described which appear to contain marginal C-H...O

interactions (10). The compounds of interest are the hydrochloride salts of cytosine derivatives, and all crystal structures of these salts have so far exhibited the same relationship between cytosine rings and chloride ions shown in Fig. 1. Clearly the major contributor to the stability of this arrangement is the network of N-H...Cl- hydrogen bonds. But secondarily there are the close C-H...O contacts involving C(6)-H and C(5)-H of a cytosine ring and O(2) of a translationally equivalent cytosine ring. In addition there is the possible interaction between C(5)-H and the chloride ion, indicated by the letter k. It was not clear from the evidence available how strongly attractive these C-H...O contacts were.

Because of the uncertainty as to the nature of the C-H...O contacts and possible hydrogen bonds involving a C-H group as a donor, it was suggested that the crystal structure of the hydrochloride salt of 2-amino-5-methyl-pyridine by investigated by X-ray diffraction methods (11). This compound is to a first approximation the "deoxy" form of 1-methylcytosine hydrochloride which is contained in the list of compounds generically referred to above as "hydrochloride salts of cytosine derivatives."

The crystal structure of 2-amino-5-methylpyridine hydrochloride (2A5MP.HCl) has been determined (12) and is shown in a stereoview in Fig. 2. Of particular note is the lack of similarity between the packing in this case and

Fig. 1. The translational arrangement of ions as exhibited by hydrochloride salts of cytosine derivatives.

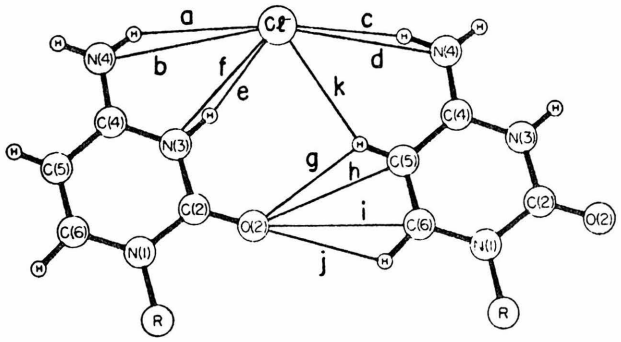
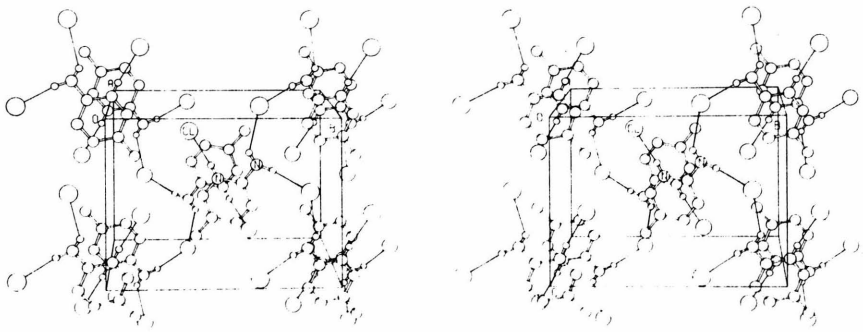


Fig. 2. Crystal packing as found in the structure of 2-amino-5-methylpyridine hydrochloride (2A5MP.HCl).

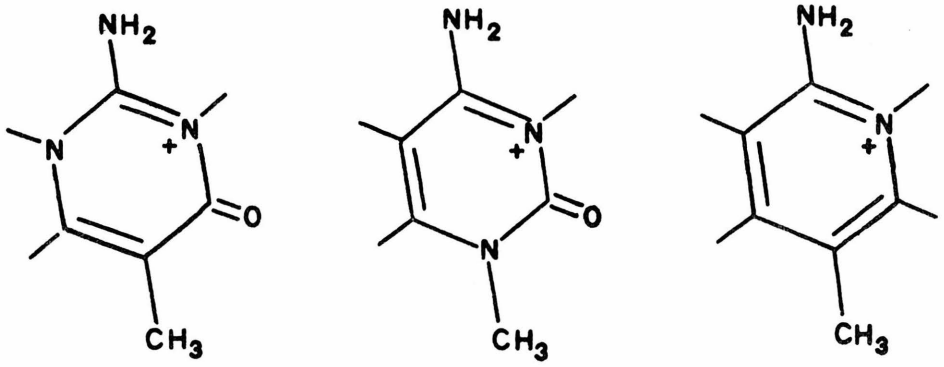


in the structure of 1-methylcytosine hydrochloride (13; see Fig. 1). One can say then that when the opportunity for C-H...O interactions in the structure of 1-methylcytosine hydrochloride is removed (by replacing that cation by the 2-amino-5-methylpyridinium cation), the molecules must rearrange to a more favorable crystal structure.

I now propose to do the crystal structure of 5-methylisocytosine hydrochloride to further investigate the nature of the packing forces in this system of structures. A drawing of the 5-methylisocytosine cation is shown in Fig. 3 along with the similar 1-methylcytosine and 2A5MP cations. A simplistic way of describing the isocytosine molecule is to say that it is 1-methylcytosine with the C(5)-H grouping replaced by an N(5)-H group.

Placing an N-H group in that position now makes possible N-H...O and N-H...Cl- hydrogen bonds. For the structure of 2A5MP.HCl it was hinted that the C-H...O interactions and the C-H...Cl- interaction (g, j, k in Fig. 1) contribute to the crystal packing forces of the hydrochloride salts of cytosine derivatives. If the crystal structure of 5-methylisocytosine hydrochloride (5MIC.HCl) is isostructural with that of 1-methylcytosine hydrochloride, and there seems to be a good possibility that it may, it would be too great a temptation to ignore the implication that the C-H...O and C-H...Cl- interactions in the latter

Fig. 3. (a) 5-methylisocytosine cation; (b) 1-methylcytosine cation; (c) 2-amino-5-methylpyridinium cation.



case mimic the N-H...O and N-H...Cl- hydrogen bonds in the former case.

If the crystal structure of 5MIC.HCl is isostructural with that of 1-methylcytosine hydrochloride, we will have solid evidence that the C-H group can, in certain environments, act as a hydrogen bond donor. That the structure of 2A5MP.HCl is not isostructural with that of 1-methylcytosine hydrochloride indicates that the C-H...O and possibly the C-H...Cl- interactions were of some importance in stabilizing the structure of 1-methylcytosine hydrochloride (and the other hydrochloride salts of cytosine derivatives). If 5MIC.HCL is isostructural with 1-methylcytosine hydrochloride, we will then have evidence that the C-H...O and C-H...Cl- interactions were definitely of an attractive nature and can be called weak hydrogen bonds.

Footnotes for Proposition I

(1) W.C. Hamilton and J.A. Ibers, "Hydrogen Bonding in Solids," W.H. Freeman and Co., San Francisco, 1968.

(2) L. Pauling, "The Nature of the Chemical Bond," 3rd Ed., Cornell University Press, Ithaca, 1960.

(3) J. Donohue, "Structural Chemistry and Molecular Biology," N. Davidson and A. Rich, Ed., W.H. Freeman and Co., San Francisco, 1968.

(4) D.J. Sutor, Acta Crystallogr., 11, 453(1958).

(5) G.S. Parry, ibid., 7, 313(1954).

(6) T.J. Kistenmacher, D.J. Hunt, and R.E. Marsh, ibid., Sect. B, 28, 3352(1972).

(7) S.L. Edwards, J.S. Sherfinski, and R.E. Marsh, in preparation(1973).

(8) S.-H. Kim, H.M. Berman, N.C. Seeman, and M.D. Newton, Acta Crystallogr., Sect. B, 29,703(1973).

(9) P. T'so, N. Kondo, M. Schweizer, and D. Halles, Biochemistry, 8, 997(1969).

(10) J.S. Sherfinski and R.E. Marsh, Acta Crystallogr., Sect B, 29, 192(1973).

(11) J.S. Sherfinski, Candidacy Report and Propositions, California Institute of Technology, 1971.

(12) J.S. Sherfinski, Ph. D. Thesis, California Institute of Technology, 1973. Also, J.S. Sherfinski and R.E. Marsh, in preparation(1973).

(13) B.L. Trus and R.E. Marsh, Acta Crystallogr., Sect. B, 28, 1834(1972).

Proposition II.

In 1965 the active component of a fermentation broth was isolated and characterized as anthramycin (1,2; see Fig. 1). This compound has been found to exhibit a wide range of antitumor, antimicrobial and chemosterilant properties (3,4). Early work showed that anthramycin's principal mode of action is its complexation with double-helical DNA to prevent DNA and RNA synthesis (5). Further work, described in more detail below, has indicated that anthramycin's interaction with DNA consists possibly of the formation of a covalent bond and at least partly involves the phosphates of the polymeric nucleic acid. There is also some evidence that anthramycin can dimerize to form an actinomycin-like chromophore, which may in turn be the active form of the drug. I propose to investigate the interaction of anthramycin with helical DNA to determine the active form and mode of operation of the drug.

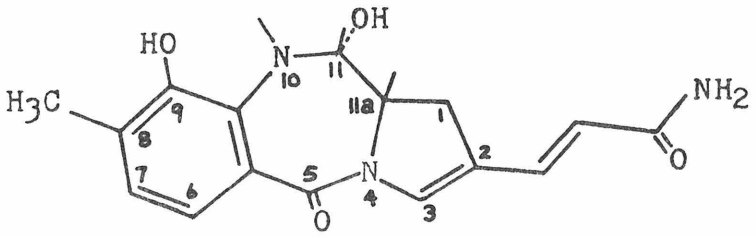


Fig. 1. Anthramycin.

A number of studies have demonstrated the rather slow rate of reaction of anthramycin with DNA. Kohn and Spears have investigated the rate of complexation as a function of pH (6), and found a catalytic effect by the hydrogen ion. Effects of ionic strength have also been studied(5-7). Monovalent cations were shown to slow the complexation reaction by an amount one would predict from ionic strength considerations. Divalent cations such as Mg++ and Mn++ showed a much greater effect than might be predicted on the basis of ionic strength alone. It is known that Mg++ and Mn++ both prefer an oxygen coordination shell and are probably coordinating with the phosphate oxygens(8). The The anomalous effect of Mn++ and Mg++ on the rate of complexation is evidence that anthramycin also coordinates to DNA through the phosphates of the polymer.

The intercalating model proposed for the actinomycin-DNA complex (9,10), although attractive in some respects, can be rejected as a reasonable model for the anthramycin-DNA complex. Studies of the competitive complexation of anthramycin and actinomycin with DNA have been reported (6), and show that both antibiotics are able to complex to the same DNA polymer and that the existence of one type of complex does not preclude the formation of the other.

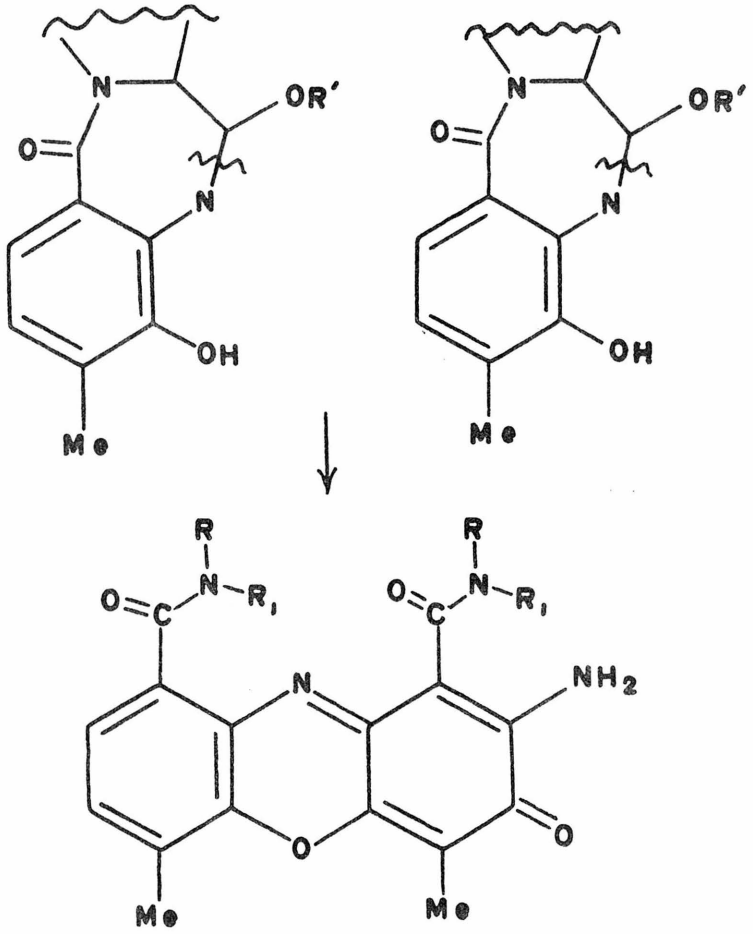
On the basis of the high stability of the anthramycin-DNA complex, it has been proposed that the interaction of anthramycin and DNA is at least in part a covalent one(5,6).

Dialysis experiments have shown that less than 20% of complexed anthramycin is dialyzed over a period of 14 days when under the same conditions free anthramycin is 95% dialized in 24 hours. Basic solution strong enough to denature double-stranded DNA did not dissociate the anthramycin-DNA complex, even though the DNA in the complex is single-stranded under those conditions. Reaction of free anthramycin with single-stranded DNA is slow at best, however, indicating that the tertiary structure of DNA is more critical to the complex formation than to the complex stability. Increasing the ionic strength to a point where single-stranded DNA is thought to assume a non-random tertiary structure similar to double-helical DNA increases the rate of complexation (5,6).

The stability of the anthramycin-DNA complex in the presence of certain destabilizing agents was compared to that of the actinomycin-DNA complex. Under conditions where the actinomycin complex dissociates (alcoholic precipitation, sodium lauryl sulphate treatment, Ag^+ treatment), the anthramycin complex was stable (6).

The experimental evidence mentioned above all tends to discount an intercalative model for the interaction of anthramycin with DNA, even though one can outline a dimerization reaction of anthramycin which produces the actinomycin chromophore (Fig. 2). In addition, Leimgruber has observed that anthramycin will dimerize under mildly

Fig. 2. A schematic outline of an anthramycin dimerization yielding the actinomycin chromophore.



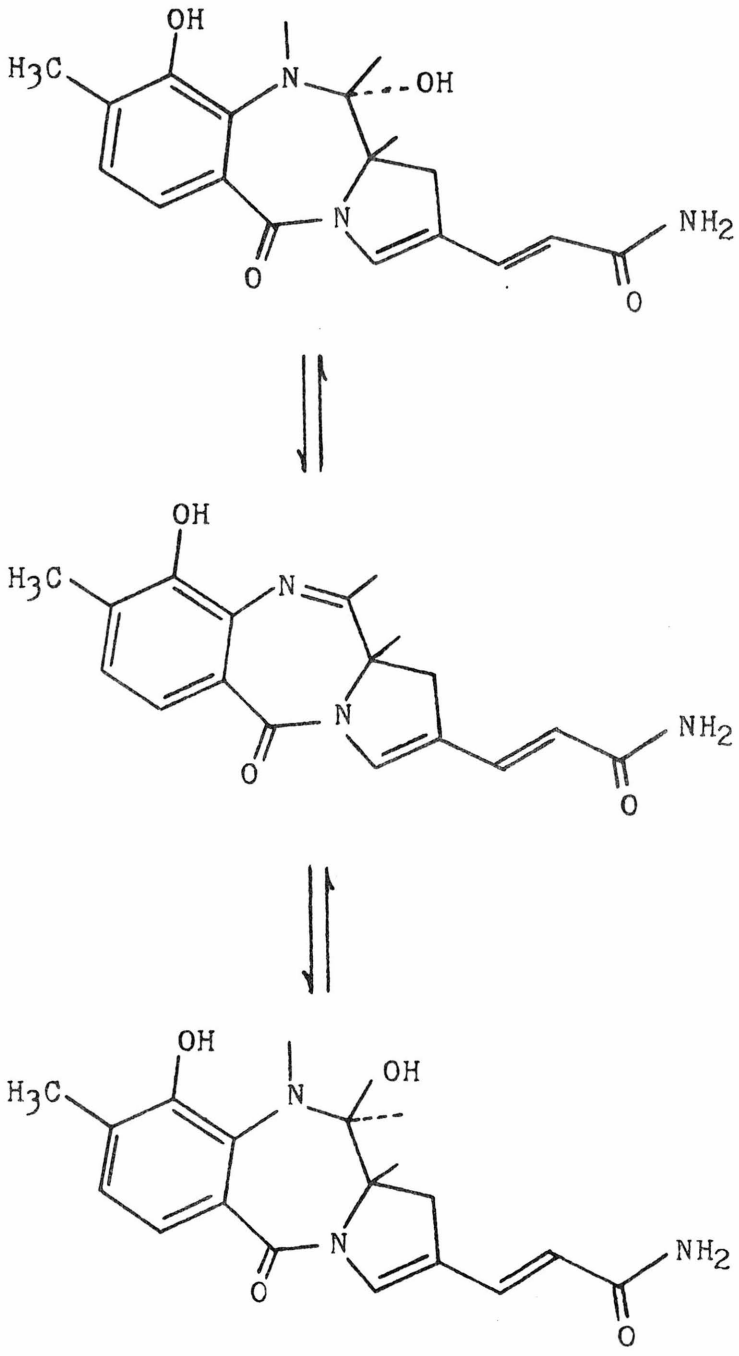
oxidative conditions to yield a compound containing the actinomycin chromophore (2). Therefore, one can not discount completely the possibility of such a dimer forming and interacting with DNA.

None the less, it seems likely that anthramycin as a monomer interacts with double-helical DNA to form a complex which in itself is quite stable yet can reversibly dissociate to free anthramycin and DNA. I propose that a possible mechanism for the complexation of anthramycin is the formation of a phosphate ester through O(11) of anthramycin, and that the -OH, NH, and acrylamide groups of anthramycin, necessary for complex formation (3), promote and preserve the proper stereochemistry necessary for the ester formation.

The observed behavior of the anthramycin-DNA complex is consistent with this mechanism: reaction rate is dependent on pH, but only catalytically (6); strongly acidic conditions can cause the complex to dissociate (6); ions known to complex with nucleic acid phosphates adversely effect the anthramycin-DNA complexation (7); the anthramycin-DNA complex is stable even if the double helix is denatured (5,6); anthramycin in aqueous solution readily undergoes the addition and elimination of water (Fig. 3; ref. 1).

Also, substitution studies have been performed (3) which show that the hydroxyl groups bonded to C(9) and C(11)

Fig. 3. The equilibrium reactions of anthramycin in water.



N(10) are associated with the active site of the molecule. Placing a keto group at C(11) or acetylating N(10) makes the anthramycin derivative inactive. Substituting a methyl ether at C(9) reduces considerably the potency of the drug, as does altering the acrylamide grouping at C(2). These observations are consistent with the phosphate ester hypothesis by requiring these side groups as aids in coordination and orientation.

I propose that investigations be carried out to prove or disprove that formation of a phosphate ester linkage is the mode of interaction of anthramycin with DNA. It has been noted(6) that Ag+ ions react with and decompose the anthramycin molecule, yet are unable to affect the antibiotic in complex form. Experiments should be started to determine the site(s) of attack of Ag+ on anthramycin to more clearly define the part of the drug involved in the complex. It should also be worthwhile to study the possible ester formation with phosphate ions alone and in the form of double helical oligonucleotides. The investigation of these reactions using such techniques as ^{31}P NMR and oxygen labeling may indicate if ester formation is taking place.

Footnotes for Proposition II

(1) W. Leimgruber, V. Stefanović, F. Schenker, A. Karr, and J. Berger, J. Amer. Chem. Soc.,87, 5791(1965).

(2) W. Leimgruber, A.D. Batcho, and F. Schenker, ibid., 87, 5793(1965).

(3) S.B. Horwitz, "Progress in Molecular and Subcellular Biology," Vol. 2, F. Hahn, Ed., Springer-Verlag, Berlin, 1971.

(4) S.B. Horwitz, S.C. Chang, A.P. Grollman, and A.B. Bořkovec, Science, 174, 159(1971).

(5) K.W. Kohn, V.H. Bono, and H.E. Kann, Biochim. Biophys. Acta, 155, 121(1968).

(6) K.W. Kohn and C.L. Spears, J. Mol. Biol., 51, 551(1970).

(7) V. Stefanović, Biochem. Pharmacol., 17, 315(1968).

(8) N.R. Sarker and A.L. Dounce, Biochim. Biophys. Acta, 49, 160(1961).

(9) W. Muller and D.M. Crothers, J. Mol. Biol., 35, 251(1968).

(10) H. M. Sobell, "Progress in Nucleic Acid Research and Molecular Biology," Vol. 13, J.N. Davidson and W.E. Cohn, Ed., Academic Press, New York, N.Y., 1973, p 153.

Proposition III.

Biochemical studies in recent years have included the reactions of macromolecules - notably polynucleotides, proteins, and lipid bilayers - to better understand the mechanisms of metabolism. Unfortunately little is known about these reactions on a molecular level, and thus a complete understanding of most mechanisms in question is impossible at this time. Application of crystallographic techniques to macromolecules has of late mushroomed to the point where now few proteins are safe from this type of investigation, whether or not a three-dimensional analysis of such a protein is worth the time and effort necessary. A disadvantage of protein crystallography is that the resulting structure is that of the protein in a static state, and any conclusions as to the dynamics of the protein mechanism must include some hypothesis of conformation change.

None the less, there are certain systems which when analyzed by protein crystallographic techniques can yield an abundance of information. I propose to carry out a complete structural analysis of one such system using protein crystallographic techniques. The system is an intermolecular complex of an aminoacylated tRNA (transfer ribonucleic acid), GTP (guanosine -5'-triphosphate), and a protein factor called T_u. The significance of this complex is at least two-fold: first, it is, as a complex, the form

in which amino acids are presented one by one to a synthesizing polypeptide chain; second, the structure of this complex would be of tremendous importance in understanding the general nature of protein-nucleic acid interfaces, and also give us some idea as to the rigidity of nucleic acids.

To understand more fully the role this complex assumes in protein synthesis, a short description of the process will be given. Proteins are composed of polypeptide chains and often other prosthetic groups such as hemes. What will be discussed here is the synthesis of the polypeptide chain and not the assembly of component parts to form a total protein. For a more complete discussion of protein synthesis, see references 1-3.

The basic "assembler" of polypeptide chains is a unit called the ribosome. The ribosome is composed of two smaller subunits, each of which contains ribonucleic acids (RNA's) and a number of proteins. Ribosomes construct polypeptide chains according to instructions delivered to them by messenger RNA (mRNA) molecules, which in turn have obtained the information from the cell's genetic material. Messenger RNA molecules are polymers of ribofuranosyl-nucleotides, where the bases of the nucleotides may be adenine, guanine, uracil, or cytosine. These nucleotides are arranged along the mRNA polymer in series of three,

where each three-nucleotide sequence specifies a particular amino acid. Thus a series of the three-base sequences (codons) will specify a particular amino acid series which results in a particular protein sequence.

Transfer RNA molecules are relatively small (about 100 nucleotides long) nucleic acids which are able to interpret the codons of the mRNA and insure that the proper amino acid is presented to the growing polypeptide chain. The tRNA molecules have one exposed end that is called the "anticodon loop." This anticodon is able to hydrogen bond Watson-Crick fashion to specific codons of the messenger RNA. To the other end of the tRNA is attached the amino acid appropriate for the particular anticodon. For instance the codon for phenylalanine is uridine-uridine-uridine(UUU). The anticodon on phenylalanine tRNA is then adenine-adenine-adenine (AAA). These codons and anticodons are able to hydrogen bond specifically together to insure that a UUU codon implies a phenylalanine amino acid.

It is the function of the ribosome to synchronize and allow for the specific interpretation of the mRNA codons. Associated with the ribosome are two binding sites (acceptor and donor sites) which expose two adjacent three-nucleotide codons of a mRNA. The region of the mRNA about these two codons is protected by the ribosome and is not exposed to the surroundings. To the donor site is attached a tRNA

molecule which is hydrogen bonded to the codon exposed there and which has attached to it the growing polypeptide chain. To the codon exposed at the acceptor site is attached a tRNA molecule with the proper anticodon and amino acid. The polypeptide chain is transferred from the tRNA at the donor site and forms a peptide bond with the amino acid bonded to the tRNA at the acceptor site. The uncharged tRNA is then expelled, the ribosome moves along the mRNA molecule so that the tRNA that was at the acceptor site is now at the donor site, and a new codon is exposed at the acceptor site. Another tRNA molecule binds, and the process continues.

When the aminoacylated tRNA (AA-tRNA) is presented to the synthesizing ribosome, it is in the form of a complex with GTP and the protein factor T_u. With the aid of a second protein factor called T_s, this complex releases the AA-tRNA to be bound to the acceptor site of the ribosome and a T_u-T_s dimer is formed, all at the expense of GTP $\rightarrow$ GDP+P_i.

A significant contribution to understanding the chemistry of polypeptide synthesis on the molecular level has been the elucidation of the three-dimensional structure of a tRNA molecule (4). Results so far show its anticodon exposed at the long end of an "L" shaped molecule, while the aminoacylation site is at the short end of the "L." There are the complimentary helices predicted by the proposed cloverleaf structure (5), but the resolution of the structure at present is just great enough to allow recognition of the

phosphate backbone. In addition, the successful crystallization of the protein factor T_u has recently been reported (6). I propose that the next logical step in understanding the mechanism of polypeptide synthesis is to study the three-dimensional structure of the T_u-AAtRNA-GTP complex. With this one structure one could see just how much the tRNA molecule must contort itself to become part of the complex. New insights as to the nature of protein-nucleic acid interactions would also be possible. I propose to do such a crystal structure analysis.

There are of course certain problems associated with the project. One is that, although both tRNA and T_u protein have been crystallized separately (4,6), the origins of the two macromolecules were different. The tRNA was from yeast and the T_u from E. coli. The compatibility of the nucleic acid and protein coming from two different organisms is yet to be demonstrated. Efforts should be made to isolate and purify the two components from the same organism if possible. Then there is the ever-present problem in protein crystallography of "will I be able to get crystals?" That the two components crystallize separately is encouraging, but the stability of the complex in crystalline form has not been determined. In conjunction with the problem of growing crystals of the native complex, one must also obtain isomorphous crystals containing heavy-atom ions or complex ions to obtain phase information necessary to Fourier

transform the measured structure factor magnitudes to obtain an electron density map (7). The successful growth of isomorphous crystals is often a major problem in macromolecular crystallography. Regardless, these problems exist for all large molecule crystallography projects; the possible results of this project seem important enough to engage in the endeavor at hand.

Footnotes for Proposition III

(1) G. Attardi, Ann. Rev. Microbiol., 21, 383(1967).

(2) P. Lengyel and D. Soll, Bacteriol. Revs., 33, 264(1969).

(3) J.L.- Lenard and F. Lipmann, Ann. Rev. Biochem., 40, 409(1971).

(4) S.H. Kim, G.J. Quigley, F.L. Suddath, A. McPherson, D. Sneden, J.J. Kim, J. Weinzierl, and A. Rich, Science, 179, 285(1973).

(5) F. Cramer, "Progress in Nuceic Acid Research and Molecular Biology," Vol. 11, J.N. Davidson and W.E. Cohn,Ed., Academic Press, New Tork, N.Y., 1971, p 391.

(6) D. Sneden, D.L. Miller, S.H. Kim, and A. Rich, Nature(London), 241, 530(1973).

(7) D. Eisenberg, "The Enzymes," Vol. I, 3rd ed., P.D. Boyer, Ed., Academic Press, New York, N.Y., 1970, p 1.

Proposition IV.

In the past several years there has been a rapid increase in the number of investigations to find a way or ways of defeating cancer. Besides trying to discover what actually causes cancer, and therefore being able to prevent it, researchers are also working hard to develop methods of combatting a replicating cancer. These methods of attack are related, since once the mechanism(s) of cancer initiation is known, it should be much easier to design a means of stopping the spread of cancers. In general though, one can choose a method of treatment which will stop or retard normal cell growth, since cancer cell growth will also be affected. One common method is that of radiation treatment. The effectiveness of this technique depends on the fact that radiation adversely affects replicating genetic material, and for most cases it is the cancer that has an abundance of replicating cells relative to the normal surrounding tissue. In fact, the basis of all cancer treatment rests on the premise that whatever one does to inhibit cell growth, it is the growth of cancer cells that will be most affected.

Chemotherapy is another means of treating malignancies. A patient may be given regular doses of a certain drug with the hope that the cancer will be killed before much normal tissue is destroyed. This is not always the case, however,

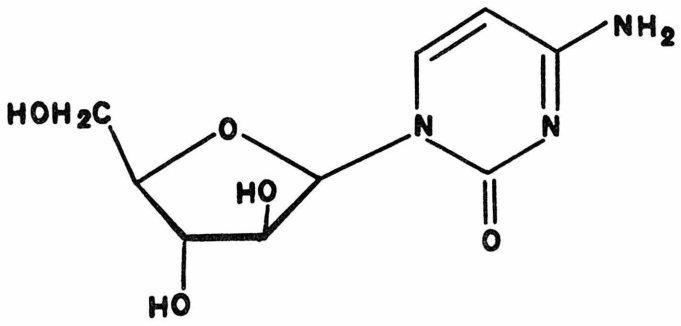
as oftentimes the huge doses needed to affect cancer growth affect other normal cells adversely. Such a reaction is manifested in symptoms like loss of hair and nausea.

A class of chemotherapeutic agents now being investigated is that of compounds which mimic the naturally occurring bases, nucleosides, and nucleotides. The mimicry can be of several types. One may be that in which the sugar of a natural nucleoside is altered slightly, such as in cytosine arabinoside (Fig. 1a). Alternatively the base of the nucleoside (or the free base) could be changed slightly, as in 8-azaguanine (Fig. 1b).

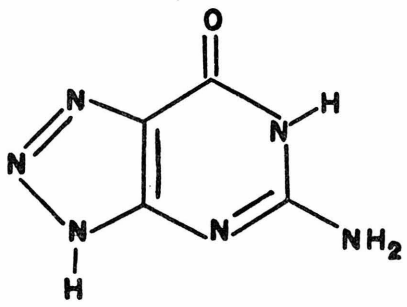
Metabolic inhibition by nucleoside analogues can be by several different mechanisms. The analogue may mimic the natural nucleoside such as to inhibit certain enzymatic reactions which require that nucleoside as a substrate. The analogue may also be incorporated into synthesizing nucleic acid where it may either upset DNA replication or RNA replication and ultimately protein synthesis. In addition, since many metabolic reactions are feedback-controlled by their reaction products or products resulting from reactions further along a metabolic pathway, analogues may, by mimicing reaction products, be able to inhibit or promote reactions which should either continue or slow down (1-3).

Recently there has been published the syntheses of several nucleoside analogues which in preliminary testing

Fig. 1. (a) arabinofuranosylcytosine; (b) 8-azaguanine.



(a)



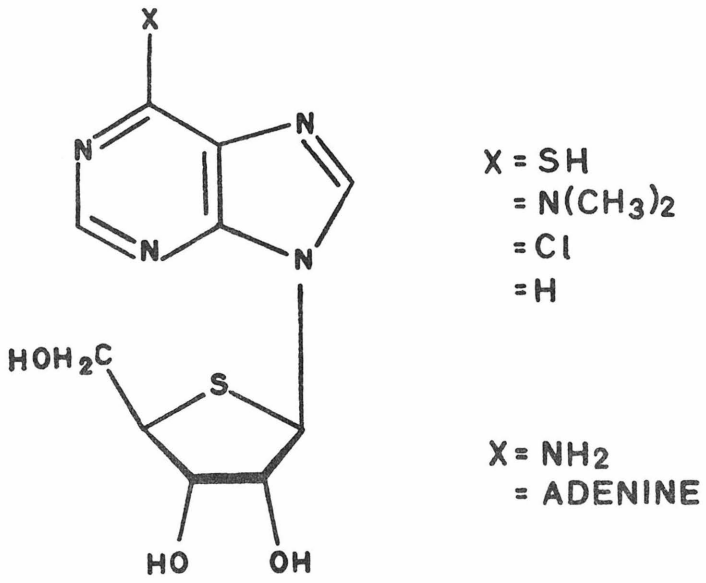
(b)

demonstrated in vitro activity against a number of cancer types (4,5). These compounds are based on the ribosyl analogue 4'-thioribofuranose, where a sulfur has substituted for O(1') in the sugar, and have so far been synthesized as a number of different sugar-adenine derivatives. I propose that the use of this new class of compounds as a possible new source of chemotherapeutic drugs be investigated.

The compounds so far tested are shown in Fig. 2. Note that all of these nucleoside analogues combine the characteristics of the sulfur-substituted sugar and a purine base which has by itself been shown effective in retarding cancer growth. All of the analogues shown (except the adenine derivative) were tested, and the results compared with the activities of the corresponding ribosyl derivatives. In some cases it was the ribosyl purine that was more effective, while in others it was the thioribosyl drug which was more potent.

The need for effective chemotherapeutic agents is obvious. I believe that investigations should be made to determine the efficacy of these thioribosyl compounds, as antimetabolites, with both natural and synthetic bases attached. In addition to the adenine derivatives, other base substitutions should be tested; it is not uncommon that for a specific type of analogue one base derivative may be completely inactive while another may be a potent antimetabolite (an example is arabinosylcytosine vs

Fig. 2. 4'-thioribosyl analogs of adenosine which have been synthesized and tested.



arabinosyluracil; see this thesis, p 3). Different organisms should also be used as test sites. Very often what adversely affects virus growth may not affect bacterial growth, and protozoan growth may be influenced in a third way, all different from the effects on mammalian cell growth. Different cancer types should also be tested, as they often behave differently to various antimetabolites (1-3).

In addition to the immediate benefit of perhaps discovering an effective anti-cancer drug, this project should be undertaken for the purpose of gaining insight into the intricacies of metabolic reaction pathways. Nucleotide analogues have been very useful in isolating a particular step or steps in biological reactions which had gone previously unrecognized (1). Antimetabolites are also of use in reducing the immune responses of tissues after transplants (6).

To help in determining the mode of action of these drugs, one should investigate what biochemical reactions they can and cannot undergo. Do the thioribosides affect nucleic acid synthesis directly? What type of nucleic acid? If DNA, must the drug be introduced as the deoxy form, or can it be reduced enzymatically? Must the drug be phosphorylated first, or can it be phosphorylated enzymatically? Is the drug incorporated into a synthesizing nucleic acid polymer, and if so, where? What happens to

polymerization when this happens? What about subsequent replication? If it is found that any of these reactions are inhibited by the thioribosides (or deoxythioribosides), can the inhibition be reversed by the administration of an excess of the corresponding natural nucleoside? Answers to such questions would do much to define the mechanism(s) of the thioribosides and could yield valuable knowledge to that we have of various metabolic reactions.

Footnotes for Proposition IV

(1) P.R.- Burman, "Analogues of Nucleic Acid Components," Springer-Verlag, New York, N.Y., 1970.

(2) R.J. Suhadolnik, "Nucleoside Antobiotics," Wiley-Interscience, New York, N.Y., 1970.

(3) M.E. Balis, "Antagonists and Nucleic Acids," North-Holland Publishing Co., Amsterdam, 1968.

(4) M. Bobeck, R.L. Whistler, and A. Block, J. Med. Chem., 13, 411(1970).

(5) M. Bobeck, R.L. Whistler, and A. Block, ibid., 15, 168(1972).

(6) G.B. Elion and G.H. Hitchings, "Advances in Chemotherapy," Vol. 2, A. Goldin, F. Hawkin, and R.J. Schnitzer, Ed., Academic Press, New York, N.Y., 1965, p 91.

Proposition V.

Circuitry common to X-ray diffractometry counting systems employs, in order, a scintillation device which outputs photons in the visible spectrum when X-ray photons are received, a photomultiplier which outputs an electronic pulse for each packet of visible photons received, a preamplifier, a linear amplifier which shapes the pulse, a pulse height analyzer which discards all those pulses coming from the linear amplifier which are either too strong or too weak, and a scaler-counter which records logic pulses output by the pulse height analyzer.

Inherent to the observation and recording of any phenomenon, not just X-ray photons, is the finite amount of time necessary to observe and record an event. This time period varies from detector to detector, and in a sense can be called "dead time," since if another event occurs while the system is recording the first, the second will not be recorded. In the case of X-ray diffractometry, as the intensity of the diffracted beam being measured increases, the probability of second and third X-ray photons reaching the detector while the first is registering also increases, and may increase to the point where the number of X-ray photon counts lost is a considerable fraction of those to be counted.

Each of the above-mentioned components of a detector circuit has associated with it a dead time, which is supposedly an indication of how long that component is incapacitated after recording an event. In the same fashion one can assign to a complete counting system a value for a dead time which is then an estimate of the counting efficiency of the system in toto. A number of methods are available to determine a value for the dead time of a diffractometer counting system (1-3). Employing such a dead time parameter, a correction factor can be derived to account for dead time losses in counting rates:

$$u_t = u_o/(1-u_o\tau)$$

where u_t is the true counting rate (number of events to be counted per unit time), u_o is the observed counting rate, and τ is the system dead time estimate. This correction assumes that to a first approximation u_o is of the same magnitude as u_t. The validity of this assumption is questionable if one considers not only the number of events that one thinks are contributing to the total dead time (u_o) but also the number of events due to instrumental and electronic noise which are filtered out by the pulse height analyzer. If one reduces the lower discriminator level of the pulse height analyzer (PHA) to zero and records the number of counts reaching the scaler, the counting rate as indicated is orders of magnitude greater than that due to the diffracted X-ray beam alone (4). This high counting

rate suggests that many more events are occurring in the detector circuitry than reach the recorder, and so the number of events actually recorded is many times less than the number of events contributing to the dead time of the system. Since derivation of the dead time correction factor given above assumes that only recorded counts contribute to dead time losses, not only is the functional form of the factor wrong, but also the values commonly used for τ have no physical significance.

I propose that steps be taken to determine the origin of these low energy pulses and if possible to minimize their occurrence. Any correction to be used to account for dead-time losses must be determined with their existence in mind.

One possible source of the low energy pulses (LEP's) is the photomultiplier tube. This tube consists of a series of dynodes which amplify the initial pulse of electrons from the photocathode. The latter is energized by the visible light photons emitted by the scintillation crystal. Thermal electrons emitted by any of these dynodes will be multiplied to produce an output pulse of low potential but which will still be shaped and analyzed by the linear amplifier and PHA. They will still contribute to the total dead time of the system. Thermal emission could be checked as a possible source of these LEP's by operating the photomultiplier tube

at different temperatures and observing the effect on the LEP rate at the recorder.

There exists a commercial instrument that, according to the manufacturer, will output random or periodic tail pulses at a variety of frequencies (up to 1 MHz) and rise and fall times (5). If the specifications are true, one should be able to use this instrument to simulate input pulses for the preamplifier, linear amplifier, and possibly the pulse height analyzer. By adjusting the lower discriminator level of the PHA and observing the counting rate, one could learn if the origin of the LEP's is the scintillation crystal-photomultiplier, preamp, or linear amplifier. In any case, the source of the low enrgy pulses precedes the pulse height analyzer in the counting circuitry.

Given that the dead time correction factor, as commonly used, is of the wrong form and is not physically meaningful in that form, a better functional form should be developed. I propose that one such factor may be a polynomial expansion of the observed counting rate, where coefficients could be determined by the method of least squares ($u_t = P_n(u_o)$). The order $\underline{n}$ of the polynomial would be dependent on the number of observations made and the degree of accuracy desired. If by increasing $\underline{n}$ by 1 no significant improvement is noticed in the least-squares analysis, one could assume he had as many terms as were needed.

The advantages of using a polynomial expression to approximate the dead time correction are several-fold. First, such a function makes no assumption as to the real source of counting losses and is thus an unbiased correction factor. This is important since no corrections now in use consider the existence of the LEP's as contributing to dead-time losses. Second, a polynomial approximation for the true counting rate would allow the effective dead time to be a function of the counting rate and would not restrict it to a constant value. Third, a polynomial expansion would be easily adaptable to on-line computor use. By analyzing an intensity profile using the step-scan technique, dead-time corrections could be applied to each step count, and the sum of corrected step counts would then be the corrected scan count. Variances for intensities measured and corrected in this fashion would be adjusted to account for the the uncertainty in the correction factor.

Footnotes for Proposition V

(1) M.B. Blackman and J.L. Michiels, Proc. Phys. Soc., 60, 549(1948).

(2) "International Tables for X-Ray Crystallography," Vol. III, Kynoch Press, Birmingham, 1962.

(3) U.W. Arndt and B.T.M. Willis, "Single Crystal Diffractometry," University Press, Cambridge, 1966.

(4) This phenomenon has been observed on the G.E. quarter-circle diffractometer on Tube Stand 11, Caltech, and no doubt occurs on all others here as well.

(5) "Random Pulse Generator," from Berkeley Nucleonics Corp., Berkeley. Specifications give rise time limits of 0.1-20 μs, fall time of 5-100 μs, and frequency from 100Hz to 1MHz. Both random and periodic frequencies are selectable.