

- 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

a position of minimum energy was found. The success of the calculations could be judged in three ways:

1. When starting the computation with the molecules in the actual positions in the unit cell, the position of minimum repulsion should be identical or should occur in close vicinity.

2. When starting the computation with the molecules in arbitrary positions in the unit cell, the final position should correspond to the crystal structure. With several energy minima occurring, the correct one should be the lowest.

3. The same potential energy function parameters should apply to different but related structures.

The Williams *PACK5* computer program, designed to search for the minimum energy position with the help of least-squares procedures, was used throughout this work. The program was modified for the IBM 1620-II computer.

The interatomic nonbonded repulsive potential energy of a pair of atoms ( $i, j$ ),  $\varepsilon_r$ , was approximated by (Williams, 1969):

$$\varepsilon_r = w(d_o - d_{ij})^2, \quad d_{ij} < d_o \quad (1a)$$

$$\varepsilon_r = 0, \quad d_{ij} > d_o \quad (1b)$$

where  $d_{ij}$  is the interatomic distance,  $d_o$  is an empirical cut-off distance, and  $w$  is a weighting factor. All pairs of atoms of one kind had the same  $d_o$  and  $w$ , meaning that all such atoms were treated equally and polarization effects were not taken into account.

The parameters  $d_o$  and  $w$  for the five nonbonded interactions O---O, O---C, O---H, C---H and H---H were obtained by trial and error and are shown in Table 1. In this manner, it was found that only the H---H, C---H and O---H contacts were important in determining the energy minima of carbohydrates. The most consistent results were obtained with the listed parameters for these three interactions. The remaining contacts were not balanced and are therefore somewhat arbitrary. No C---C contact less than 3.7 Å was found. The cut-off distances for H---H and C---H determined in this fashion were in good agreement with those most recently reported by Williams (1970). In addition to the nonbonded interactions, the possible presence of a hydrogen bond was taken into account as follows. An O---O contact was considered to be a hydrogen bond when the interatomic distance was less than 3.0 Å and greater than 2.5 Å and the corresponding oxygen-hydrogen distance O---H---O was less than 2.5 Å. The applicable  $d_o$  and  $w$  parameters are also listed in Table 1. It was found expedient not to use equation (1b) for the hydrogen bond; rather equation (1a) was used for all values of  $d_{ij}$ . This procedure had the effect of forcing the hydrogen bond towards the length  $d_o$  as it simulated an attractive hydrogen-bonding force. The attractive potential was, however, always negligible compared with the total repulsion energy. The influence of H---H interaction ( $d_{ij} < 2.5$  Å) between hydroxyl hydrogens involved in a hydrogen bond was found to be insignificant and this interaction was

dropped in the hydrogen bond calculation. This constituted an advantage because the hydroxyl hydrogens are generally not easily located due to possible rotations about the C-O axis.

Table 1. *Parameter set used for the calculation of repulsion energy*

Interaction	Cut-off distance $d_o$ (Å)	Scaling factor $w$
O---O	3.40	1.25
O---C	3.40	1.00
O---H	3.25	1.80
C---H	3.30	1.35
H---H	3.20	0.50
O---O (H-bond)	2.80	10.00

All molecules were considered to be rigid because the limited computer core available did not permit incorporation of substituent rotations into the program. This is a serious limitation and must be removed before these calculations are extended to more complex carbohydrates such as dimers, oligomers or polysaccharides.

As a measure of the deviation between actual and predicted atomic coordinates, an average deviation,  $\varphi$ , was calculated for each structure:

$$\varphi = \left( \sum_{i=1}^N \Delta_i^2 / 3N \right)^{1/2} \quad (2)$$

where  $N$  is the number of atoms in the molecule and  $\Delta_i$  is the difference between corresponding atom positions in the predicted and actual structures. For each actual and predicted structure the total repulsive energy,  $E_r$ , was calculated by summing over all pairwise energies  $\varepsilon_r$ .

It was expected that starting the calculations with the molecules in arbitrary and random starting positions, a number of false energy minima in addition to the correct minimum would be found. The following strategy was finally developed for searching out the correct minimum-energy position.

Random starting positions were generated by translations and rotations of the molecule relative to some fixed point. Initial calculations from these positions were carried out by omitting either all O---H or all H---H interactions. This resulted in a much faster calculation, and more importantly, in far fewer minimum energy positions when compared with a full calculation including all interactions. Each minimum thus reached constituted a new starting position for a subsequent full calculation. In most cases, all of the false minima were easily eliminated from consideration in this fashion; however, in a few instances the choices were more difficult. Three criteria were finally selected for choosing the 'correct' minimum positions:

1. The correct structure should have a 'very low' repulsive packing energy.

2. No unreasonably short contacts should occur and the closest contacts should not decrease to unreasonable values when the potential parameter set was slightly changed.

3. The number of hydrogen bonds must be maximized.

In the following discussion, the predicted position of the molecule in the unit cell is reported in terms of the rotation matrix  $\mathbf{R}$  and the translation vector  $\mathbf{S}$ , relative to the rotation point  $\mathbf{P}$ . The latter is given in ångströms relative to the origin of the space group. An identity matrix for  $\mathbf{R}$  and a zero vector for  $\mathbf{S}$  indicate perfect coincidence with the actual structure.

## Results and discussion

### 1. $\alpha$ -D-Glucose

The crystal structure of  $\alpha$ -D-glucose was a most useful model structure for the calculations because its hydrogen positions were accurately determined by neutron diffraction (Brown & Levy, 1965). The reported C-H bond distances varied from 1.092 Å to 1.106 Å and the O-H distances from 0.964 to 0.973 Å. These distances are approximately 0.15–0.20 Å longer than the corresponding distances in another carbohydrate, D-glucono-(1,5)-lactone, for which the hydrogen positions are also accurately known. As will be shown throughout this work, the hydrogen coordinates are most important in determining the minimum energy structure, as well as the value of the total repulsive energy  $E_r$ , and a comparison of the two carbohydrates has proved to be most informative.

The final predicted structure of  $\alpha$ -D-glucose was almost identical with the actual structure, as shown by the rotation matrix  $\mathbf{R}$ , the vector  $\mathbf{S}$  and the deviation  $\varphi = 0.0325$  Å with all atoms included in the calculation.

$$\mathbf{R} = \begin{pmatrix} 0.99999 & -0.00126 & 0.00221 \\ 0.00125 & 0.99999 & 0.00379 \\ -0.00222 & 0.00378 & 0.99999 \end{pmatrix}$$

$$\mathbf{P} = (4.0, 8.0, 0.0) \text{ Å}$$

$$\mathbf{S} = (-0.02, -0.05, -0.02) \text{ Å}$$

For carbon and oxygen atoms only the value of  $\varphi = 0.0327$  Å, insignificantly different from that with all atoms included. Little change in deviation was also found when the function parameters shown in Table 1 were changed; for example, the value of  $\varphi$  for all atoms decreased to 0.0314 Å when the cut-off distance for H---H was increased to 3.30 Å. It decreased by a similar amount when the corresponding O---H parameter was changed. However, such increases in these parameters for general use were not warranted because it was subsequently found that they did not produce improvements in all structures.

Drastic changes in  $\varphi$  did not occur even when one of the main interactions, such as H---H, H---C, or H---O, was omitted. The corresponding  $\varphi$  values were  $\varphi_{(\text{H} \cdots \text{H})} = 0.0345$  Å,  $\varphi_{(\text{H} \cdots \text{O})} = 0.0935$  Å, and  $\varphi_{(\text{H} \cdots \text{C})} = 0.1116$  Å. Even with the weighting term for the H---O contacts and the potential term for the hydrogen bond both set equal to zero, a good fit for the calculated structure was found. But now the previously stable minimum of repulsive energy was no longer stable and the calculated structure began to oscillate about the minimum, forcing an arbitrary termination of the computation. These results indicated that the packing of the structure was not extremely sensitive to the type of functions used to describe the repulsive forces and, perhaps more importantly, that the minimum energy position was determined by a relatively small number of nonbonded interactions.

The breakdown of short contacts by type in both the actual and the calculated structure is shown in Table 2, along with the corresponding repulsive energies. The O---O and O---C contacts were of no importance either in number or in repulsive energy. The main contribution to the repulsive energy arose from the O---H, C---H and H---H interactions which almost dupli-

Table 2. Distribution of short contacts in actual and computed crystal structures of  $\alpha$ -D-glucose

All hydrogen bonds are included.

Interaction	Total number of repulsive interactions (and H-bonds)	$E_r$ for observed structure	Atoms*	Closest contacts (Å)	
				Obs.	Pred.
O---O	4	0.04	O(1)---O(5)	3.28	3.28
O---C	10	0.16	C(6)---O(2)	3.22	3.27
O---H	36	8.98	O(1)---H(2)	2.29	2.30
			H(6B)---O(2)	2.73	2.77
			H(1')---O(6)	2.80	2.83
			O(1)---H(4)	2.82	2.83
C---H	24	5.22	C(6)---H(2')	2.67	2.64
			C(2)---H(3')	2.72	2.75
			C(4)---H(4')	2.73	2.72
			H(1')---C(1)	2.82	2.72
H---H	48	5.52	H(1)---H(6B)	2.37	2.33
			H(5)---H(4)	2.44	2.45
			H(5)---H(4')	2.44	2.47
			H(4')---H(4')	2.49	2.49
O---O (H-bond)	10	0.40	O(2)---O(3)	2.70	2.70
			O(6)---O(3)	2.71	2.73
			O(4)---O(4)	2.77	2.75
			O(6)---O(2)	2.77	2.76
			O(1)---O(5)	2.84	2.75

\* The hydroxyl hydrogens are denoted by a prime and the same number as the hydroxyl oxygen.

cated those found in the actual structure, including a rather close O---H contact of 2.29 Å which remained nearly unchanged in the calculated structure. The five unique hydrogen bonds are also listed in Table 2 (the other five are symmetry related) and as can be seen, the predicted bond lengths are, with one exception, close to the bond lengths found in the actual structure. The significance of the one exception (difference of 0.09 Å) can at present not be estimated. Only minor changes in the fit of the calculated structure were observed when the scaling factor for the hydrogen bond was increased by a factor of 2.

For the actual structure of  $\alpha$ -D-glucose,  $E_r=20.32$  (in arbitrary units), which included an attractive part of 0.04 for the hydrogen bond. For the computed structure,  $E_r=20.01$ . The agreement between the two values further points out the small differences between the actual and computed structures.

## 2. $\beta$ -D-Glucose

The crystal structure of  $\beta$ -D-glucose had been determined by X-ray diffraction (Chu & Jeffrey, 1968) but

without refinement of hydrogen positions. The predicted structure occurred at

$$\mathbf{R} = \begin{pmatrix} 0.99958 & -0.10404 & -0.02529 \\ 0.01471 & 0.99954 & 0.02637 \\ 0.02491 & -0.02673 & 0.99933 \end{pmatrix}$$

$$\mathbf{P} = (1.0, 2.0, 2.0) \text{ \AA}$$

$$\mathbf{S} = (0.01, 0.08, -0.05) \text{ \AA}$$

and was characterized by  $\varphi=0.0653$  Å with all atoms included in the calculation. With hydrogens excluded,  $\varphi=0.0617$  Å. By increasing the cut-off distance for H---H repulsion to 3.30 Å,  $\varphi$  decreased to 0.0644 Å for all atoms.

The breakdown of the repulsive energies and the closest contacts for both the actual and the computed structure are listed in Table 3. As was the case with  $\alpha$ -D-glucose, the predominant interactions were O---H, C---H, and H---H, with O---O and O---C interactions contributing insignificantly to the total repulsive energy. Two very short contacts in the actual crystal structure, an O---H contact of 2.32 Å and an H---H contact of 2.06 Å, were preserved in the predicted structure.

Table 3. *Distribution of short contacts in actual and computed crystal structures of  $\beta$ -D-glucose*

All hydrogen bonds are included.

(a) Real minimum		$E_r$ for observed structure	Closest contacts (Å)		
Interaction	Total number of repulsive interactions (and H-bonds)		Atoms*	Obs.	Pred.
O---O	2	0.04	O(4)---O(2)	3.26	3.37
O---C	2	0.02	O(3)---C(2)	3.29	—
O---H	40	10.23	H(4')---O(2)	2.32	2.45
			O(4)---H(5)	2.54	2.56
			O(1)---H(6A)	2.70	2.81
			O(4)---H(3)	2.79	2.62
C---H	26	4.14	C(3)---H(2')	2.68	2.77
			C(6)---H(1')	2.82	2.95
			C(2)---H(2')	2.87	2.96
			H(3')---C(1)	2.90	2.81
H---H	48	5.77	H(4)---H(6A)	2.06	2.03
			H(2)---H(2')	2.36	2.46
			H(4')---H(5)	2.44	2.37
			H(4')---H(6')	2.46	2.51
O---O (H-bond)	8	0.79	O(6)---O(1)	2.66	2.71
			O(3)---O(2)	2.68	2.74
			O(2)---O(6)	2.70	2.81
			O(3)---O(5)	2.77	2.72
(b) False minimum					
O---O	8	0.39	O(4)---O(5)		3.00
			O(2)---O(4)		3.38
			O(3)---O(1)		3.38
O---C	2	0.00	C(3)---O(1)		3.37
O---H	38	10.97	H(3)---O(1)		2.37
			O(2)---H(4')		2.42
			O(4)---H(2)		2.70
			H(6A)---O(6)		2.71
C---H	22	1.80	H(2')---C(6)		2.82
			H(1')---C(6)		2.92
			C(1)---H(4')		3.00
H---H	50	6.66	H(6B)---H(2)		2.10
			H(3')---H(4')		2.12
			H(3)---H(1')		2.33
			H(1)---H(4')		2.36
O---O (H-bond)	6	0.10	O(2)---O(6)		2.79
			O(1)---O(6)		2.81
			O(2)---O(3)		2.86

\* See footnote at the bottom of Table 2.

In general, it was apparent that the differences between the closest contacts of the actual and the computed structure were larger in  $\beta$ -D-glucose than in  $\alpha$ -D-glucose. It is likely that this was due to the less accurate positions of the hydrogens in the former structure. The total repulsive energies for  $\beta$ -D-glucose were similar to those for the  $\alpha$  anomer:  $E_r=21.00$  for the actual and 19.40 for the completed structure.

An interesting pseudo-minimum was found in the course of calculation of  $\beta$ -D-glucose. The pseudo-minimum occurred at

$$\mathbf{R} = \begin{pmatrix} -0.8686 & 0.4955 & 0.0051 \\ 0.4950 & 0.8681 & -0.0367 \\ -0.0226 & -0.0293 & -0.9993 \end{pmatrix}$$

$$\mathbf{S} = (0.85, 1.88, -0.50) \text{ \AA}$$

The closest contacts and the corresponding repulsive energies for the resulting structure are also listed in Table 3. The  $E_r=19.97$  for this structure was only slightly higher than that for the 'correct' calculated structure, 19.40. The closest contacts were, in general, no worse than for actual structure, as can be seen from Table 3. But only six hydrogen bonds to the asymmetric unit were found, compared with eight in the lower energy structure, and the hydrogen-bonding scheme in the latter appeared considerably more reasonable.

### 3. $\beta$ -Lyxose

As an example of a pentose, the minimum energy structure of  $\beta$ -lyxose was computed. Its crystal structure had been determined by X-ray analysis, without refinement of hydrogen positions (Hordvik, 1966). The predicted structure was characterized by a relatively large deviation  $\varphi=0.0920 \text{ \AA}$  ( $\varphi=0.0901 \text{ \AA}$  with hydrogens excluded) which was mostly due to the apparently systematic shift in the  $z$  coordinate of all atoms by approximately  $-0.12 \text{ \AA}$  as shown by the translation vector  $\mathbf{S}$ .

$$\mathbf{R} = \begin{pmatrix} 0.99917 & -0.04035 & 0.00383 \\ 0.04036 & 0.99918 & -0.00225 \\ -0.00374 & 0.00240 & 0.99999 \end{pmatrix}$$

$$\mathbf{P} = (3.0, 3.0, 0.00) \text{ \AA}$$

$$\mathbf{S} = (-0.01, -0.10, -0.12) \text{ \AA}$$

The individual short contacts and the corresponding repulsive energies for both actual and computed structure of lyxose are shown in Table 4. No C---H contacts below  $2.96 \text{ \AA}$  were observed. No very short O---H contacts were found either, yet this interaction maintained its importance in determining the repulsive energy. Due to the shift in the  $z$  coordinates for the calculated structure, larger deviations in the closest contacts between the calculated and actual structure were observed, especially in the hydrogen bond distances. The effect of the  $z$  shift was also seen in the drop of  $E_r=12.67$  for the actual structure to 11.74 for the calculated one.

### 4. Methyl $\alpha$ -D-glucopyranose

The crystal structure of methyl  $\alpha$ -D-glucopyranose (Berman & Kim, 1968) was of interest largely because of the presence of the methyl substituent. The computation was performed with a rigid molecule. Although relatively large errors in the hydrogen positions were present, the results were excellent, as shown by  $\mathbf{R}$ ,  $\mathbf{S}$  and the small deviations  $\varphi=0.0574 \text{ \AA}$  with all atoms included, and  $\varphi=0.0583 \text{ \AA}$  for the oxygen and carbon atoms only.

$$\mathbf{R} = \begin{pmatrix} 0.99997 & -0.00316 & 0.00572 \\ 0.00306 & 0.99983 & 0.01802 \\ -0.00577 & 0.01801 & 0.99982 \end{pmatrix}$$

$$\mathbf{P} = (5.5, 4.5, 3.0) \text{ \AA}$$

$$\mathbf{S} = (0.07, -0.03, 0.05) \text{ \AA}$$

The repulsive energies and the closest contacts are listed in Table 5. In this structure, the H---H interactions were the most significant, followed by the O---H

Table 4. Distribution of short contacts in actual and computed crystal structures of  $\beta$ -lyxose

All hydrogen bonds are included

Interactions	Total number of repulsive interactions (and H-bonds)	$E_r$ for observed structure	Closest contacts ( $\text{\AA}$ )		
			Atoms*	Obs.	Pred.
O---O	2	0.05	O(2)---O(4)	3.26	3.24
O---C	2	0.00	O(5)---C(4)	3.38	—
O---H	38	7.96	O(1)---H(5B)	2.68	2.66
			H(2')---O(4)	2.70	2.68
			H(5A)---O(3)	2.71	2.87
			H(3)---O(2)	2.72	2.75
C---H	14	0.98	C(1)---H(4')	2.96	3.06
			C(5)---H(4')	3.03	2.97
H---H	38	2.97	H(1)---H(4)	2.26	2.32
			H(3')---H(3)	2.59	2.68
			H(1')---H(5A)	2.66	2.65
O---O (H-bond)	8	0.75	O(5)---O(4)	2.69	2.72
			O(3)---O(4)	2.76	2.71
			O(1)---O(3)	2.81	2.67
			O(1)---O(2)	2.96	2.85

\* See footnote at the bottom of Table 2.

interaction. The closest contacts were comparable to those of  $\beta$ -lyxose rather than to those found in either  $\alpha$ - or  $\beta$ -D-glucose. No C---O contact less than 3.40 Å was found and only two O---O contacts with 3.25 Å were below the cut-off distance.

Many numerically larger differences in the closest contacts between the actual and the computed structure were observed in this carbohydrate in comparison with those previously discussed. This was especially true for the hydrogen bond. Nonetheless, for a total of 130 repulsive contacts, the repulsive packing energies were reasonably low:  $E_r = 15.28$  for the actual structure and 14.40 for the computed structure.

### 5. D-Glucono-(1,5)-lactone

The recently reported structure of D-glucono-(1,5)-lactone (Hackert & Jacobson, 1971) presented a good test for the computation of carbohydrate structures in several ways. First, the hydrogen positions were refined in the X-ray analysis. Secondly, the C-H and O-H bond lengths were much shorter and more uniform in this structure when compared with other carbohydrates (0.91–0.96 Å for C-H, and 0.71–0.78 Å for O-H). Finally, there were some short O---O contacts.

As the rotation matrix **R** and the translation vector **S** show, the minimum energy structure was found in a position nearly identical with that of the actual structure. This also resulted in a relatively small value for the average deviation:  $\varphi = 0.0697$  Å for all atoms and  $\varphi = 0.0706$  Å for carbons and oxygens only.

$$\mathbf{R} = \begin{pmatrix} 0.99972 & -0.02188 & 0.00852 \\ 0.02212 & 0.99933 & -0.02881 \\ -0.00788 & 0.02899 & 0.99954 \end{pmatrix}$$

$$\mathbf{P} = (1.5, 4.0, 6.0) \text{ \AA}$$

$$\mathbf{S} = (-0.02, 0.07, -0.02) \text{ \AA}$$

(The *z* coordinate of atom O(3) in the original paper is in error and should be 0.9267 instead of 0.9627.)

The repulsive energies and the closest contacts are listed in Table 6. Here the O---H contacts were predominant in repulsive packing energy, even though one very short distance in the actual structure was considerably lengthened in the minimum energy structure. The H---H interactions contributed only a small fraction to the total energy, which may be a reflection of the short C-H and O-H bond lengths. The C---H interaction was also nearly negligible. The values for  $E_r$  were low for both actual and computed structures: 14.02 and 13.15 respectively, again reflecting the presence of short C-H and O-H bond lengths.

### 6. $\beta$ -Arabinose

An interesting situation developed in the calculation of the minimum energy structure of this carbohydrate. The global energy minimum was not found in the vicinity of the reported crystal structure (Hordvik, 1961), but close to a position obtained by shifting the molecule by  $-a/4$ . Subsequent checking of the original data revealed that the reported coordinates were not in agreement with the convention of the *International Tables for X-ray Crystallography* (1952) (wrong origin was used) and furthermore, that discrepancies in both signs and magnitudes of the structure factors of  $0k0$  reflections in *a* and *c* projections were present. For these reasons, as well as for the purpose of observing whether the minimum energy structure could serve as an initial model for crystallographic refinement, the latter process was carried out with the reported structure factors, using the procedure outlined below. The *ORFLS* computer program (Busing, Martin & Levy, 1962) was used for refinement.

The *a* and *c* projections were refined separately as was done originally (Hordvik, 1961). After two cycles of refinement the *R* values were  $R_{hk0} = 0.060$  and  $R_{0kl} = 0.076$ , with hydrogens included and using isotropic temperature factors. Because of the small amount of

Table 5. Distribution of short contacts in actual and computed crystal structures of methyl- $\alpha$ -D-glucopyranoside

Inter-action	Total number of repulsive interactions (and H-bonds)	$E_r$ for observed structure	All hydrogen bonds are included		
			Atoms*	Closest contacts (Å)	Pred.
O---O	2	0.05	O(4)---O(5)	3.25	3.32
O---C	—	—	—	—	—
O---H	38	5.45	H(4')---O(5)	2.68	2.75
			O(3)---H(3)	2.72	2.83
			H(2)---O(4)	2.78	2.71
			H(6A)---O(2)	2.78	2.82
C---H	24	3.10	C(3)---H(6')	2.79	2.86
			C(6)---H(2')	2.81	2.72
			C(2)---H(3')	2.82	2.93
			H(4)---C(7)	2.83	2.85
H---H	58	5.78	H(2)---H(5)	2.22	2.19
			H(7C)---H(7B)	2.42	2.42
			H(6A)---H(7C)	2.42	2.48
			H(1)---H(7A)	2.51	2.48
O---O (H-bond)	8	0.88	O(6)---O(2)	2.70	2.69
			O(2)---O(3)	2.70	2.80
			O(3)---O(6)	2.70	2.77
			O(4)---O(6)	2.92	2.83

\* See footnote at the bottom of Table 2.

Table 6. *Distribution of short contacts in actual and computed crystal structures of D-glucono-(1,5)-lactone*

All hydrogen bonds are included.

Interaction	Total number of repulsive interactions (and H-bonds)	$E_r$ for observed structure	Closest contacts (Å)		
			Atoms*	Obs.	Pred.
O---O	2	0.31	O(1)---O(6)	3.04	3.24
O---C	4	0.05	O(2)---C(6)	3.28	3.31
O---H	34	10.05	O(1)---H(6')	2.59	2.83
			O(4)---H(4)	2.60	2.64
			O(2)---H(6A)	2.66	2.67
C---H	18	0.90	C(6)---H(6A)	3.01	2.99
			C(1)---H(6')	3.01	3.21
H---H	32	2.14	H(6B)---H(6A)	2.31	2.35
			H(3)---H(6B)	2.62	2.54
			H(4')---H(4)	2.70	2.73
O---O (H-bond)	8	0.58	O(3)---O(6)	2.68	2.82
			O(2)---O(3)	2.72	2.72
			O(1)---O(4)	2.84	2.73
			O(2)---O(6)	2.88	2.84

\* See footnote at the bottom of Table 2.

data available, high correlation coefficients occurred between some coordinates and temperature factors. This necessitated fixing several  $y$  coordinates and the  $x$  and  $y$  coordinates of C(4) in order to obtain agreement between both projections and to maintain a suitable C(3)–C(4)–C(5) bond angle. After relocating the hydrogens to within reasonable bond lengths and angles, two more cycles of refinement with the averaged  $y$  coordinates produced  $R_{hko}=0.065$  and  $R_{okl}=0.085$ . The fractional coordinates of this structure are listed in Table 7 and the bond lengths and angles in Table 8.

Table 7. *Revised atomic coordinates and temperature factors for  $\beta$ -arabinose*Unit cell:  $a=6.535$ ,  $b=19.467$ ,  $c=4.841$  Å. Space group:  $P2_12_1$ 

The standard errors for the Cartesian coordinates are about the same for  $x$ ,  $y$ , and  $z$ , and are 0.01 Å for oxygen and 0.015 Å for carbon atoms; the standard error for the temperature factors is 0.30 Å<sup>2</sup>.

	Fractional coordinates			Temperature factor (Å <sup>2</sup> )	
	$x$	$y$	$z$	$B_{hko}$	$B_{okl}$
O(1)	0.3125	0.4100	−0.0800	3.50	3.15
O(2)	0.3560	0.3017	0.2705	3.30	3.30
O(3)	−0.0603	0.2545	0.2350	4.50	3.40
O(4)	−0.2525	0.3765	0.4260	4.90	5.15
O(5)	0.0765	0.4633	0.2095	3.10	3.00
C(1)	0.2430	0.4170	0.1865	3.50	2.65
C(2)	0.1830	0.3468	0.2950	3.10	2.80
C(3)	0.0010	0.3200	0.1300	3.10	2.70
C(4)	−0.1780	0.3711	0.1475	4.10	3.95
C(5)	−0.0975	0.4420	0.0540	4.30	4.10
H(1')	0.395	0.448	−0.120	3.50	3.15
H(2')	0.320	0.256	0.310	3.30	3.30
H(3')	−0.020	0.220	0.110	4.50	3.40
H(4')	−0.300	0.333	0.490	4.90	5.15
H(1)	0.370	0.438	0.295	3.50	2.65
H(2)	0.140	0.350	0.510	3.10	2.80
H(3)	0.040	0.315	−0.085	3.10	2.70
H(4)	−0.305	0.355	0.020	4.10	3.95
H(5A)	−0.220	0.478	0.100	4.30	4.10
H(5B)	−0.070	0.448	−0.165	4.30	4.10

Table 8. *Bond lengths and angles of the new crystal structure of  $\beta$ -arabinose*

C(1)–C(2)	1.516 Å	C(2)–C(1)–O(1)	108.7°
C(1)–O(1)	1.375	C(2)–C(1)–O(5)	110.3
C(1)–O(5)	1.417	O(1)–C(1)–O(5)	113.0
C(2)–C(3)	1.525	C(1)–C(2)–C(3)	109.2
C(2)–O(2)	1.436	C(1)–C(2)–O(2)	108.6
		C(3)–C(2)–O(2)	111.2
C(3)–C(4)	1.538	C(2)–C(3)–C(4)	110.1
C(3)–O(3)	1.430	C(2)–C(3)–O(3)	109.7
		C(4)–C(3)–O(3)	110.1
C(4)–C(5)	1.545	C(3)–C(4)–C(5)	107.6
C(4)–O(4)	1.437	C(3)–C(4)–O(4)	110.9
		C(5)–C(4)–O(4)	109.0
C(5)–O(5)	1.425	C(4)–C(5)–O(5)	112.1
		C(1)–O(5)–C(5)	112.7

Using these coordinates, a new minimum energy structure was computed which yielded an average deviation  $\varphi=0.077$  Å ( $\varphi=0.070$  Å with only carbons and oxygens present) and which was in the position:

$$\mathbf{R} = \begin{pmatrix} 0.9988 & 0.0319 & -0.0366 \\ -0.0322 & 0.9995 & -0.0072 \\ 0.0364 & 0.0084 & 0.9993 \end{pmatrix}$$

$$\mathbf{P} = (0.20, 7.00, 0.80) \text{ Å}$$

$$\mathbf{S} = (-0.07, 0.07, 0.00) \text{ Å}$$

In comparison with the refined structure, the predicted structure yielded  $R_{hko}=0.26$  and  $R_{okl}=0.25$ . It is obvious that a good correspondence between the new crystal structure and the predicted structure of  $\beta$ -arabinose was obtained. The breakdown of repulsive energies and closest contacts is shown in Table 9. As was true in all other cases, the O---H and H---H interactions predominated. Quite large discrepancies in the short contacts between the actual and predicted structures were evident, which was undoubtedly due to the low precision in the atomic coordinates. The total repulsive energies for both structures were  $E_r=23.84$  for the actual structure and  $E_r=22.59$  for the computed one.

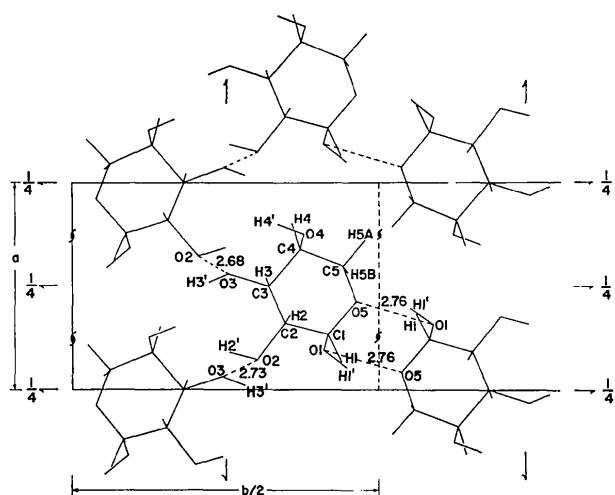


Fig. 1. Packing of  $\beta$ -arabinose viewed along the  $c$  axis of the unit cell. Hydrogen bonds are indicated by dashed lines.

The new structure of  $\beta$ -arabinose is shown in Fig. 1. It is essentially in agreement with the previously reported structure (Hordvik, 1961), with the exception of the O(2)---O(4) hydrogen bond which appears less probable in the present structure because of unfavorable position of H(4').

### Conclusions

The most significant result of this study was the accuracy which was achieved in the prediction of the structures of the six monosaccharides, notwithstanding the approximate nature of the procedures involved. It was also remarkable that the uncertainties in the positions of the hydrogens exerted little effect on the outcome of the calculations, even though the short contacts involving the hydrogens were the most important in all cases.

It was of interest in this connection that the average deviations calculated both with and without hydrogens were little different for a given structure. This meant that the inaccuracies in the positions of the hydrogens and the heavier atoms of a predicted structure were of the same order of magnitude, in contrast with the usual results of X-ray structure determination.

In view of the varying accuracy in the atomic coordinates and bond lengths of the different carbohydrates, no significance should be attached to the absolute values of  $E_r$ . Rather, the differences in  $E_r$  for the actual and predicted structures should be viewed as significant.

A comparison of  $\alpha$ -D-glucose and D-glucono-(1,5)-lactone, for which accurate bond length data were available, provided more information. The value of  $E_r$  for the lactone was considerably smaller than that for  $\alpha$ -D-glucose and the distribution of short contacts was different in both structures. This undoubtedly reflected the observed differences in C-H and O-H bond lengths for the two compounds.

Projecting the use of this method to more complex carbohydrates such as di-, oligo- and polysaccharides, or to structures with large rotatable substituents, more difficulties with false minima should be expected. The choice of strategy and the use of more accurate potential energy functions, possibly including attractive terms, remains to be determined. However, as the case of  $\beta$ -arabinose demonstrates, it is likely that the choice between a small number of alternative models could easily be made through standard X-ray refinement procedures. In addition, there may be other means available to assist in the solving of these problems (Neuman, 1970).

This work has been supported by the National Science Foundation Research Grant No. GP-27997. One of the authors (PZ) thanks the Deutsche Forschungsgemeinschaft for a travel grant.

Table 9. Distribution of short contacts in the revised crystal structure and computed structure of  $\beta$ -arabinose

All hydrogen bonds are included.

Inter-action	Total number of repulsive interactions (and H-bonds)	$E_r$ for observed structure	Closest contacts ( $\text{\AA}$ )		
			Atoms*	Obs.	Pred.
O---O	4	0.32	O(4)---O(2)	3.03	3.08
			C(3)---O(2)	3.20	3.31
			H(4')---O(2)	2.55	2.52
O---H	46	13.29	O(1)---H(2)	2.56	2.66
			H(5B)---O(4)	2.69	2.79
			H(3)---O(2)	2.72	2.84
			C(3)---H(2')	2.84	2.95
			H(1')---C(1)	2.93	2.92
C---H	22	2.04	C(5)---H(5A)	2.94	2.86
			H(4)---H(3')	2.11	2.32
			H(4')---H(2')	2.13	2.25
			H(3)---H(2)	2.17	2.11
H---H	56	7.75	H(3)---H(2')	2.27	2.41
			O(3)---O(2)	2.68	2.69
			O(3)---O(2)	2.73	2.80
			O(1)---O(5)	2.76	2.71
O---O (H-bond)	6	0.35			

\* See footnote at the bottom of Table 2.



## References

- BERMAN, H. M. & KIM, S. H. (1968). *Acta Cryst.* **B24**, 897.
- BROWN, G. M. & LEVY, H. A. (1965). *Science*, **147**, 1038.
- BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1962). *ORFLS*. Report ORNL-TM-305, Oak Ridge National Laboratory, Oak Ridge, Tennessee.
- CHU, S. C. & JEFFREY, G. A. (1968). *Acta Cryst.* **B24**, 830.
- CLARK, E. S. & GEERING, G. T. (1970). Papers presented at the ACA Summer Meeting, Aug. 16–22, Carleton Univ., Ottawa, Canada.
- HACKERT, M. L. & JACOBSON, R. A. (1971). *Acta Cryst.* **B27**, 203.
- HORDVIK, A. (1961). *Acta Chem. Scand.* **15**, 16.
- HORDVIK, A. (1966). *Acta Chem. Scand.* **20**, 1943.
- International Tables for X-ray Crystallography* (1952). Vol. I. Birmingham: Kynoch Press.
- NEUMAN, M. A. (1970). *Trans. ACA*, **6**, 111.
- WARSHEL, A. & LIFSON, S. (1970). *J. Chem. Phys.* **52**, 582.
- WILLIAMS, D. E. (1965). *Science*, **147**, 605.
- WILLIAMS, D. E. (1969). *Acta Cryst.* **A25**, 464.
- WILLIAMS, D. E. (1970). *Trans. ACA*, **6**, 21.
- WOBSE, G. & BLASENBREY, S. (1970). *Kolloid Z. Z. Polym.* **241**, 985.

*Acta Cryst.* (1972). **B28**, 3166

## Modified Tangent Procedures\*

BY HERBERT HAUPTMAN AND CHARLES M. WEEKS

*The Medical Foundation of Buffalo, 73 High Street, Buffalo, New York 14203, U.S.A.*

(Received 19 April 1972)

Two procedures for extending a basic set of known phases are described which, unlike the tangent formula, depend on the conditional probability distribution of  $\cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})$  as well as the distribution of  $\sin(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})$ . One of these modified tangent procedures has been applied to one unsolved and three solved crystal structures, and it has been found, in all instances, to be superior to the simple tangent formula when the initial set of phases is small. In addition, the modified tangent procedure, but not the simple tangent formula, is capable of making an enantiomorph selection in space group  $P2_1$  when the input set consists only of the origin-fixing phases and those determined by means of  $\Sigma_1$ .

### 1. Introduction

One of the most useful relationships which may be derived from the conditional probability distribution of the pair of phases  $\varphi_{\mathbf{k}}$  and  $\varphi_{-\mathbf{h}-\mathbf{k}}$ , where  $\mathbf{h}$  is a fixed vector having an unknown phase and  $\mathbf{k}$  ranges over all vectors in reciprocal space such that the magnitudes of  $|E_{\mathbf{k}}|$  and  $|E_{-\mathbf{h}-\mathbf{k}}|$  have fixed values, is the tangent formula (Karle & Hauptman, 1956). Not only is the tangent formula a widely used technique for the solution of equal atom problems in noncentrosymmetric space groups, but it is also useful for refining and extrapolating approximate phases derived in any way, e.g. when phases are calculated from a known fragment of a crystal structure (Karle, 1968).

The simple tangent formula depends on the conditional probability distribution of  $\sin Y$ , where  $Y = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}$ , but not on the distribution of  $\cos Y$ . The purpose of the present paper is to discuss tangent procedures which make more effective use of the probability distribution of the pair of phases  $\varphi_{\mathbf{k}}$  and  $\varphi_{-\mathbf{h}-\mathbf{k}}$  by forcing two of its parameters, the average values of

$\sin Y$  and  $\cos Y$  to agree as closely as possible with their theoretical expected values.

### 2. Derivation of the modified tangent procedures

Suppose that the space group is  $P1$  and that there are  $N$  identical atoms in the unit cell. Define  $A_{\mathbf{k}}$  by

$$A_{\mathbf{k}} = \frac{2}{N^{1/2}} |E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{h}+\mathbf{k}}|. \quad (2.1)$$

Assume that the reciprocal vector  $\mathbf{h}$  is fixed and that  $\mathbf{k}$ , the primitive random variable, ranges uniformly over that region of reciprocal space for which  $A_{\mathbf{k}}$  has a fixed, specified value. Under these circumstances the conditional probability distribution of the pair of phases  $\varphi_{\mathbf{k}}$ ,  $\varphi_{-\mathbf{h}-\mathbf{k}}$ , given  $A_{\mathbf{k}}$ , is known to be (Cochran, 1955; Hauptman, 1972a)

$$P(\Phi_0, \Phi_1 | A_{\mathbf{k}}) \simeq \frac{1}{4\pi^2 I_0(A_{\mathbf{k}})} \times \exp \{A_{\mathbf{k}} \cos(\varphi_{\mathbf{h}} + \Phi_0 + \Phi_1)\} \quad (2.2)$$

where  $I_0$  is the modified Bessel function. From (2.2) it follows readily that (Hauptman, 1966, 1972a; Karle & Karle, 1966)

\* Portions of this paper were presented at the meetings of the American Crystallographic Association at Columbia, South Carolina, February 1971, and Ames, Iowa, August 1971.