

*Acta Cryst.* (1972). B28, 3145

## The Crystal Structure of $\alpha,\alpha$ -Trehalose Dihydrate from Three Independent X-ray Determinations\*

By G. M. BROWN

*Chemistry Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, U.S.A.*

D. C. ROHRER AND B. BERKING

*Department of Crystallography, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, U.S.A.,*

AND C. A. BEEVERS, R. O. GOULD AND R. SIMPSON

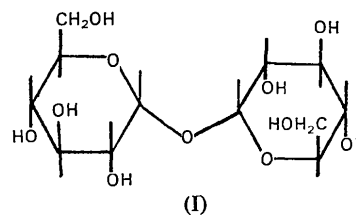
*Department of Chemistry, University of Edinburgh, Edinburgh EH9 3JJ, Scotland*

(Received 1 February 1972 and in revised form 11 May 1972)

The structure of the dihydrate of the nonreducing disaccharide  $\alpha,\alpha$ -trehalose ( $\alpha$ -D-glucopyranosyl  $\alpha$ -D-glucopyranoside) has been determined independently by direct methods in three different laboratories. There are four units  $C_{12}H_{22}O_{11} \cdot 2H_2O$  in an orthorhombic cell ( $P2_12_12_1$ ) with  $a=12.230$  (2),  $b=17.890$  (3), and  $c=7.596$  (1) Å. The discrepancy indices  $R(F)$  are: 0.057 (based on 3293 Mo  $K\alpha$  diffractometer data); 0.041 (1611 Cu  $K\alpha$  diffractometer data); 0.055 (2149 Cu  $K\alpha$  photographic data). Comparison of coordinates suggests that their standard errors may have been underestimated by a factor of  $\sim 2$ . The  $\alpha,\alpha$ -trehalose molecule has approximate  $C_2$  symmetry, but there are some significant structural differences between the two chemically equivalent halves. The angles at the ring oxygen atoms are  $114.0$  and  $114.2^\circ$ ; the angle at the glycosidic oxygen is  $115.8^\circ$ . The two rings have  $C1$  conformations; the conformations about the linking C-O bonds are stabilized by the conformational 'anomeric effect' and also by a complex system of hydrogen bonds involving all of the hydroxyl groups, both water molecules, and one ring oxygen. By hydrogen bonding each glucopyranosyl moiety is linked to the other in the same molecule indirectly through two different water molecules. The helical arrangement of hydrogen bonds about the screw axes parallel to  $c$  suggests the possibility of observing ferroelectric behavior.

### Introduction

$\alpha,\alpha$ -Trehalose (I), or  $\alpha$ -D-glucopyranosyl  $\alpha$ -D-glucopyranoside, is one of three isomeric D-glucopyranosyl D-glucopyranosides collectively known as the trehaloses (Birch, 1963). It is a nonreducing disaccharide composed of two  $\alpha$ -D-glucosyl units linked by a glycosidic oxygen bridge between their two anomeric carbon atoms, C(1) and C'(1) in the usual numbering system. The isomeric  $\alpha,\beta$ -trehalose and  $\beta,\beta$ -trehalose have been synthesized, but only the  $\alpha,\alpha$  isomer occurs naturally. Known also as mycose or mushroom sugar,  $\alpha,\alpha$ -trehalose is widespread in nature, where it is found in bacteria, insect blood, fungi, algae, lichens, and some higher plants. It accounts for as much as 15% of the weight of mushrooms and dried yeast, but apparently it is not present in more basic foodstuffs. As the major blood sugar of many, but not all, insects (Wyatt & Kalf, 1957), it is important in insect metabolism. It is specifically hydrolyzed by the enzyme trehalase isolated from the wax moth (Kalf & Rieder, 1958).



This paper reports the results of three X-ray crystal-structure determinations of  $\alpha,\alpha$ -trehalose dihydrate carried out independently and simultaneously in our three separate laboratories. This is the first structure reported of a carbohydrate oligomer containing a glycosidic oxygen atom which is axial with respect to two sugar rings. A comparison of the two nonequivalent D-glucosyl moieties of the molecule in the crystal is interesting because the significant differences between the two must result from the packing of the molecules, which in isolation have chemically equivalent moieties and inherent twofold axes of symmetry.

### Experimental

$\alpha,\alpha$ -Trehalose dihydrate (m.p.  $96-97^\circ\text{C}$ ) is very soluble in water and in aqueous ethanol, but nearly insoluble in anhydrous ethanol. It can be crystallized in anhydrous form from dry pyridine, but from aqueous ethanol

\* Research sponsored in part by the U.S. Atomic Energy Commission under Contract with the Union Carbide Corporation, and in part by the National Institutes of Health, United States Public Health Service, under Grant GM 11293.

it comes out as a beautifully crystalline dihydrate. Groth (1910) described the dihydrate as orthorhombic bisphenoidal and gave the axial ratios  $a:b:c=0.6814:1:0.4171$ , on the basis of goniometric observations dating back to 1857.

X-ray photography and diffractometry confirm that the dihydrate is orthorhombic and establish  $P2_12_12_1$  as the space group, the observed extinctions being reflections  $h00$ ,  $0k0$ , and  $00l$  for odd values of  $h$ ,  $k$ , and  $l$  respectively. The largest of the crystals available at Edinburgh showed a feeble piezoelectric response.

The three sets of unit-cell translations are in remarkably good agreement (see Table 1, which presents a synopsis of the experimental work). The Oak Ridge cell parameters were derived by the least-squares method (Busing & Levy, 1967; Busing, 1970) from the  $2\theta$  angles of 12 resolved Mo  $K\alpha_1$  reflections in the range  $40.8^\circ$  to  $44.4^\circ$   $2\theta$  recorded with the Oak Ridge computer-controlled diffractometer (Busing, Ellison, Levy, King & Roseberry, 1968). The 12 reflections were 6 pairs of Friedel mates; the  $2\theta$  angles of the two reflections of each pair were measured in opposite senses to nullify any possible error in the zero of the  $2\theta$  scale of the diffractometer. The Pittsburgh parameters were derived by a least-squares method using  $2\theta$  angles for 20 general Cu  $K\alpha$  reflections in the  $2\theta$  range  $45$  to  $60^\circ$ . The  $2\theta$  values were obtained on a Picker FACS-I diffractometer from measurements on both the positive and negative sides of the  $2\theta$  circle. The Edinburgh translations were derived by extrapolation procedures from angle data obtained with a knife-edged camera which provides a standard high angle.

The recording of intensity data (see Table 1) with the Oak Ridge computer-controlled diffractometer and preliminary data reduction followed procedures vir-

tually identical with those used in a study of 6-mercaptapurine monohydrate (Brown, 1969). In particular, the  $\omega$  step-scan technique devised by Levy (1966) and described briefly in the 6-mercaptapurine paper was used for reflections below  $22^\circ$   $2\theta$  to provide more reliable background determinations on the low-angle sides of the Bragg peaks than is possible with  $\theta$ - $2\theta$  scans. The recording of reflections in the highest angle range,  $60$  to  $65^\circ$   $2\theta$ , was not completed, because the crystal spontaneously split into two pieces, possibly as the effect of contraction of a thin film of glue on the base of the crystal upon cracks produced when the crystal was cut from a larger one. There had been no indication in the intensity values recorded periodically (every 20 reflections) for a reference reflection that deterioration of the crystal was occurring.

For each reflection a statistical standard error  $\sigma_c(|F^2|)$  was calculated according to the equation

$$\sigma_c^2(|F^2|) = (AS/Lp)^2(T + k^2B),$$

where  $S$  is the scale factor on the intensities,  $A$  is the correction factor for absorption,  $L$  is the Lorentz factor,  $p$  is the polarization factor,  $T$  is the total count  $B$  is the background count, and  $k$  is the ratio of times spent measuring  $T$  and  $B$ . The standard errors were used in calculating the weights in the subsequent least-squares refinement (see below). However, the low-angle background measurements were not considered reliable even in the  $\omega$ -scan method for  $2\theta < 12^\circ$ . Accordingly, the 34 reflections of lowest angle were later given zero weights in the final refinement cycles, along with two other reflections for which the intensities were discovered to have been improperly recorded. Otherwise, all reflections were considered as observed.

Table 1. *Experimental data for  $\alpha,\alpha$ -trehalose dihydrate*

Orthorhombic, space group  $P2_12_12_1$ ; asymmetric unit  $C_{12}H_{22}O_{11} \cdot 2H_2O$ , formula weight 378.2; cell volume  $1662 \text{ \AA}^3$ , observed density  $1.512$ , calculated density  $1.511 \text{ g.cm}^{-3}$ ,  $Z=4$ ;  $\mu(\text{Cu } K\alpha)=12.2$ ,  $\mu(\text{Mo } K\alpha)=1.49 \text{ cm}^{-1}$ .

Cell parameters	Oak Ridge	Pittsburgh	Edinburgh
$a$	12.229 (2) $\text{\AA}$	12.230 (2) $\text{\AA}$	12.233 (3) $\text{\AA}$
$b$	17.890 (3)	17.890 (3)	17.889 (4)
$c$	7.597 (1)	7.595 (1)	7.596 (2)
$\lambda$	0.70926	1.54178	1.5405, 1.5443
Temperature	$23.5 \pm 0.5^\circ\text{C}$	$17.0 \pm 1.0^\circ\text{C}$	$18.0 \pm 1.0^\circ\text{C}$
Intensity data			
Crystal source	Mann Research*	Dr D. French†	Dr G. G. Birch‡
shape	Polyhedron	Polyhedron	Spheres
size (mm)	$0.3 \times 0.3 \times 0.4$	$0.3 \times 0.4 \times 0.4$	$0.3$ - $0.4$ radii
Data collection technique	Diffractometry, $\omega$ and $\theta$ - $2\theta$ scans	Diffractometry, $\theta$ - $2\theta$ scans	Weissenberg photography
Radiation	Mo $K\alpha$	Cu $K\alpha$	Cu $K\alpha$
$(\sin \theta/\lambda)_{\text{max}}$ ( $\text{\AA}^{-1}$ )	0.758	0.599	0.645
Independent data	3329	1634	2149
Absorption correction	Yes§	No	Yes

\* Mann Research Laboratories, 136 Liberty Street, New York, U.S.A.

† Iowa State University, Ames, Iowa, U.S.A.

‡ National College of Food Technology, Weybridge, Surrey, England.

§ The range of correction factors was 1.033 to 1.056.

The least-squares method of Thiessen, Levy & Brown (1970*a, b*) provided normalized structure factors (Hauptman & Karle, 1953), the scale factor, and a preliminary over-all anisotropic temperature factor. Statistical tests clearly indicated the centric distribution for the zonal intensities and the acentric one for the general intensities.

The Pittsburgh intensity data were measured on a Picker FACS-I diffractometer up to  $2\theta = 130^\circ$ , using the  $\theta$ - $2\theta$  scan mode with a scan speed of  $2^\circ 2\theta$  per minute. The scan width used for each intensity measurement was  $1.7^\circ$  plus a dispersion term for Ni-filtered Cu  $K\alpha$  radiation. Background counts of 10 sec duration were recorded at the beginning and end points of each scan. Three reference reflections recorded every 49 reflections varied less than 2.5% over the period of the data collection. Therefore, it seems that no crystal deterioration was occurring.

The intensity data were reduced to structure factors and normalized to  $E$ 's by the programs of Shiono (1971). Using the criterion that a measured intensity less than two standard deviations be considered unobserved, 23 of the 1634 independent reflections were considered unobserved and assigned weights of zero. No corrections for absorption were made.

The Edinburgh intensity data were obtained by the equiinclination Weissenberg method. Photographs were taken with the crystal rotating about the  $a$  axis ( $h=0-9$ ), the  $b$  axis ( $k=0-14$ ), and the  $c$  axis ( $l=0-5$ ). Multiple-film packs were used, and multiple exposures were made. The intensity comparisons and the interfilm and interlayer scaling were done very carefully for a total of 2149 independent reflections. This total, which includes 52 unobserved reflections, is no less than 98.5% of the accessible reflections for Cu  $K\alpha$  radiation. A few 'Renninger' reflections were noticed.

Absorption corrections appropriate for spherical crystals (*International Tables for X-ray Crystallography*, 1959, p. 299) were applied. The data generally had an internal consistency characterized by a 5% standard deviation, except for the very small and very large  $F$ 's, which were less reliably determined. The three axial projections were confirmed as centrosymmetrical by intensity statistics.

### Solution

In each of the three determinations the solution was achieved by direct methods. At Oak Ridge the tangent formula (Karle & Hauptman, 1956; Karle & Karle, 1966) was used in reiterative phasing, with the aid of Drew's program *PHASEM* (see Drew, Templeton & Zalkin, 1969; Drew, 1969). The starting set of phases (Table 2) included four phases set arbitrarily to specify the origin and enantiomorph (Karle & Hauptman, 1956) and seven others for which values were indicated with tentatively acceptable probabilities of correctness by a calculation with the  $\sum_1$  formula (Hauptman & Karle, 1953; Karle & Hauptman, 1956). The phasing

process converged for the 258 reflections with  $|E| < 1.66$  in 20 cycles. Of the highest 26 peaks in the  $E$  map computed with 258 phased  $E$ 's, 25 peaks were found to represent quite well the 25 carbon and oxygen atoms of the asymmetric unit, the 24th peak in order of decreasing height being spurious.

The acceptance of seven phases from the  $\sum_1$  calculation was perhaps not cautious; but since the chemical structure of  $\alpha, \alpha$ -trehalose is well known, no difficulty was expected in recognizing a false solution. As it turned out, for the intensity invariants with  $|E| \geq 1.50$  for which the  $\sum_1$  calculation was made, the 11 phase indications with highest probabilities of correctness ( $P \geq 0.717$ ) were all correct.

The tangent formula and the phase extension method of Hall (1967, 1970) were used to solve the structure in Pittsburgh. Six zonal reflections were used as the starting phase set\* (Table 2). These were extended to 227 phases with a minimum  $|E|$  of 1.4. The  $E$  map calculated using these phases showed only 25 major peaks corresponding to the entire trehalose molecule and the two water oxygens. No spurious peaks of significant height were generated.

At Edinburgh a prolonged attempt was made to solve the structure from Patterson projections and by three-dimensional Patterson methods. One ring (primed atoms in our numbering system) was reasonably accurately located, but the position of the other was off by about 0.5 Å (as seen in retrospect), so that a successful refinement was not possible.

After Patterson methods were abandoned in favor of direct methods, the solution was achieved by the symbolic addition procedure (Karle & Karle, 1966) and by the method of phasing with the tangent formula described by Germain, Main & Woolfson (1970, 1971), as implemented in the program *MULTAN* (Main, Woolfson & Germain, 1971). Readily interpretable  $E$  maps were obtained using 291 reflections with  $|E| \geq 1.38$  phased by the first procedure and 265 reflections with  $|E| \geq 1.50$  phased by the second procedure. The starting phases for both procedures are given in Table 2. The two sets of phases were about equally good, with about 90% of the zonal phases correct and with an average error of about  $35^\circ$  in the others. As at Oak Ridge, the  $\sum_1$  relation was found to work very well; although only two signs were indicated with high probability, they were both correct, as were the next half-dozen less probable indications.

\* The indices of these reflections have been transformed and the phases shifted appropriately from the original values, which were based on a unit cell with the translations  $a$  and  $b$  interchanged from the assignment specified in Table 1. Phase shifts are required because for space group  $P2_12_12_1$  the set of permissible origins (Hauptman & Karle, 1956) for one cell description does not coincide with the set for the other description, given that both determinations used the same form for the structure-factor expression. The origin shifts is  $m/4, n/4, p/4$ , where  $m, n$ , and  $p$  are odd integers whose particular values depend on the phase assignments made to fix the origin.

Table 2. Starting sets of phases  $\varphi_{tr}$  and  $\varphi_{sa}$  used in solving the structure by the tangent-refinement and symbolic addition methods, respectively

	<i>h</i>	<i>k</i>	<i>l</i>	<i>E</i>	$\varphi_{tr}$	$\varphi_{sa}$
Oak Ridge	5	0	2	4.05	0	
	9	5	0	3.41	$\pi/2$	
	0	17	1	3.34	$\pi/2$	
	14*	0*	7*	3.78	$\pi/2$	
	10	0	0	2.11	$\pi$	
	4	6	0	2.43	$\pi$	
	0*	12*	10*	1.94	$\pi$	
	0	14	0	2.39	0	
	12*	0*	8*	2.49	$\pi$	
	0	4	8	2.15	0	
	2*	26*	0*	2.19	$\pi$	
Pittsburgh	5	0	2	2.95	0	
	9	5	0	2.95	$\pi/2$	
	0	4	3	2.20	$\pi$	
	4	9	0	2.27	0 ( $\pi$ ) $\ddagger$	
	0	17	1	2.98	<i>a</i> ( $\pi/2$ ) $\ddagger$	
	0	9	2	2.85	<i>b</i> ( $\pi/2$ ) $\ddagger$	
Edinburgh	5	0	2	3.65	0	0
	9	5	0	3.20	$\pi/2$	
	0	17	1	3.14	$\pi/2$	
	11	4	0	2.32	$\pi/2$	
	10	0	0	2.07	$\pi$	
	0	18	6	1.61	0	0
	0	4	3	2.78	0, $\pi$ $\ddagger$	
	1	8	1	2.13	$\pi/4, -\pi/4$ $3\pi/4, -3\pi/4$ $\ddagger$	
	0	9	2	2.54		$\pi/2$
	6	8	3	2.82		<i>a</i> (0.11 $\pi$ ) $\ddagger$
4	13	4	2.41		<i>b</i> (-0.37 $\pi$ ) $\ddagger$	

\* Reflection beyond the Cu  $K\alpha$  limiting sphere.

$\ddagger$  Phase shifted from 0 to  $\pi$ .

$\ddagger$  Phases established for symbols and phases selected for reflections for which more than one starting phase was entered.

### Refinement

In the three determinations, (I), (II), and (III), denoting Oak Ridge, Pittsburgh, and Edinburgh, respectively, all 26 hydrogen atoms were located approximately by difference syntheses following some preliminary cycles of least-squares refinement. In the final cycles of each determination anisotropic thermal parameters were used for the carbon and oxygen atoms and isotropic parameters for the hydrogen atoms. In determination (I), the hydrogen thermal parameters were adjusted; in (II), held fixed at values for the parent C and O atoms reached in preliminary refinement; in (III), held fixed at 2.5 Å<sup>2</sup>. Hydrogen coordinates were adjusted in all three determinations. The scattering factors were from the following sources: (I) Hanson, Herman, Lea & Skillman (1964) for C and O and Stewart, Davidson & Simpson (1965) for H; (II) Cromer & Waber (1965) for C and O and Stewart *et al.* for H; (III) Berghuis *et al.* (1955) for C, Freeman (1959) for O, and McWeeny (1951) for H.

The full-matrix least-squares method was used in (I), with minimization of  $\sum w(\Delta|F|^2)^2$ ; the block-

diagonal method was used in (II) and (III), with minimization of  $\sum w(\Delta|F|)^2$ . For (I), except for zero weights noted already, the weights were set according to the equation

$$w(|F_o^2|) = 1/[\sigma_c^2(|F_o|^2) + (0.03|F_o|^2)^2],$$

where  $\sigma_c^2(|F_o|^2)$  is the purely statistical variance already defined and the term  $(0.03|F_o|^2)^2$  is an empirical correction term (Peterson & Levy, 1957). For (II), the basis of the weighting was essentially the same as for (I), except that the factor 0.03 was replaced by 0.05 and the equation appropriate for  $w(|F_o|)$  is

$$w(|F_o|) = 4|F_o|^2/[\sigma_c^2(|F_o|^2) + (0.05|F_o|^2)^2].$$

The scheme used in (III) was:

$$\begin{aligned} w(|F_o|) &= 1, & |F_o| < 14; \\ w(|F_o|) &= 14/|F_o|, & 14 \leq |F_o| \leq 50; \\ w(|F_o|) &= 700/|F_o|^2, & |F_o| > 50. \end{aligned}$$

The final measures of goodness of fit\* were: (I)  $R(F) = 0.057$  (excluding only data with zero weights),  $R_w(F^2) = 0.081$ , and  $\sigma_1 = 1.209$ ; (II)  $R(F) = 0.041$  for observed data only and 0.057 for all data,  $R_w(F^2) = 0.087$ , and  $\sigma_1 = 1.401$ ; (III)  $R(F) = 0.055$  (all data).

The final parameters and standard errors from all three determinations are shown in Table 3, which also includes a set of weighted-average coordinates and standard errors. The weight used for each coordinate was the reciprocal of the square of its standard error; the standard error reported for each average is the square root of the reciprocal of the sum of weights for that average. The coordinates describe the  $\alpha, \alpha$ -trehalose molecule in correct absolute configuration.

The observed and calculated structure factors ( $\times 10$ ) for all three determinations are given in Table 4. For a given reflection  $hkl$  the first entry gives the Oak Ridge data; the second, if present, gives the Edinburgh data; and the third, if present, gives the Pittsburgh data. $\ddagger$  The Table also includes in each entry for Oak Ridge either the standard error  $\sigma(|F_o|) \times 100$  [ $\sigma(|F_o|)$  being calculated as  $\sigma(|F_o^2|)/2|F_o|$ ], or, if  $|F_o|^2 \leq \sigma(|F_o^2|)$ , the error  $\sigma(|F_o^2|) \times 10$  (in which case a minus sign precedes the error). The 34 low-angle reflections given zero weight in final refinement are indicated by negative signs on the  $|F_o|$  values. Two other reflections given zero weights because they were not properly determined are marked with  $U$ 's. The 'unobserved' data among the entries for Pittsburgh are also marked with  $U$ 's. The unobserved Edinburgh data are represented by zero  $|F_o|$  values.

\* These measures of goodness-of-fit are explicitly defined by Brown (1969), among others.

$\ddagger$  Since the space group is noncentrosymmetric, it is important, at least in principle, to note that the Pittsburgh data are actually for reflections  $hkl$  rather than for  $h\bar{k}l$ . The negative sign on  $h$  arises in the index transformation mentioned in an earlier footnote.











Table 4 (cont.)

K	FO	FC	SI	GM	K	FO	FC	SI	GM	K	FO	FC	SI	GM	K	FO	FC	SI	GM	K	FO	FC	SI	GM
1	00	00	00	00	1	00	00	00	00	1	00	00	00	00	1	00	00	00	00	1	00	00	00	00
2	06	70	61	00	2	06	70	61	00	2	06	70	61	00	2	06	70	61	00	2	06	70	61	00
3	06	70	61	00	3	06	70	61	00	3	06	70	61	00	3	06	70	61	00	3	06	70	61	00
4	06	70	61	00	4	06	70	61	00	4	06	70	61	00	4	06	70	61	00	4	06	70	61	00
5	06	70	61	00	5	06	70	61	00	5	06	70	61	00	5	06	70	61	00	5	06	70	61	00
6	06	70	61	00	6	06	70	61	00	6	06	70	61	00	6	06	70	61	00	6	06	70	61	00
7	06	70	61	00	7	06	70	61	00	7	06	70	61	00	7	06	70	61	00	7	06	70	61	00
8	06	70	61	00	8	06	70	61	00	8	06	70	61	00	8	06	70	61	00	8	06	70	61	00
9	06	70	61	00	9	06	70	61	00	9	06	70	61	00	9	06	70	61	00	9	06	70	61	00
10	06	70	61	00	10	06	70	61	00	10	06	70	61	00	10	06	70	61	00	10	06	70	61	00
11	06	70	61	00	11	06	70	61	00	11	06	70	61	00	11	06	70	61	00	11	06	70	61	00
12	06	70	61	00	12	06	70	61	00	12	06	70	61	00	12	06	70	61	00	12	06	70	61	00
13	06	70	61	00	13	06	70	61	00	13	06	70	61	00	13	06	70	61	00	13	06	70	61	00
14	06	70	61	00	14	06	70	61	00	14	06	70	61	00	14	06	70	61	00	14	06	70	61	00
15	06	70	61	00	15	06	70	61	00	15	06	70	61	00	15	06	70	61	00	15	06	70	61	00
16	06	70	61	00	16	06	70	61	00	16	06	70	61	00	16	06	70	61	00	16	06	70	61	00
17	06	70	61	00	17	06	70	61	00	17	06	70	61	00	17	06	70	61	00	17	06	70	61	00
18	06	70	61	00	18	06	70	61	00	18	06	70	61	00	18	06	70	61	00	18	06	70	61	00
19	06	70	61	00	19	06	70	61	00	19	06	70	61	00	19	06	70	61	00	19	06	70	61	00
20	06	70	61	00	20	06	70	61	00	20	06	70	61	00	20	06	70	61	00	20	06	70	61	00
21	06	70	61	00	21	06	70	61	00	21	06	70	61	00	21	06	70	61	00	21	06	70	61	00
22	06	70	61	00	22	06	70	61	00	22	06	70	61	00	22	06	70	61	00	22	06	70	61	00
23	06	70	61	00	23	06	70	61	00	23	06	70	61	00	23	06	70	61	00	23	06	70	61	00
24	06	70	61	00	24	06	70	61	00	24	06	70	61	00	24	06	70	61	00	24	06	70	61	00
25	06	70	61	00	25	06	70	61	00	25	06	70	61	00	25	06	70	61	00	25	06	70	61	00
26	06	70	61	00	26	06	70	61	00	26	06	70	61	00	26	06	70	61	00	26	06	70	61	00
27	06	70	61	00	27	06	70	61	00	27	06	70	61	00	27	06	70	61	00	27	06	70	61	00
28	06	70	61	00	28	06	70	61	00	28	06	70	61	00	28	06	70	61	00	28	06	70	61	00
29	06	70	61	00	29	06	70	61	00	29	06	70	61	00	29	06	70	61	00	29	06	70	61	00
30	06	70	61	00	30	06	70	61	00	30	06	70	61	00	30	06	70	61	00	30	06	70	61	00
31	06	70	61	00	31	06	70	61	00	31	06	70	61	00	31	06	70	61	00	31	06	70	61	00
32	06	70	61	00	32	06	70	61	00	32	06	70	61	00	32	06	70	61	00	32	06	70	61	00
33	06	70	61	00	33	06	70	61	00	33	06	70	61	00	33	06	70	61	00	33	06	70	61	00
34	06	70	61	00	34	06	70	61	00	34	06	70	61	00	34	06	70	61	00	34	06	70	61	00
35	06	70	61	00	35	06	70	61	00	35	06	70	61	00	35	06	70	61	00	35	06	70	61	00
36	06	70	61	00	36	06	70	61	00	36	06	70	61	00	36	06	70	61	00	36	06	70	61	00
37	06	70	61	00	37	06	70	61	00	37	06	70	61	00	37	06	70	61	00	37	06	70	61	00
38	06	70	61	00	38	06	70	61	00	38	06	70	61	00	38	06	70	61	00	38	06	70	61	00
39	06	70	61	00	39	06	70	61	00	39	06	70	61	00	39	06	70	61	00	39	06	70	61	00
40	06	70	61	00	40	06	70	61	00	40	06	70	61	00	40	06	70	61	00	40	06	70	61	00

## Discussion

It seems appropriate to describe the structure primarily on the basis of the weighted-average coordinates of Table 3. In view of the probability that the estimated standard deviations of the individual determinations are unrealistically low, as is often the case (Hamilton & Abrahams, 1970), the deviations of the individual coordinates from their average values are hardly to be regarded as significant. In fact, if we multiply the apparent individual standard deviations by a factor of two, as has been suggested (Hamilton, 1965) as a rule of thumb, the extreme ratio of coordinate difference to corresponding standard deviation is only 3.1; and this extreme ratio is for a hydrogen coordinate, the  $x$  coordinate of H(O'6). Thus, although the agreement between the three sets of coordinates is not ideal, it appears to be about as good as is usually obtained when the same crystalline substance is subjected to X-ray analysis in different laboratories (Hamilton & Abrahams, 1970).

Fig. 1(a) shows the individual lengths and the average length for each C-C and C-O bond and the average apparent lengths of the C-H and O-H bonds.\* Since the differences among angle values have about the same magnitudes in relation to the standard errors as those of the bond lengths, we report for the valence angles and torsion angles (Klyne & Prelog, 1960) involving carbon and oxygen atoms only the values calculated from the averaged coordinates [see Fig. 1(b) and (c)]. The average angles C-O-H are also shown in Fig. 1(b). The average angles (not shown) involving the C-H bonds all lie in the range 106 to 114°. The shape of the molecule is shown in the stereoscopic drawing in Fig. 2, in which the carbon and oxygen atoms are represented by their thermal ellipsoids of 50% probability (Johnson, 1970).

For the average molecular parameters in Fig. 1, the standard errors, including the arbitrary factor of two (see above), are as follows: C-O and C-C bonds, 0.003 to 0.004 Å; C-H bonds, 0.03 to 0.04 Å; O-H

bonds, 0.04 to 0.05 Å; C-C-C, C-C-O, C-O-C, and torsion angles, 0.2°; C-C-H and O-C-H angles, 2°; C-O-H angles, 3 to 4°.

The C-C bond lengths in Fig. 1(a) based on the average coordinates have a mean of 1.523 Å and a range of 0.017 Å; the C-O bond lengths excluding those involving O(1), O(5), and O'(5) have a mean of 1.422 Å and a range of 0.011 Å. These values are in excellent agreement with those reported previously (Strahs, 1970) for  $\alpha$ -pyranosyl systems, namely, 1.524 Å (range 0.021) and 1.425 Å (range 0.016). According to Strahs, in  $\alpha$ -pyranosyl systems the anomeric effect on bond lengths (Berman, Chu & Jeffrey, 1967) leads on the average to deviations from the mean C-O length 1.425 Å of  $-0.024$ ,  $-0.004$ , and  $+0.015$  Å for the C(1)-O(1), C(1)-O(5), and C(5)-O(5) bonds, respectively. In trehalose, the corresponding deviations are  $-0.010$ ,  $-0.004$ , and  $+0.008$  Å for the unprimed system and  $-0.003$ ,  $-0.021$ , and  $+0.006$  Å for the primed system. The only major difference in these C-O bonds between the two halves of the molecule is the more marked shortening of C'(1)-O'(5), which may be related to the fact that O'(5) does not take part in hydrogen bonding whereas O(5) does.

Both of the pyranosyl rings have the C1 conformation (Reeves, 1949), with the torsion angles defined by the six ring bonds ranging in magnitude from 48.1 to 62.5° for one ring (unprimed atoms) and from 49.3 to 59.6° for the other. These are normal variations, showing some flexibility of the ring structure.

The nearly identical angles, 114.0 and 114.2°, at the two ring oxygen atoms are in the normal range for such angles in pyranose rings (see, for example, Sundaralingam, 1968). That the angle of 115.8° at O(1) is similarly in the normal range for glycosidic oxygen atoms is a fact of some interest, since, as already noted,  $\alpha$ -trehalose is the first carbohydrate oligomer to have its structure determined which has a bridging oxygen atom axial to two rings. It seems that the demands of the conformational 'anomeric effect' (de Hoog, Buys, Altona & Havinga, 1969; Angyal, 1968) and of hydrogen bonding (see discussion below) are satisfied through adoption by the molecule of conformations about the bonds C(1)-O(1) and C'(1)-O'(1) which produce an

\* In Figs. 1 and 2 the atom labels have been abbreviated by dropping the parentheses used in Table 3 and in the text.

almost maximally extended structure with no interactions tending to increase the angle at O(1).

The torsion angles about C(1)–O(1) and C'(1)–O(1) which are specified in Fig. 1(c) clearly describe for each of the two glucopyranosyl–O(1) groups the *A1* conformer (de Hoog, *et al.*, 1969; Eliel, 1969), which is favored by the anomeric effect and which has often been found (Eliel & Giza, 1968) in crystalline pyranosides. Specifically, the torsion angles 74.8 and 61.7° describe *gauche* or +*syn-clinal* (Klyne & Prelog, 1960)

conformations, and the angles –165 and –177° describe *trans* or *anti-periplanar* conformations.

To specify the relative conformation of the two glucopyranosyl rings, it is better (Rohrer, 1972) to consider pseudotorision angles about the line between C(1) and C'(1) (see Fig. 3) rather than the torsion angles about the two bonds of the glycosidic linkage. The difference of only 6.8° between the pseudotorision angles (C(2)–C(1)···C'(1)–O'(5) and C'(2)–C'(1)···C(1)–O(5) and the general appearance of Fig. 2 show

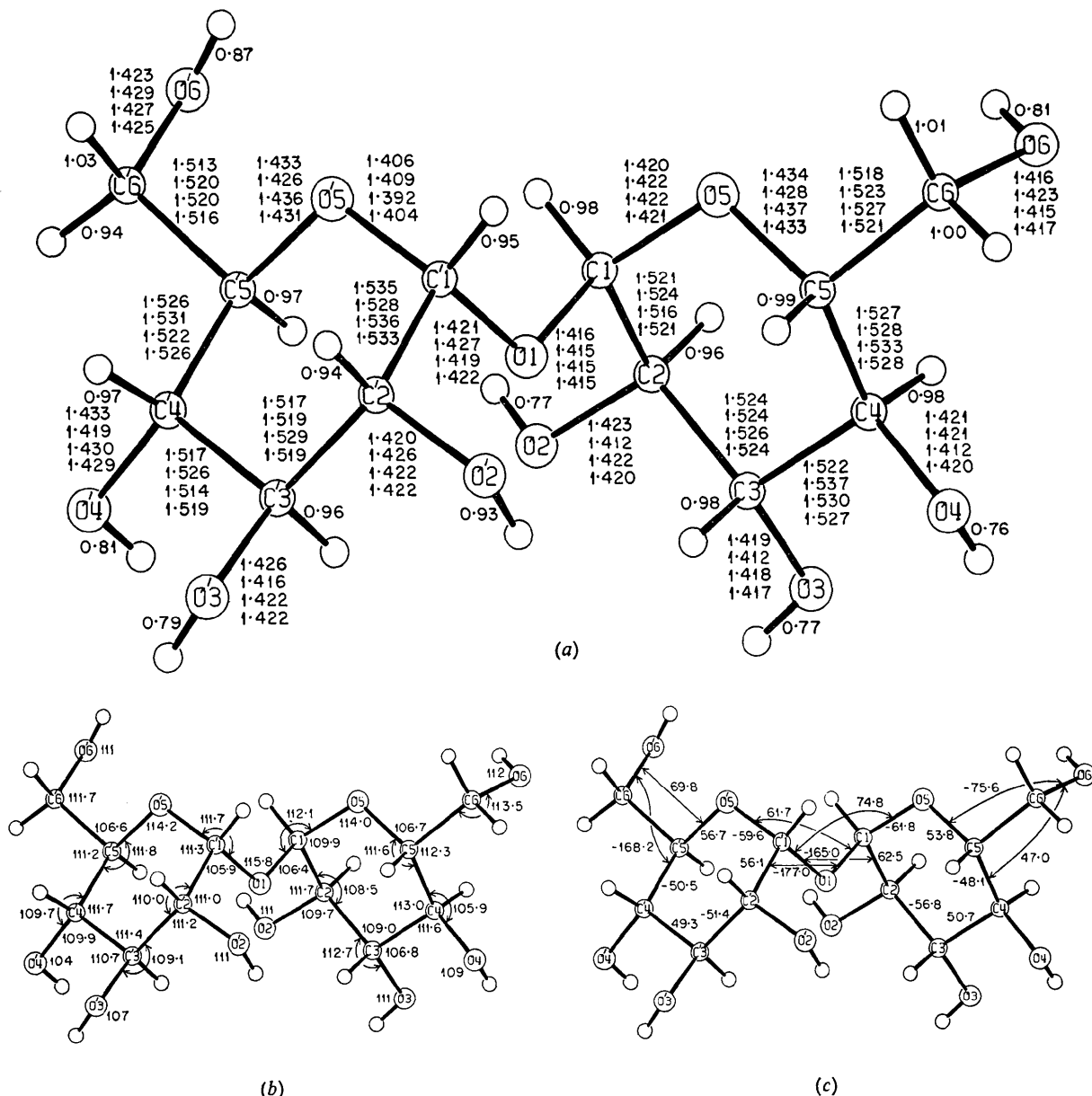


Fig. 1. Bond lengths, valence angles, and torsion angles in the molecule of  $\alpha,\alpha$ -trehalose. (a) For each C–C and C–O bond, the individual lengths from the three determinations and the average length (corresponding to the averaged coordinates) are shown. The order is, from top to bottom: Oak Ridge, Pittsburgh, Edinburgh, average. For the C–H and O–H bonds, only the average values are shown. (b) Average valence angles C–C–C, C–C–O, C–O–C, and C–O–H. (c) Average torsion angles. For each ring bond the torsion angle defined by the two adjacent ring bonds is written next to the bond. Torsion angles about other specific bonds are defined by double-headed arrows pointing in each case to two bonds attached at opposite ends of a central bond.

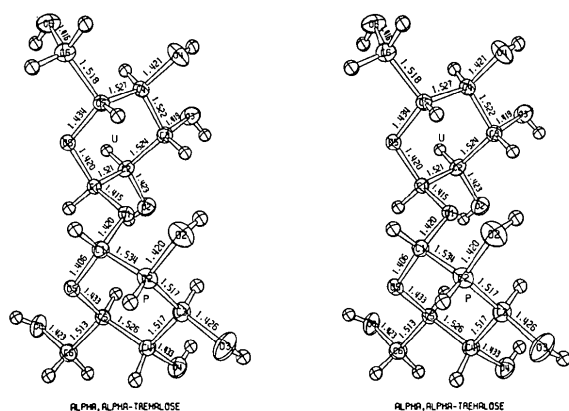


Fig. 2. Stereoscopic drawing of the molecule of  $\alpha, \alpha$ -trehalose based on the Oak Ridge parameters. Carbon and oxygen atoms are represented by their ellipsoids of 50% probability and hydrogen atoms by spheres of an arbitrary radius. The unprimed and primed atoms are indicated by the letter *U* in one ring and the letter *P* in the other.

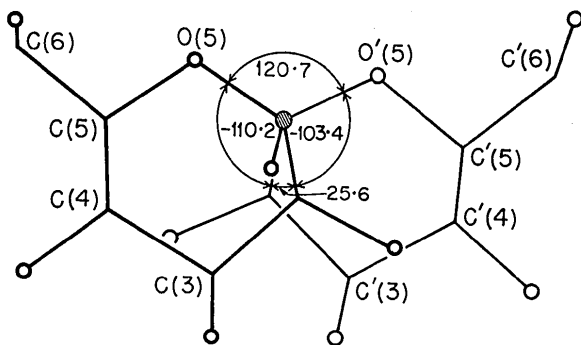


Fig. 3. View of the  $\alpha, \alpha$ -trehalose molecule along the line between  $C(1)$  and  $C'(1)$  and some pseudotorsion angles about this line.

that the departure of the molecule from the symmetry of point group  $C_2$  is small.

The hydroxymethyl groups show the largest conformational difference between the two glucose moieties. The torsion angles about the  $C(5)-C(6)$  bond involving  $O(6)$  with  $O(5)$  and with  $C(4)$  are  $-75.6$  and  $47.0^\circ$ , respectively, describing a *gauche-gauche* arrangement. The corresponding angles about the bond  $C'(5)-C'(6)$  are  $69.8$  and  $-168.2^\circ$ , describing a *gauche-trans* conformation. Both of these conformations have been previously found in crystal structures containing  $\alpha$ -D-glucosyl units (Fries, Rao & Sundaralingam, 1971). A third possible arrangement, *trans-gauche*, is thought to be forbidden because of interaction between  $O(6)$  and  $O(4)$ .

The molecules of  $\alpha, \alpha$ -trehalose and water are held together in the crystal by a complex system of twelve crystallographically independent hydrogen bonds  $O-H \cdots O$  (Table 5 and Figs. 4-6). Each hydroxyl group of the sugar molecule is both a donor and an acceptor in hydrogen-bond formation. The ring oxygen  $O(6)$  is an acceptor, but  $O'(5)$  and  $O(1)$  are not involved in hydrogen bonding. Each of the water molecules is a donor in the formation of two different hydrogen bonds. Molecule  $W(1)$  is an acceptor for two bonds and atom  $O(W1)$  is tetrahedrally coordinated, whereas molecule  $W(2)$  is an acceptor for only one hydrogen bond and atom  $O(W2)$  displays pyramidal coordination.

There is no direct intramolecular hydrogen bonding such as is found in sucrose (Brown & Levy, 1963), cellobiose (Chu & Jeffrey, 1968), methyl- $\beta$ -cellobioside-methanol (Ham & Williams, 1970),  $\beta$ -maltose monohydrate (Quigley, Sarko & Marchessault, 1971), methyl- $\beta$ -maltoside monohydrate (Chu & Jeffrey, 1967), and  $\beta$ -lactose monohydrate (Fries, Rao & Sundaralingam, 1971). However, water molecule  $W(1)$  is

Table 5. Hydrogen bonds in the crystal structure of  $\alpha, \alpha$ -trehalose dihydrate

The symmetry operations are those which generate the coordinates of the acceptor oxygen atoms. For each operation the first digit specifies one of the following: (1)  $x, y, z$ ; (2)  $\frac{1}{2} - x, \bar{y}, \frac{1}{2} + z$ ; (3)  $\frac{1}{2} + x, \frac{1}{2} - y, \bar{z}$ ; (4)  $\bar{x}, \frac{1}{2} + y, \frac{1}{2} - z$ . The last three digits specify a lattice translation.

Hydrogen bond O-H...O	Symmetry operation	Distances (Å)				Angles (°)	
		O-H	H...O	O...O	O-H...O	C-O-H or H-O-H	C-O...O or O...O...O
O(2)—H(O2) ... O(W2)	[1,000]	0.77	1.97	2.732	172	111	112.8
O(3)—H(O3) ... O'(3)	[2,00 $\bar{1}$ ]	0.77	2.01	2.750	159	111	125.7
O(4)—H(O4) ... O(5)	[3,001]	0.76	2.23	2.882	144	109	122.1
O(6)—H(O6) ... O'(4)	[4,000]	0.81	1.94	2.728	165	112	121.7
O'(2)—H(O'2) ... O(W1)	[2,000]	0.93	1.87	2.770	163	111	121.5
O'(3)—H(O'3) ... O(4)	[2,00 $\bar{1}$ ]	0.79	2.12	2.906	176	107	108.2
O'(4)—H(O'4) ... O(W1)	[1,000]	0.81	2.09	2.882	169	104	99.9
O'(6)—H(O'6) ... O(6)	[3, $\bar{1}$ 01]	0.87	1.83	2.704	174	111	115.3
O(W1)—H(W1) ... O(2)	[1,000]	0.83	1.94	2.766	177		
O(W1)—H'(W1) ... O'(2)	[1,00 $\bar{1}$ ]	0.79	1.92	2.697	169	106	108.6
O(W2)—H(W2) ... O(3)	[3, $\bar{1}$ 00]	0.76	2.05	2.803	170		
O(W2)—H'(W2) ... O'(6)	[1,000]	0.82	1.96	2.764	168	107	103.4
		0.04	0.04	0.003	4	3	0.1
		to	to	to	to	to	to
	Standard errors	0.05	0.05	0.004	5	5	0.2

linked by hydrogen bonding both to O(2) and to O'(4), and molecule *W*(2) is linked both to O(2) and to O'(6), all within the same trehalose molecule. The same kind of indirect intramolecular hydrogen bonding through water molecules occurs in methyl- $\beta$ -maltoside monohydrate, planteose dihydrate (Rohrer, 1972) and raffinose pentahydrate (Berman, 1970). Direct intramolecular hydrogen bonding is less likely in crystals of oligosaccharides as the number of water molecules per monosaccharide moiety increases.

A particularly interesting feature of the hydrogen bonding pattern is the bonding involving *W*(1) and O'(2)H. This bonding produces about the axis  $2_1$  at  $\frac{1}{4}, 0, z$  the helix of hydrogen bonds shown schematically in Fig. 5 and also shown rather clearly between sugar molecules 6 and 7 in the stereoscopic drawing of Fig. 6. The helix around the  $2_1$  axis at  $\frac{3}{4}, \frac{1}{2}, z$  is identical, but its vectors  $O(W1) \rightarrow H'(W1)$  point in the opposite direction along the *c* axis. It is possible that because of these helices the crystal may show an abnormally high electric polarizability in the *c* direction and that in some temperature range it may show ferroelectric behavior. Thus, if in half the helices the hydroxyl O'(2)H were to reorient by rotating  $\sim 180^\circ$  around bond C'(2)-O'(2) and if at the same time the water molecule *W*(1) were to rotate  $\sim 120^\circ$  around the bond O(*W*1)-H(*W*1), possibly with some additional minor adjustments in the orientation of *W*(1), the result would be a highly polar structure of symmetry  $P2_1$  with about the same stability as the  $P2_12_12_1$  structure. From the X-ray analysis of Beevers & Hughes (1941) and the neutron analysis of Frazer (1962) (see also Hamilton & Ibers, 1968), a similar mechanism is thought to explain the ferroelectric behavior of Rochelle salt,  $\text{NaK}(\text{C}_4\text{H}_4\text{O}_6) \cdot 4\text{H}_2\text{O}$ .

Another interesting feature of the hydrogen bonding is the spatial disposition of the bond O(4)-H(O4)  $\cdots$  O(5) relative to the C(1)-O(5)-C(5) group of atoms. In nearly every previously known case (Sundaralingam, 1968) the hydrogen bond to the ring oxygen of a pyranose system is equatorially oriented at the acceptor oxygen. In the present case, as can be seen in Fig. 6, the orientation is neither equatorial nor axial, but about halfway between; that is, the configuration of atoms C(1), O(5), C(5), and H(O4) is very nearly planar. The angles C(1)-O(5)-H(O4) and C(5)-O(5)-H(O4) are  $104(1)$  and  $141(1)^\circ$ , and the sum of angles about O(5) is  $359^\circ$ .

### References

- ANGYAL, S. J. (1968). *Aust. J. Chem.* **21**, 2737.  
 BEEVERS, C. A. & HUGHES, W. (1941). *Proc. Roy. Soc. A* **177**, 251.  
 BERGHUIS, J., HAANAPPEL, IJ. M., POTTERS, M., LOOPSTRA, B. O., MACGILLAVRY, C. H. & VEENENDAAL, A. L. (1955). *Acta Cryst.* **8**, 478.  
 BERMAN, H. M. (1970). *Acta Cryst.* **B26**, 290.  
 BERMAN, H. M., CHU, S. S. C. & JEFFREY, G. A. (1967). *Science*, **157**, 1576.

- BIRCH, G. G. (1963). *Advanc. Carbohydrate Chem.* **18**, 201.  
 BROWN, G. M. (1969). *Acta Cryst.* **B25**, 1338.  
 BROWN, G. M. & LEVY, H. A. (1963). *Science*, **141**, 921.  
 BUSING, W. R. (1970). In *Crystallographic Computing*, p. 319. Edited by F. R. AHMED. Copenhagen: Munksgaard.

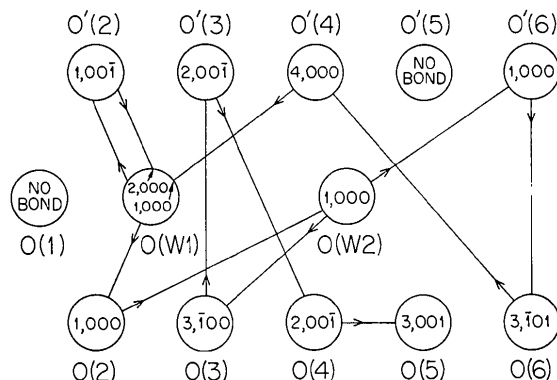


Fig. 4. Schematic diagram of hydrogen bonding in  $\alpha,\alpha$ -trehalose dihydrate. The large circles represent oxygen atoms; the arrows denote directions of donation of hydrogen atoms. Each oxygen atom functioning as a donor is to be understood to belong to the reference asymmetric unit (symmetry transformation [1,000] in the notation of Table 5). The symmetry transformation for each acceptor oxygen is indicated.

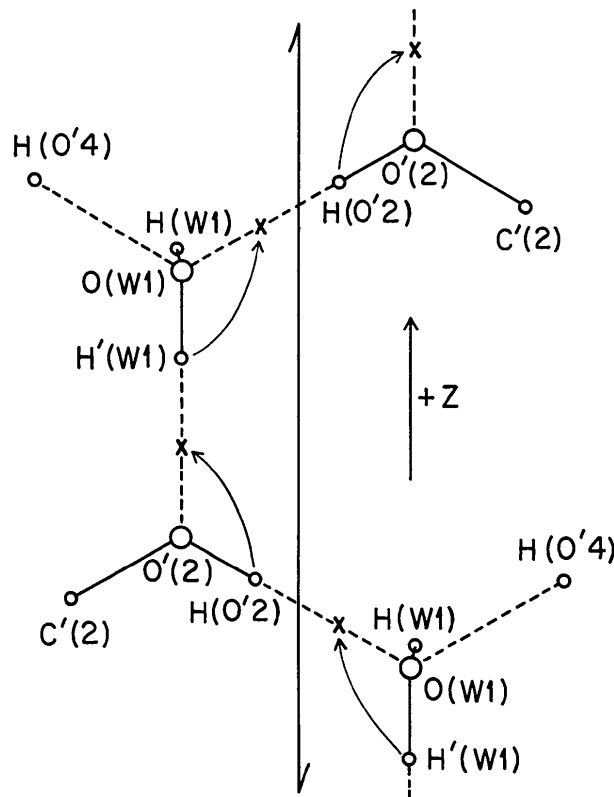


Fig. 5. Schematic drawing of helical arrangement of hydrogen bonds about axis  $\frac{1}{4}, 0, z$ .

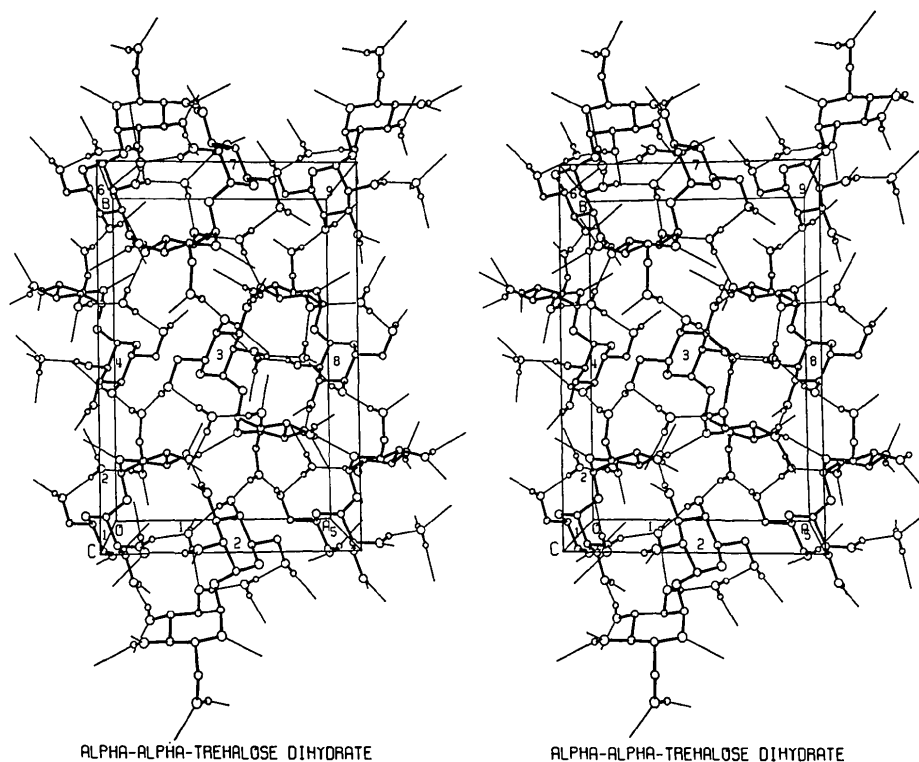


Fig. 6. Stereoscopic view of the packing of the sugar and water molecules in the crystal structure of  $\alpha,\alpha$ -trehalose dihydrate. Hydrogen atoms attached to carbon atoms are omitted. The thin lines connecting hydrogen atoms and oxygen atoms show the hydrogen bonding. Within the ring of primed atoms of each sugar molecule is a numerical label which identifies the symmetry operation generating that molecule according to the scheme used in Table 5: 1=[1,000]; 2=[2,000]; 3=[3,001]; 4=[4,000]; 5=[1,100]; 6=[1,010]; 7=[2,010]; 8=[4,100]; 9=[1,110]. Two independent water molecules are labelled 1 and 2. Note, however, that by inadvertence water molecule 1 here is not the molecule  $W(1)$  whose coordinates are given in Table 3 but rather the molecule  $W(1)$  [2,000].

- BUSING, W. R., ELLISON, R. D., LEVY, H. A., KING, S. P. & ROSEBERRY, R. T. (1968). *The Oak Ridge Computer-Controlled X-ray Diffractometer*. Report ORNL-4143, Oak Ridge National Laboratory, Oak Ridge, Tennessee.
- BUSING, W. R. & LEVY, H. A. (1967). *Acta Cryst.* **22**, 457.
- CHU, S. S. C. & JEFFREY, G. A. (1967). *Acta Cryst.* **B23**, 1038.
- CHU, S. S. C. & JEFFREY, G. A. (1968). *Acta Cryst.* **B24**, 830.
- CROMER, D. T. & WABER, J. T. (1965). *Acta Cryst.* **18**, 104.
- DREW, M. G. B. (1969). *Acta Cryst.* **B25**, 1320.
- DREW, M. G. B., TEMPLETON, D. H. & ZALKIN, A. (1969). *Acta Cryst.* **B25**, 261.
- ELIEL, E. L. (1969). *Svensk Kem. Tidskr.* NR/6-7, p. 22.
- ELIEL, E. L. & GIZA, C. A. (1968). *J. Org. Chem.* **33**, 3754.
- FRAZER, B. C. (1962). *J. Phys. Soc. Japan*, **17**, supp. B-II, p. 376.
- FREEMAN, A. J. (1959). *Acta Cryst.* **12**, 261.
- FRIES, D. C., RAO, S. T. & SUNDARALINGAM, M. (1971). *Acta Cryst.* **B27**, 994.
- GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* **B26**, 274.
- GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1971). *Acta Cryst.* **A27**, 368.
- GROTH, P. (1910). *Chemische Kristallographie*, Part 3, p. 450. Leipzig: Engelmann.
- HALL, S. R. (1967). *Direct Phasing Method, UWAC-17*. The University of Western Australia. Revised for IBM 7090 by H. M. BERMAN.
- HALL, S. R. (1970). In *Crystallographic Computing*, p. 66. Edited by F. R. AHMED. Copenhagen: Munksgaard.
- HAM, J. T. & WILLIAMS, D. G. (1970). *Acta Cryst.* **B26**, 1373.
- HAMILTON, W. C. (1965). *Trans. ACA*, **1**, 17.
- HAMILTON, W. C. & ABRAHAMS, S. C. (1970). *Acta Cryst.* **A26**, 18.
- HAMILTON, W. C. & IBERS, J. A. (1968). *Hydrogen Bonding in Solids*, p. 241. New York: Benjamin.
- HANSON, H. P., HERMAN, F., LEA, J. D. & SKILLMAN, S. (1964). *Acta Cryst.* **17**, 1040.
- HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal*. ACA Monograph No. 3. Pittsburgh: Polycrystal Book Service.
- HAUPTMAN, H. & KARLE, J. (1956). *Acta Cryst.* **9**, 45.
- HOOG, A. J. DE, BUYS, H. R., ALTONA, C. & HAVINGA, E. (1969). *Tetrahedron*, **25**, 3365.
- International Tables for X-ray Crystallography* (1959). Vol. II. Birmingham: Kynoch Press.
- JOHNSON, C. K. (1970). *ORTEP*. Report ORNL-3794 (2nd. rev.), Oak Ridge National Laboratory, Oak Ridge, Tennessee.
- KALF, G. F. & RIEDER, S. V. (1958). *J. Biol. Chem.* **230**, 691.
- KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.
- KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849.
- KLYNE, W. & PRELOG, V. (1960). *Experientia*, **16**, 521.
- LEVY, H. A. (1966). Unpublished.
- MCWEENY, R. (1951). *Acta Cryst.* **4**, 513.

- MAIN, P., WOOLFSON, M. M. & GERMAIN, G. (1971). *MULTAN, A Computer Program for the Automatic Solution of Crystal Structures*.
- PETERSON, S. W. & LEVY, H. A. (1957). *Acta Cryst.* **10**, 70.
- QUIGLY, G. J., SARKO, A. & MARCHESSAULT, R. H. (1971). Private communication.
- REEVES, R. E. (1949). *J. Amer. Chem. Soc.* **71**, 215.
- ROHRER, D. C. (1972). *Acta Cryst.* **B28**, 425.
- SHONO, R. (1971). *Technical Report*, TR-71-2. Department of Crystallography, University of Pittsburgh, Pittsburgh, Pennsylvania.
- STEWART, R. F., DAVIDSON, E. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 3175.
- STRAHS, G. (1970). *Advanc. Carbohydrate Chem.* **25**, 53.
- SUNDARALINGAM, M. (1968). *Biopolymers*, **6**, 189.
- THIESSEN, W. E., LEVY, H. A. & BROWN, G. M. (1970a). *Abstracts*, ACA Winter Meeting, New Orleans, Louisiana, p. 22.
- THIESSEN, W. E., LEVY, H. A. & BROWN, G. M. (1970b). *Chem. Div. Annual Report*, pp. 135. ORNL Report 4581, Oak Ridge National Laboratory, Oak Ridge, Tennessee.
- WYATT, G. R. & KALF, G. F. (1957). *J. Gen. Physiol.* **40**, 833.

*Acta Cryst.* (1972). **B28**, 3158

## Packing Analysis of Carbohydrates and Polysaccharides. I. Monosaccharides

BY P. ZUGENMAIER AND A. SARKO

*Department of Chemistry, State University College of Forestry, Syracuse, New York 13210, U.S.A.*

(Received 18 November 1971 and in revised form 1 July 1972)

The minimum energy crystal structures of six monosaccharides:  $\alpha$ - and  $\beta$ -D-glucose,  $\beta$ -lyxose, methyl- $\alpha$ -D-glucopyranoside, D-glucono-(1,5)-lactone, and  $\beta$ -arabinose, were calculated using a modified Williams procedure of minimization of nonbonded repulsion energy. In all cases but  $\beta$ -arabinose, the predicted structures were in close agreement with the real structures, and could easily serve as starting models for least-squares refinement procedures. The same potential energy parameter set was applicable in all cases. In the case of  $\beta$ -arabinose a structure was predicted which differed from the reported one by a shift of  $-\frac{1}{4}a$ . Subsequent refinement of the predicted structure with X-ray data revealed errors in the reported data and produced a corrected crystal structure and good agreement with the predicted one. Of all the nonbonded interactions present in these carbohydrates, those involving the hydrogens were the most important and those involving the carbons the least important. Little difficulty with predicted structures in false energy minima was encountered.

### Introduction

The determination of polymer crystal structure is, in general, handicapped by a lack of observed diffraction data, necessitating reliance on information obtained from other sources. In principle, such information may be obtained from a computation of minimum potential energy structures which could serve as models for subsequent structure refinement with available diffraction data. This goal has not yet been achieved, but nevertheless, such calculations have been successful with low molecular weight hydrocarbons (Williams, 1969; Neuman, 1970; Warshel & Lifson, 1970) and, to some extent, with hydrocarbon polymers (Clark & Geering, 1970; Wobser & Blasenbrey, 1970). These calculations have been carried out in two different ways, either by minimization of the total potential energy or by the minimization of repulsive energy only (Williams, 1965, 1969). Calculations using the second approach have the advantage of being simpler and faster and appear to be sufficiently reliable.

The extension of the same techniques to the prediction of crystal structures of carbohydrates is somewhat more difficult owing to the presence of oxygens. First, at least three more nonbonded potential energy func-

tions must be introduced. Second, hydrogen bonding must now be considered.

As part of work with biologically important polysaccharides, it was decided to investigate in detail the forces responsible for the conformation of these molecules and their packing into crystal structures. A suitable starting point seemed to be the attempt to predict the known crystal structures of carbohydrates of low molecular weight, in an effort to determine the best set of potential energy functions applicable to all such molecules. Calculations with the same potential parameter set for other, as yet undetermined carbohydrate structures might then provide reliable predictions of their crystal structures. In this communication we describe the calculation of the most probable crystal structures of  $\alpha$ -D-glucose,  $\beta$ -D-glucose,  $\beta$ -lyxose, methyl- $\alpha$ -D-glucopyranose, D-glucono-(1,5)-lactone, and  $\beta$ -arabinose, using the method of minimization of repulsive nonbonded energies only.

### Method of calculation

The method employed was that of Williams (1969) and consisted, basically, in moving a rigid molecule and its symmetry-related mates within the fixed unit cell until