Acta Cryst. (1983). B**39**, 469–480

The Hydrogen-Bonding Patterns in the Pyranose and Pyranoside Crystal Structures

BY G. A. JEFFREY AND J. MITRA

Department of Crystallography, University of Pittsburgh, Pittsburgh, PA 15260, *USA*

(*Received* 24 *September* 1982; *accepted* 23 *November* 1982)

Abstract

The hydrogen-bonding patterns in the crystal structures of the simple cyclic carbohydrates attempt to comply with three *rules*:

(1) Maximize the hydrogen-bond energy by including all the hydroxyls and as many of the ring and glycosidic oxygens as possible.

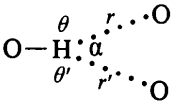
(2) Maximize the energy through the *cooperative* effect by forming infinite or long finite chains of hydrogen bonds.

(3) Take into account the *anomeric effect* by including the anomeric hydroxyls as strong donors and poor acceptors of hydrogen bonds.

These three rules are mutually incompatible. Compromises give rise to patterns of hydrogen bonding which can be divided into four distinct classes with related sub-groups. These patterns are distributed almost equally between the crystal structures of the 58 pyranoses, furanoses, pyranosides and furanosides included in this survey.

Introduction

An in-depth study of the geometry associated with O–H···O hydrogen bonding in the 24 monosaccharide and related crystal structures determined by means of neutron diffraction has been reported by Ceccarelli, Jeffrey & Taylor (1981). The emphasis in that work was on the factors which determine the positions of the hydroxyl hydrogen atoms and their relationship to the overall pattern of the hydrogen bonding in the crystal structure. This is a question that is frequently overlooked in X-ray-analyzed crystal structures for the good reason that the hydrogen positions are often not known with certainty, especially those attached to the oxygen atoms. In X-ray analyses, the $X\cdots Y$ separation of the X–H$\cdots Y$ hydrogen bond is often referred to as the hydrogen-bond length, whereas it is, in fact, a function of three structural quantities, the X–H covalent-bond length, the H$\cdots Y$ hydrogen-bond length, and the X–H$\cdots Y$ angle, which is rarely 180°. One of the more unexpected results of that survey was the high proportion of three-center (bifurcated) hydrogen bonds that were observed, *i.e.* 25%. It was noted with interest that this corresponded approximately to the ratio of acetal oxygens to hydroxyl oxygens in the data set of the molecules examined. The term 'three-center' hydrogen bond is used to describe the geometry in which the hydrogen-atom position is clearly determined by three nearest-neighbor oxygen atoms, that to which it is covalently bonded and the two next-nearest neighbors, as in (I).

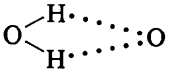


$\theta + \theta' + \alpha \simeq 360°$

$1{\cdot}4 < r < r' < 2{\cdot}85$ Å.

(I)

The cut-off value of 2·85 Å is consistent with the Hamilton & Ibers (1968) criterion for a hydrogen bond, $H\cdots Y \leq W_H + W_O - 0{\cdot}3$ Å, if the Allinger (1976) van der Waals constants are used for W_R, *i.e.* $W_H = 1{\cdot}50$ Å, $W_O = 1{\cdot}65$ Å. We prefer three-center bond to the term 'bifurcated' which was originally defined as (II) by Pimentel & McClellan (1960), and is still used in this context by theoretical chemists and spectroscopists (*cf.* Kollman & Allen, 1969).



(II)

In this paper, we examine the overall hydrogen-bonding patterns of the known crystal structures of 58 mono- and disaccharides where the hydrogen positions are certain from neutron diffraction analyses or reasonably certain from good quality X-ray analyses. We have selected molecules which contain only hydroxyl groups and acetal or hemi-acetal oxygen atoms. They include, therefore, in addition to pyranose and furanose molecules, the pyranosides, furanosides, and anhydro sugars, which contain two fused rings with two acetal ring oxygen atoms. Ethylidene and propylidene derivatives have been omitted, because we wish to analyze only those patterns in which the principal cohesive forces are the hydrogen bonds. Carbohydrates containing carbonyl groups, such as lactones and acids, were excluded, as were those

0108-7681/83/040469-12$01.50

containing other functional groups, such as amino or acetyl groups. As in the amino acids, the presence of a wider variety of hydrogen-bonded functional groups greatly increases the complexity of pattern recognition (Jeffrey & Maluszynska, 1982).

Experimental

The atomic coordinates were taken from the Cambridge Crystallographic Data File, January 1982 update (Allen *et al.*, 1979). The hydrogen-bonding patterns were generated and displayed on an MMS-X computer graphics system. This permitted the rotation of a display of the hydrogen-bonded functional groups alone, while maintaining them in the same relative positions as they occur in the crystal structure, an operation which is not possible with normal mechanical crystal-structure models. Examples of the hydrogen-bonding patterns thereby obtained are given in the reports of several of the X-ray and neutron diffraction analyses, *cf.*, Takagi & Jeffrey (1978*a*,*b*,*c*; 1979*a*,*b*).

Since X-ray analyses frequently report abnormally short observed O—H distances, as compared with the internuclear values from neutron diffraction, all O—H bond distances from the X-ray analyses were normalized to a standard covalent-bond-length value of 0·97 Å, by moving the hydrogen atom along the direction of the O—H bond. This was done to provide consistency with the neutron diffraction analyses that were included in the data set.

Table 1* gives the schematics of the 43 hydrogen-bonding schemes of the monosaccharides that were analyzed and grouped into their appropriate categories according to the rules stated below. The 15 disaccharides, which have more-complex hydrogen bonding, generally involving more than one category, are given in Table 2. The atomic notation used is that of the original paper which describes the crystal-structure analysis. In the majority of papers, standard carbohydrate notation is used; O(5) is an aldopyranose ring atom, O(6) is a ketopyranose ring atom, O(2) is a furanose ring atom, O(1) is a glycosidic linkage atom, and O(1)H is the anomeric hydroxyl of a pyranose sugar. In a reducing disaccharide, the atomic notation on the reducing residue is primed. The H···O distances from the neutron analyses are as observed, without correction for thermal motion. Those from the X-ray analyses are normalized as described above.

* In Tables 1, 2 and 3, the crystal structure is identified by its name and the Cambridge Data File REFCODE. (N) indicates a neutron diffraction study. The arrows indicate the hydrogen-bond donor direction. Interactions with H···O > 2·0 Å are dotted. The H···O distances are in Å and the O—H···O angles in degrees. Intramolecular hydrogen bonds are marked *. Where there is more than one crystallographically independent molecule in the unit cell, *A*, *B*, *C*, *etc.* are used as the supplementary atomic notation. The numbers preceding the name refer to the pattern under discussion in the text.

The general rules

There are three general controlling rules which govern the observed hydrogen-bonding patterns. These are:

(1) In order to make the hydrogen-bonding contribution to the lattice energy as large as possible, all hydroxyls and as many as possible of the ring or glycosidic oxygen atoms should be hydrogen bonded.

(2) Advantage should be taken of the *cooperativity*, or *non-additivity* property of the hydrogen bonds. This states that $E(\mathrm{H}\cdots\mathrm{O})_n > nE(\mathrm{H}\cdots\mathrm{O})$; thereby favoring the formation of infinite chains such as $\cdots\mathrm{O}(\mathrm{H})\cdots\mathrm{O}(\mathrm{H})\cdots$ or closed loops. Cooperativity increases with the length of the chain and therefore long finite chains are energetically favored over short chains or isolated bonds.

(3) The hemi-acetal hydroxyl, $-\mathrm{O}-\mathrm{C}(-)(-)-\underline{\mathrm{OH}}$ has special properties in carbohydrate chemistry. It is responsible for epimerization and for the phenomenon of mutarotation. It is attached to the *anomeric* carbon atom, which is unique in simple sugars by being the only carbon atom bonded to two oxygen atoms. The anomeric hydroxyl has special hydrogen-bonding properties which have been studied experimentally by Jeffrey, Gress & Takagi (1977) and theoretically by Tse & Newton (1977) and Newton (1983). One of these characteristics is that it is a stronger than usual hydrogen-bond donor, and a weaker than usual acceptor.

There is clearly an inconsistency between rule 1 and rules 2 and 3. Ring and glycosidic oxygens have no hydrogen atoms attached to them. Therefore, they cannot participate in infinite chains or loops. Similarly if anomeric hydroxyls are to participate in infinite chains, they must function as donors as well as acceptors. It is the way in which compromises are reached concerning these conflicting requirements that gives rise to the four principal categories of hydrogen-bonding patterns described below.

The hydrogen-bonding patterns

The hydrogen-bonding patterns fall into four general classes.

Type I. Infinite chains are formed by excluding from the hydrogen bonding all the ring or glycosidic oxygen

atoms, or by including them by means of a very weak component of a three- or four-center bond.

Type II. Finite chains are formed which terminate at a ring or glycosidic oxygen. With pyranoses or furanoses, there are two sub-sections of this class.

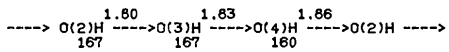
(*a*) The finite chains commence with the anomeric hydroxyl and terminate at the ring or glycosidic oxygen.

(*b*) A less desirable variation is a finite chain which starts with a hydroxyl other than the anomeric, and terminates with the ring oxygen.

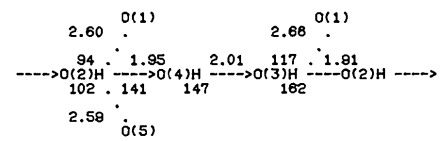
Table 1. *Patterns of hydrogen bonding in carbohydrate crystal structures, monosaccharides*

TYPE I : INFINITE CHAIN , RING AND GLYCOSIDIC OXYGENS OMITTED

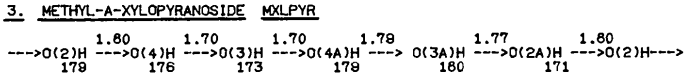
1. METHYL-B-L-ARABINOSIDE MBLAR10



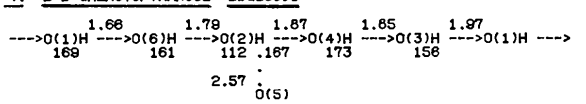
2. METHYL-B-RIBOPYRANOSIDE(N) MDRIBP02



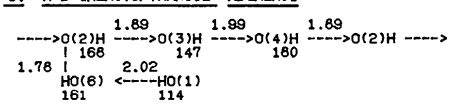
3. METHYL-A-XYLOPYRANOSIDE MXLPYR



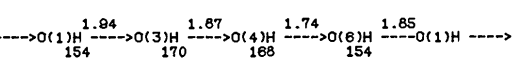
4. B-D-GALACTOPYRANOSE BDGLOS01



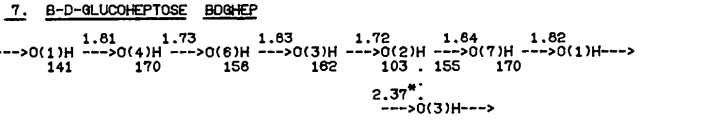
5. A-D-GALACTOPYRANOSE ADGALA01



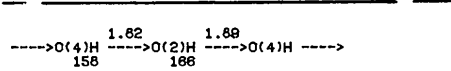
6. 2-DEOXY-2-FLUORO-B-D-MANNOPYRANOSE XFMANP



7. B-D-GLUCOHEPTOSE BDGHEP

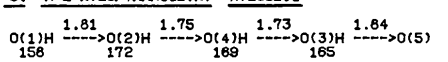


8. METHYL-3,6-ANHYDRO-A-D-GALACTOPYRANOSIDE MANGAL

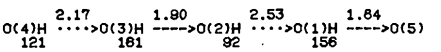


TYPE II : FINITE CHAIN TERMINATING WITH RING OXYGEN

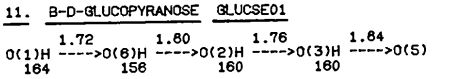
9. A-L-XYLOPYRANOSE(N) XYLOSE01



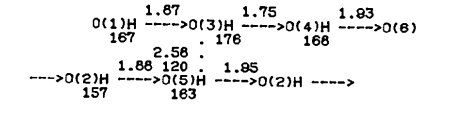
10. A-L-FUCOPYRANOSE ALFUCO



11. B-D-GLUCOPYRANOSE GLUCSE01



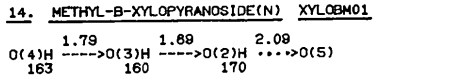
12. A-L-SORBOPYRANOSE(N) SORBOL01



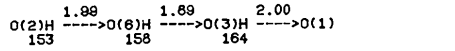
13. A-D-TAGATOPYRANOSE TAGTOS

SEE ALSO TYPE III.

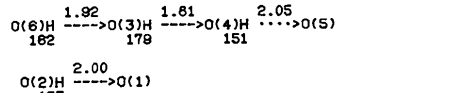
14. METHYL-B-XYLOPYRANOSIDE(N) XYLOBM01



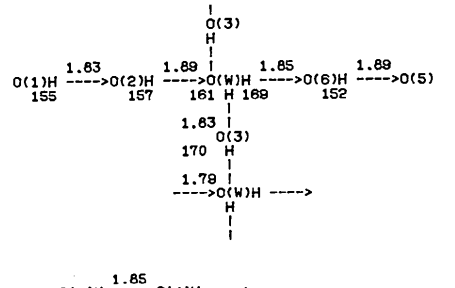
15. METHYL-4-DEOXY-4-FLUORO-A-D-GLUCOPYRANOSIDE MXFGPY



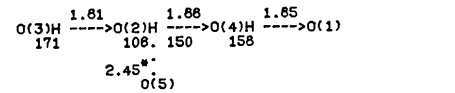
16. METHYL-A-D-MANNOPYRANOSIDE(N) MEMANP11



17. A-D-GLUCOPYRANOSE MONOHYDRATE GLUCMH11



18. 1,6-ANHYDRO-B-D-GLUCOPYRANOSE AHGLPY10



19. METHYL-B-D-GLUCOPYRANOSIDE HEMIHYDRATE MBDGPH10

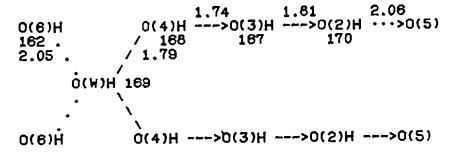
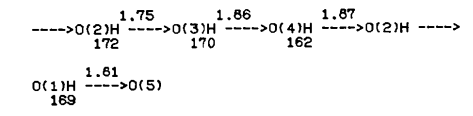


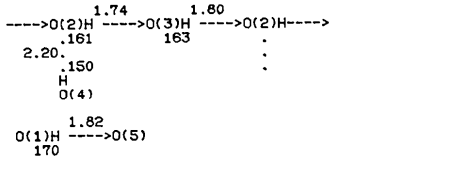
Table 1 (*cont.*)

TYPE III : INFINITE CHAINS WITH SEPARATE OH-->O(5) OR O(1)

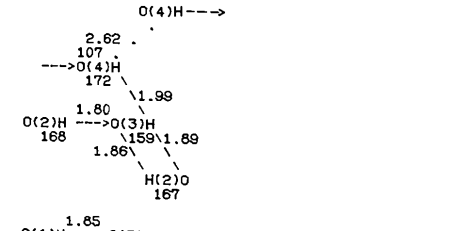
20. B-D,L-ARABINOPYRANOSE(N) ABINOR01



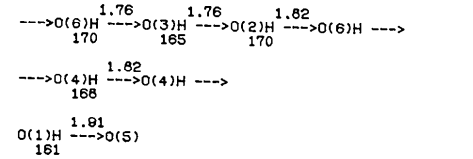
21. B-L-ARABINOPYRANOSE(N) ABINOS01



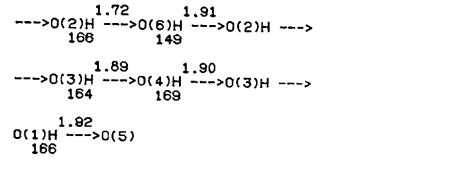
22. A-D,L-FUCOPYRANOSE ADLFUC



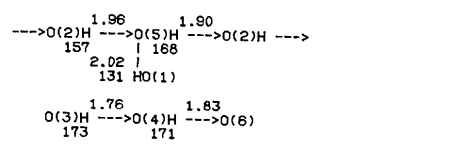
23. A-D-GLUCOPYRANOSE(N) GLUCSA01



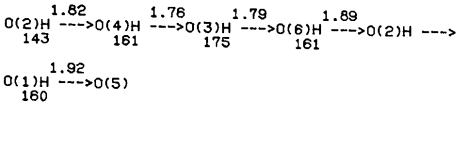
24. A-D-L-MANNOPYRANOSE ADLMAN



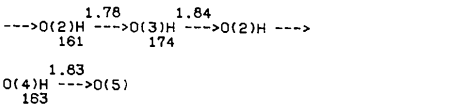
25. A-D-TAGATOPYRANOSE TAGTOS



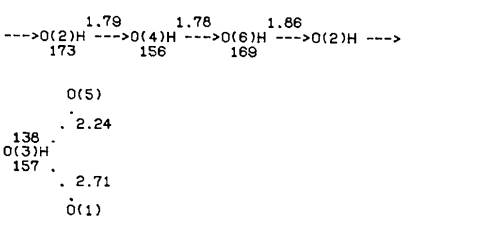
26. A-D-TALOPYRANOSE ADTALO10



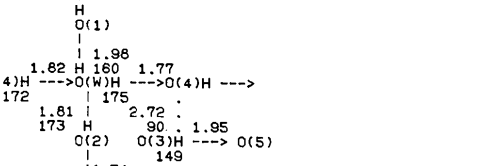
27. METHYL-A-L-ARABINOPYRANOSE MALARA10



28. METHYL-B-GALACTOPYRANOSIDE(N) MBDGAL02



29. B-D-DIGITOXOSE BDDIGX



TYPE IV : INFINITE CHAINS WITH 3-CENTER BONDS TO ACETAL OXYGEN

30. A-L-RHAMNOSE MONOHYDRATE(N) RHAMAH12

31. 1,6-ANHYDRO-B-D-GALACTOPYRANOSE AHGALP

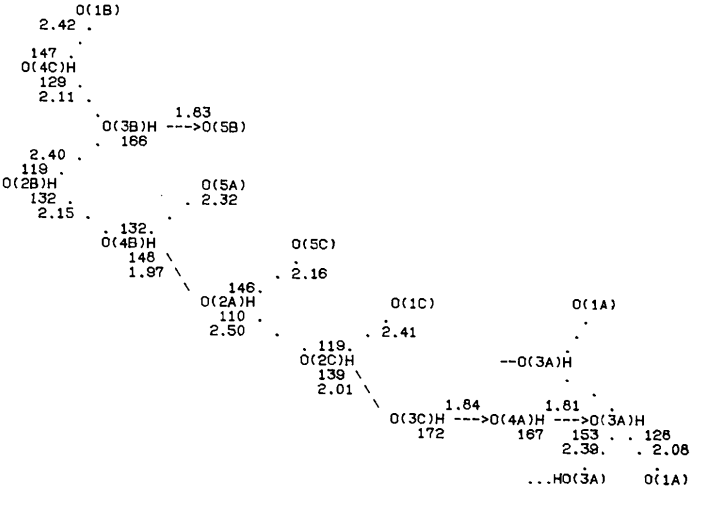
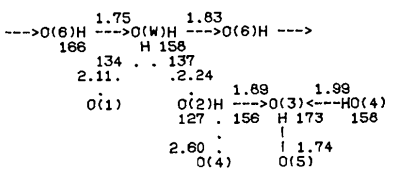
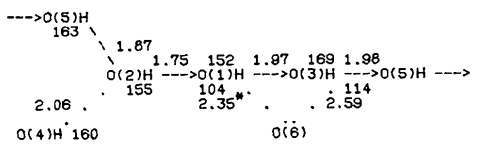


Table 1 (*cont.*)

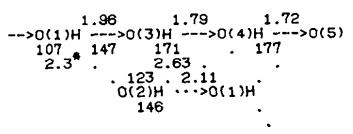
32. METHYL-A-D-GALACTOPYRANOSIDE MONOHYDRATE(N) MGALPY01



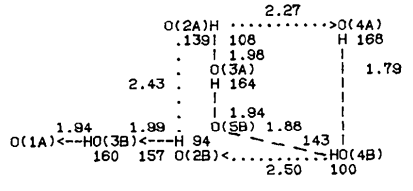
33. B-D-FRUCTOPYRANOSE(N) FRUCT002



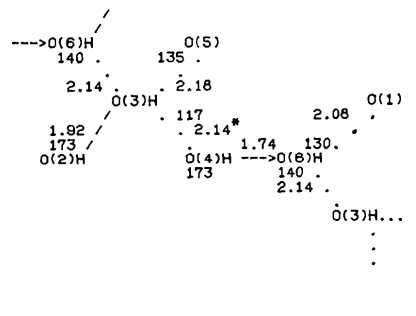
34. B-L-LYXOPYRANOSE(N) LYXOSE01



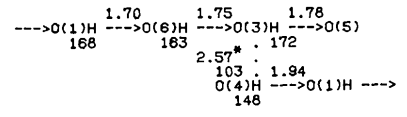
35. 1,6-ANHYDRO-B-D-MANNOPYRANOSE



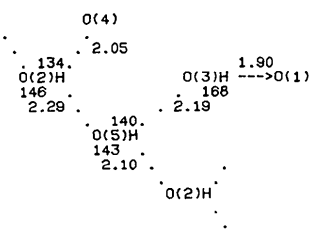
36. METHYL-A-D-ALTROPYRANOSIDE(N) MALTPY01



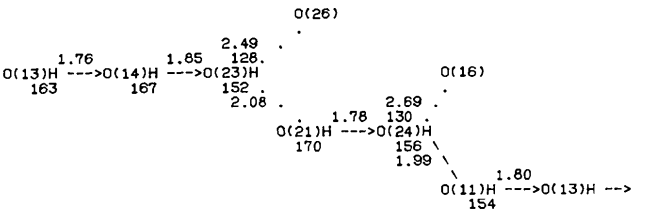
37. 2-DEOXY-D-GLUCOPYRANOSE



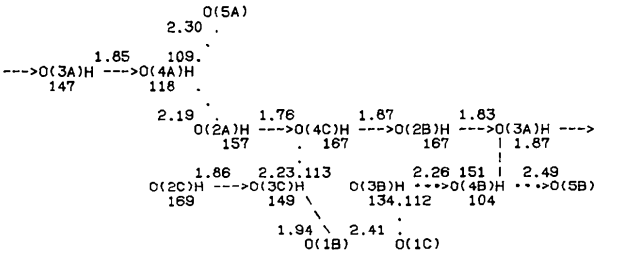
38. 1,6-ANHYDRO-B-D-MANNOFURANOSE AMANOF



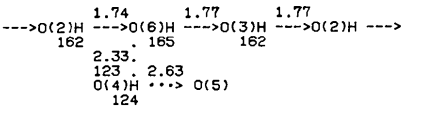
39. 2,6-ANHYDRO-B-D-FRUCTOFURANOSE ANFFUR



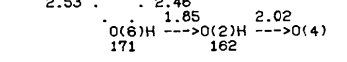
40. 1,6-ANHYDRO-B-D-ALLOPYRANOSE(ALLOSAN)



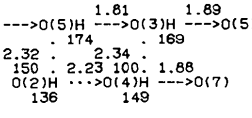
41. METHYL-A-D-GLUCOPYRANOSIDE(N) MGLUCP11



42. METHYL-A-D-GALACTOFURANOSIDE MGALAF



43. 1,6-ANHYDRO-L-IDITOL AHIDIT



In the methyl pyranosides, furanosides, 1,6-anhydro sugars, and non-reducing residues of disaccharides, there are no anomeric hydroxyls, and two other variations are observed.

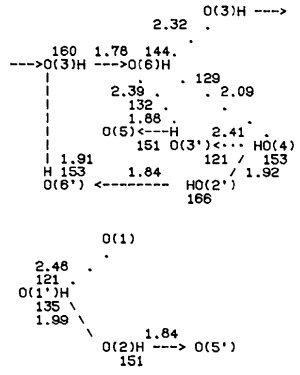
(*c*) A finite chain terminates at the ring oxygen, but omits the glycosidic oxygen, or *vice versa*.

(*d*) A finite chain terminates with a ring *or* glycosidic oxygen and includes the other acetyl oxygen through a three-center bond.

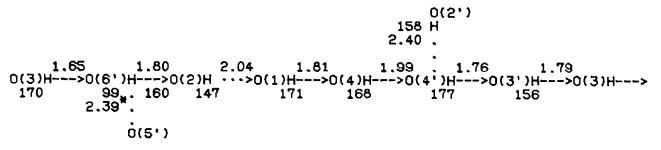
Type III. An infinite chain is formed from the majority of the hydroxyl groups, with a separate finite

Table 2. *Patterns of hydrogen bonding in carbohydrate crystal structures, disaccharides*

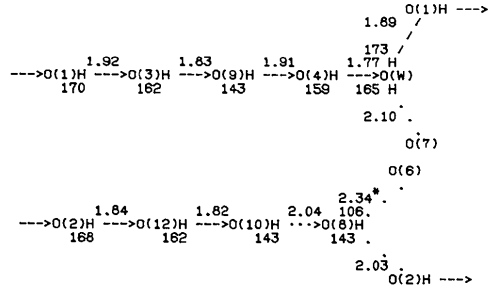
1. B-CELLOBIOSE CELLOB02



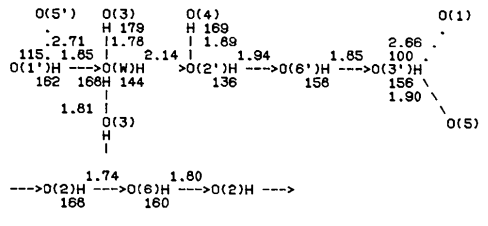
2. GENTIOBIOSE GENTBS01



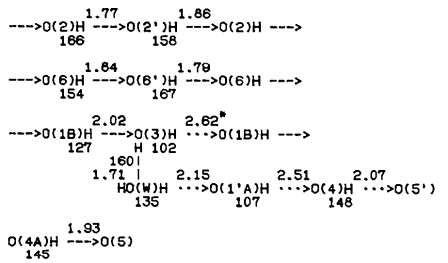
3. ISOMALTULOSE HYDRATE IMATUL



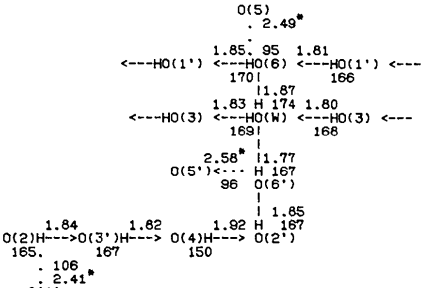
4. A-LACTOSE MONOHYDRATE LACTOS10



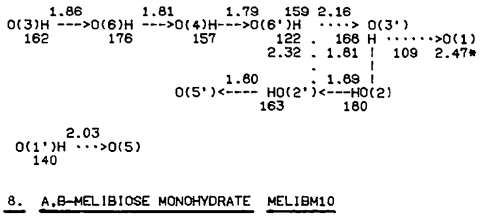
5. LAMINARABIOSE LAMBIO



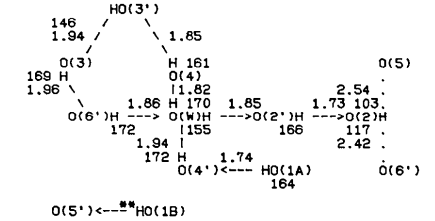
6. B-MALTOSE MONOHYDRATE(N) MALTOS01



7. A-MALTOSE MALTOT



8. A,B-MELIBIOSE MONOHYDRATE MELIBM10



9. METHYL-B-MALTOPYRANOSIDE MONOHYDRATE MMALTS

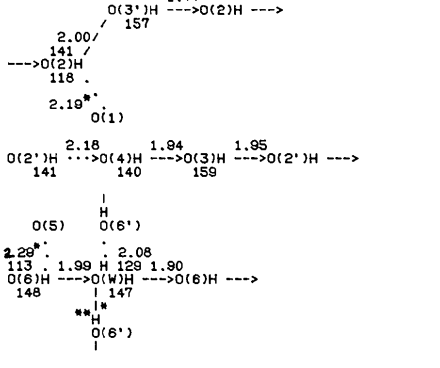
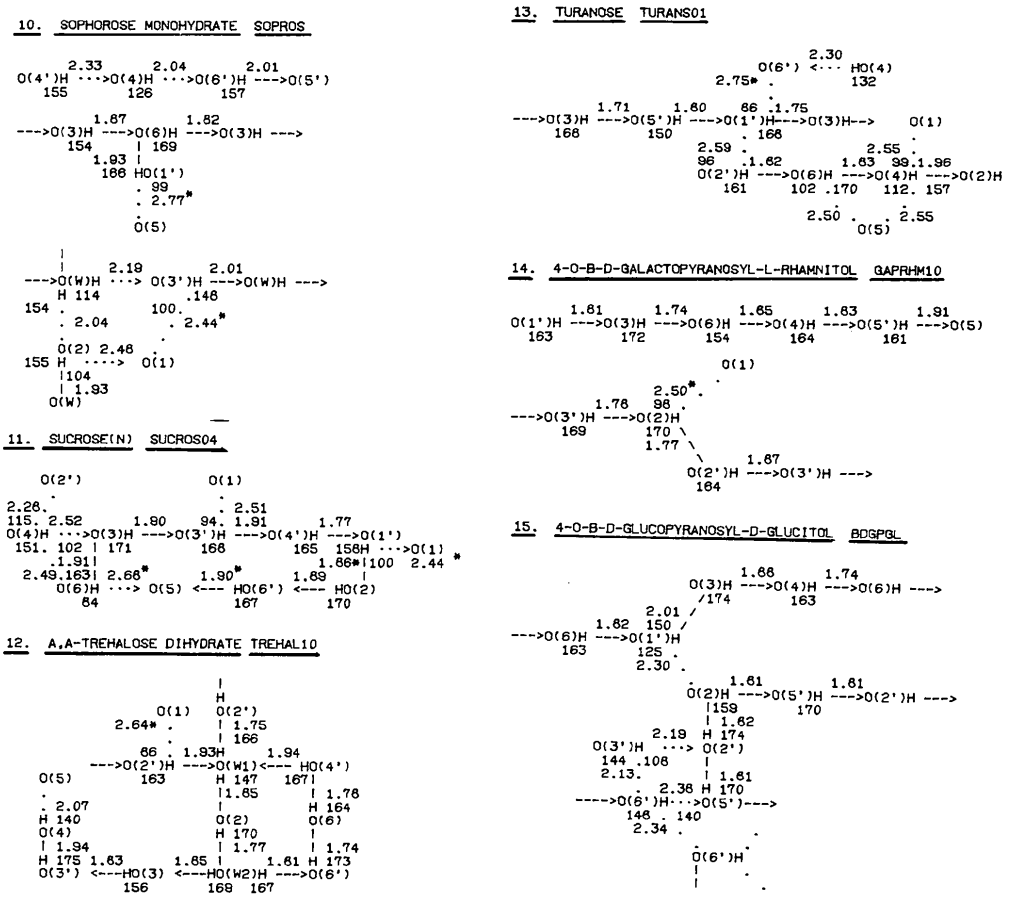


Table 2 (*cont.*)



chain which terminates in a ring or glycosidic oxygen. There are two variations.

(*a*) In pyranoses and furanoses, and reducing disaccharides, an optimum situation is to form a separate anomeric hydroxyl to ring oxygen bond, with the remaining hydroxyls forming the infinite chains.

(*b*) For pyranosides, furanosides, and non-reducing residues, a favorable case is to include both acetal oxygens in a separate three-center bond.

Type IV. Three-center hydrogen-bonding is utilized to incorporate acetal oxygens into the hydrogen-bonding scheme without disrupting the cooperative advantage of the infinite chain or closed loop. In view of the theoretical studies of Newton, Jeffrey & Takagi (1979) on the energetics of the three-center bonds, this might be expected to be the energetically most favorable pattern, but in fact it only accounts for approximately one quarter of the patterns observed.

The role of water molecules

Despite their solubility in water, carbohydrates do not form a wide variety of hydrates, as compared with aliphatic amines, for example (Pickering, 1893; Jeffrey, 1969). The exceptions are the cycloamyloses (*cf.* Lindner & Saenger, 1982), where the topology of the molecules necessarily leaves voids that are filled by the solvent species. Carbohydrates form strong hydrogen-bonded systems in the crystalline state, which generally exclude water solvent molecules unless they can fit into the carbohydrate-dominated pattern. A remarkable example of this is galactaric acid (mucic acid), $HOOC(CHOH)_4COOH$, which is insoluble in cold water owing to an exceptionally strong hydrogen-bonding system in the crystalline state (Jeffrey & Wood, 1982).

It is noticeable that hydrates are more common in disaccharides than in monosaccharides (8 out of 15 *versus* 4 out of 43). This is presumably because the more complex disaccharides pack less efficiently with their own hydrogen-bonding functional groups. Water molecules have double donor and acceptor functions when they occur in the crystal structures of carbohydrates. The water oxygen is approximately tetrahedrally coordinated with four hydrogen-bonded neighboring oxygens, two as donors and two as

acceptors. They most frequently occur therefore at points of intersection of two infinite or finite chains. The homo-, anti-, and heterodromic loops, such as observed in the heavily hydrated cycloamylose structures, are rarely observed in the lower hydrates of the mono- and disaccharides (Lindner & Saenger, 1982).

Hydrogen bonding in polyols

The preceding discussion indicates that the variety of hydrogen-bonding patterns observed in the cyclic monosaccharides is a consequence of the alternative ways of satisfying, to a greater or less degree, the conflicting requirements of the *cooperativity* and the hydrogen-bond-acceptor property of the acetal or hemi-acetal oxygen atoms which have no hydrogen atoms attached. It is relevant therefore to examine the crystal structures of carbohydrates which contain only hydroxyl groups without the ring or glycosidic oxygens. These are the pentitols, $HOH_2C(CHOH)_3CH_2OH$, and the hexitols, $HOH_2C(CHOH)_4CH_2OH$, the crystal structures of nine of which have been determined (Jeffrey & Kim, 1970; Azarnia, Jeffrey & Shen, 1972). With one exception, these were all room-temperature X-ray analyses, and, in a few cases, the hydroxyl hydrogen atoms were inferred rather than observed. The hydrogen-bond geometry is summarized in Table 3. It shows quite clearly that with hydroxyl groups as the only hydrogen-bonding functional groups present, infinite chains are invariably formed, supporting our hypothesis concerning the importance of the cooperative $(O{-}H\cdots O{-}H)_n$ bonding.

Notes and discussion of Table 1

Our description of the hydrogen-bonding schemes differs from that generally presented in the original papers in two respects:

(*a*) It is customary in descriptions of the results of a crystal-structure analysis to refer all hydrogen-bonding geometry to a central molecule; that with symmetry coordinates *x*,*y*,*z*. We trace the hydrogen bonding through successively linked molecules.

(*b*) As a result of the analysis of the neutron diffraction data by Ceccarelli, Jeffrey & Taylor (1981), we place more emphasis on the role of the three-center, and in some cases four-center, bonds, to satisfy the hydrogen-bond-acceptor characteristic of the acetal or hemi-acetal oxygen atoms.

Type I: *Infinite chains excluding ring and glycosidic oxygens.* (The number preceding the REFCODE refers to the number of the pattern in Table 1.)

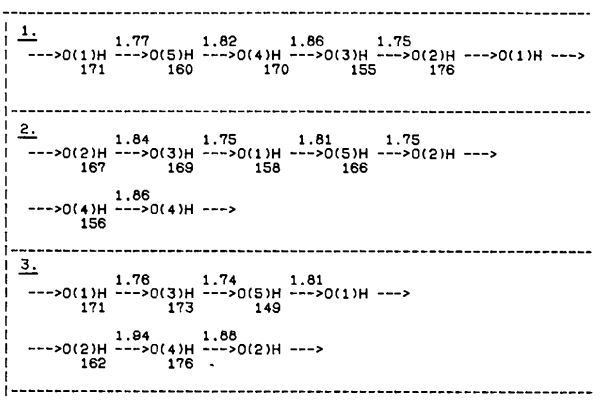
2. *MDRIBP*02. This pattern could be included in type IV. The ring and glycosidic oxygens are included by means of *very weak* components of a four-center (trifurcated) bond from O(2)H. All five atoms involved, O(2), O(1), O(4), O(5) and H(O2) are approximately in

Table 3. *Patterns of hydrogen bonding in carbohydrate crystal structures, alditols*

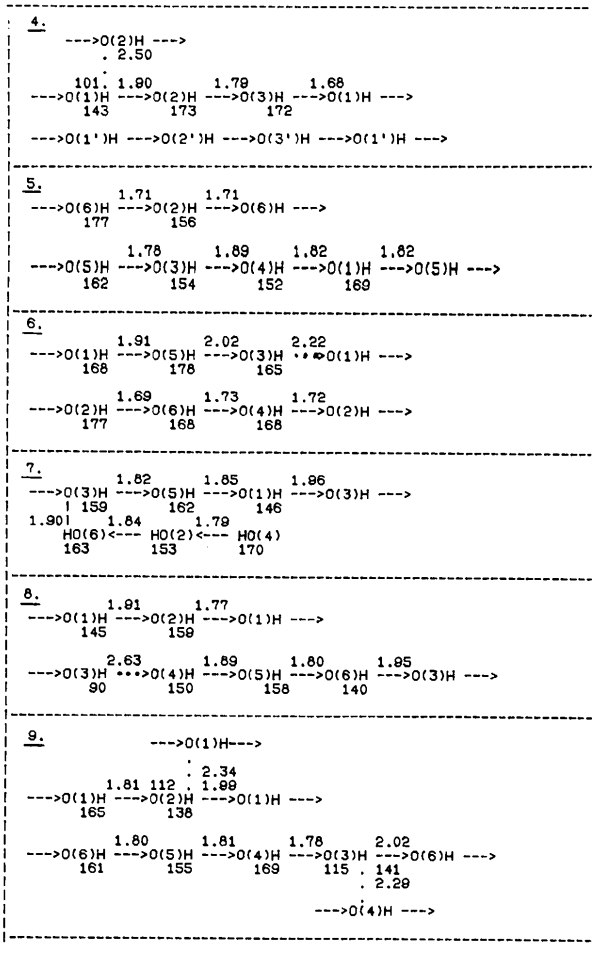
Key to molecules: (1) D,L-Arabinitol. (2) Ribitol. (3) Xylitol. (4) Allitol. (5) Galactitol. (6) D-Glucitol (N). (7) D-Iditol. (8) D-Mannitol, *B*-form. (9) D-Mannitol, *K*-form.

MOLECULE

$HOH_2C(CHOH)_3CH_2OH$



$HOH_2C(CHOH)_4CH_2OH$



one plane. O(1) also receives the weak component of a three-center bond from O(3)H.

4. *BDGLOS*01. This pattern could also be included in type IV. The ring oxygen O(5) is included by the very weak component of the three-center bond from O(2)H.

5. *ADGALA*01. The anomeric hydroxyl O(1) is a donor only, initiating a short side chain, which links to an infinite chain (Jeffrey & Shiono, 1977).

7. *BDGHEP*. The identical infinite chains are cross-linked by the weak component of a three-center bond from O(2)H.

8. *MANGAL*. Both the ring oxygens, O(3) and O(5), and the glycosidic oxygen, O(1), are omitted from the hydrogen bonding. This is in marked contrast to the other anhydro compounds discussed later.

Type II: *Finite chains terminating at ring or glycosidic oxygen atoms.*

9. *XYLOSE*01 *and* 11. *GLUCSE*01. These are classic examples of finite chains, originating at the anomeric hydroxyl and terminating at the ring oxygen.

10. *ALFUCO*. The poor geometry at O(2)H and O(4)H suggests that the hydrogen positions are incorrectly located in this analysis.

12. *SORBOL*01. O(6) is the ring oxygen in this keto sugar. This pattern consists of an infinite chain and a finite chain, weakly linked through the minor component of a three-center bond from O(5)H.

14. *XYLOBM*01. The ring oxygen O(5) is included, but the glycosidic oxygen O(1) is omitted from this pattern.

15. *MXFGPY*. This is the reverse of the preceding structure; the glycosidic oxygen O(1) is included, but the ring oxygen O(5) is omitted.

16. *MEMANP*11. In contrast to the two preceding structures, both ring and glycosidic oxygens are included, by terminating a finite chain and by a single link.

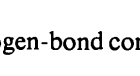
17. *GLUCMH*11. A finite chain originating at the anomeric hydroxyl O(1)H and terminating at the ring oxygen O(5) intersects at the water molecule with an infinite chain. There is also a separate infinite chain linking all the O(4) hydroxyls.

18. *AHGLPY*10. In this anhydro compound, one of the ring oxygens, O(1), terminates the finite chain, while the other, O(5), is included *via* the minor intramolecular component of a three-center bond.

19. *MBDGPH*10. Two finite chains, terminating at the ring oxygen, O(5), related by a twofold axis, intersect at the water oxygen.

Type III: *Infinite chain with separate finite chain terminating at ring or glycosidic oxygen.*

20. *ABINOR*01, 21. *ABINOS*01, 22. *ADLFUC*, 23. *GLUCSA*01, 24. *ADLMAN*, *and* 26. *ADTALO*10. These are classic examples of the anomeric hydroxyl O(1)H forming a separate link to the ring oxygen O(5), while the remainder of the hydroxyl groups form infinite chains.

22. *ADLFUC*. This structure is reported to contain the rare hydrogen-bond configuration  .(This should be confirmed by neutron diffraction.)

23. *GLUCSA*01. It is interesting to compare this pattern with that of the glucose monohydrate (No. 17). In both, the O(4)H hydroxyls form separate infinite chains.

25. *TAGTOS*. O(6) is the ring oxygen in this keto sugar.

27. *MALARA*10. The glycosidic oxygen, O(1), is omitted.

28. *MBDGAL*02. This is an interesting example in which both the ring and glycosidic oxygens are included by means of a separate three-center bond.

29. *BDDIGX*. The separate link to the ring oxygen O(5) is joined to the infinite chain by the weak component of a three-center bond.

Type IV: *Infinite chains with three-center bonds.*

31. *AHGALP*. The structure contains three crystallographically independent molecules. There are nine hydroxyls and six ring oxygens, all of which are included in the hydrogen bonding by means of three-center bonds. This permits the formation of a long finite chain originating at O(4*C*), which joins an infinite chain formed by the O(3*A*)H groups.

32. *MGALPY*01. A finite chain terminating at the ring oxygen O(5) joins an infinite chain at the water molecule by means of a three-center interaction. The glycosidic oxygen O(1) is included by means of a three-center bond from one of the water hydrogen atoms.

33. *FRUCTO*02. The ring oxygen O(6) receives two weak components of two three-center bonds, one of which is intramolecular. O(4)H is almost a non-bonded hydroxyl.

35. (*no REFCODE*). This pattern is rare in a monosaccharide. There is a four-link loop of strong bonds linked into infinite chains by the weaker components of three-center bonds (Maluszynska, Kinoshita & Jeffrey, 1982).

37. (*no REFCODE*). The infinite chain is formed between two finite chains by a weak intramolecular component of a three-center bond (Maluszynska, Ruble & Jeffrey, 1981).

38. *AMANOF*. There are two independent molecules in the structure. The anhydro ring oxygens O(16), O(26) are included, but the furanose ring oxygens are not.

40. (*no REFCODE*). As in 1,6-anhydro-β-galactopyranose, there are three independent molecules in the

structure, but not all oxygen atoms are included. One ring oxygen, O(1), is omitted from the hydrogen bonding (Norrestam, Bock & Pedersen, 1981).

41. *MGLUCP*11. One hydroxyl O(4)H makes only a weak three-center bond. One component links to the infinite chain, the other to the ring oxygen O(5). The glycosidic oxygen O(1) is omitted.

42. *MGALAF*. The finite chain which terminates at the ring oxygen O(4) is linked to an infinite chain by weak three-center bonds to and from O(6)H.

43. *AHIDIT*. The finite chain is linked to an infinite chain by two three-center bonds.

Notes and discussion of Table 2

The hydrogen-bonding patterns of the disaccharides frequently overlap two or more of the types identified in the monosaccharides. For this reason they are not separated into types.

1. *CELLOB*02. An infinite chain and a short finite chain. The infinite chain is linked to an infinite spiral. The finite chain starts with the anomeric hydroxyl O(1′)H and terminates with the ring oxygen O(5′). The other ring oxygen O(5) receives a strong bond from O(3′)H, which forms part of a weakly bonded spiral formed by the weaker component of three-center bonds.

2. *GENTBS*01. This is a relatively simple pattern with an infinite chain, which includes one of the ring oxygens O(5′) by means of a three-center bond, but excludes O(5) and the linkage oxygen O(6). The hydroxyl O(2′)H is only very weakly bonded to O(4′) with an H···O distance of 2·40 Å.

3. *IMATUL*. Two separate infinite chains. The ring oxygen O(7) is included by a weak bond from the water molecule. The other ring oxygen O(6) is included by means of a three-center bond. In this pattern the water molecule permits the continuation of the infinite chain and the inclusion of a ring oxygen by means of its double-donor property. This is not a common function. It is an exception to the usual role as an intersection point of two chains.

4. *LACTOS*10. An infinite chain and an infinite chain which intersects with a finite chain at the water oxygen. The finite chain originates at the anomeric hydroxyl O(1′)H and terminates with a three-center bond to a ring oxygen O(5) and the linkage oxygen O(1). The intersecting infinite chain is formed by the O(3)H groups only. The other ring oxygen O(5′) is linked by means of the weak component of a three-center bond from O(1′)H.

5. *LAMBIO*. Three separate infinite chains, one of which has a branch chain at the water oxygen which terminates at a ring oxygen O(5′). The other ring oxygen O(5) is bonded by a single link from O(4*A*)H. The water hydrogen atoms were not located and the bonding in that region is uncertain.

6. *MALTOS*01. Two infinite chains intersected by a finite chain at the water oxygen and at O(6)H. Both ring oxygens O(5), O(5′) are included by means of the weak *intramolecular* components of three-center bonds from the adjacent primary alcohol hydroxyls O(6)H and O(6′)H. The linkage oxygen O(1) is included by the weak *intramolecular* component of a three-center bond from O(2)H.

7. *MALTOT*. A finite chain, originating at O(3)H and terminating at the ring oxygen O(5′), with a separate link from the anomeric O(1′)H to ring oxygen O(5). The finite chain contains a four-bond loop by means of a three-center bond from O(6′)H. The link oxygen O(1) is included by means of the weak *intramolecular* component of a three-center bond.

8. *MELIBM*10. A finite chain beginning with the anomeric hydroxyl O(1*A*)H, forming a loop by intersecting with itself at the water molecule. It terminates with a three-center bond to the ring oxygen O(5) and the linkage oxygen O(6′). The minor anomeric β component in the crystal forms a separate bond to the ring oxygen O(5).

9. *MMALTS*. Four infinite chains, two of which intersect at the water molecule. The ring oxygen O(5) and the linkage oxygen O(1) are included by means of the weak component of three-center bonds, both of which are *intramolecular*.

10. *SOPROS*. Two infinite chains which intersect at the water oxygens, and separate infinite and finite chains. The finite chain originates at O(4′)H and terminates at a ring oxygen O(5′). The other ring oxygen O(5) is linked to the infinite chain by a weak intramolecular component of a three-center bond from HO(1′). The linkage oxygen O(1) is linked by two weak components of three-center bonds, one of which is intramolecular. Thus all the oxygen atoms are included. With the exception of the separate infinite chain, all the bonds are unusually weak with H···O distances greater than 2·0 Å.

11. *SUCROS*04. Without the weak interactions, this pattern is remarkably deficient in hydrogen bonding. The H---O bonds less than 2·0 Å form a finite chain which excludes both ring oxygens and the linkage oxygen and the hydroxyl O(4)H. When a weak four-center interaction from O(4)H and two weak components of three-center bonds are included, the pattern becomes a loop. There are two strong inter-residue intramolecular bonds O(1′)H to O(2) and O(6′)H to the ring oxygen O(5).

12. *TREHAL*10. Loop of strong hydrogen bonds which includes both water molecules. At O(*W*1), the loop intersects with an infinite chain formed by the O(2′)H groups. At O(*W*2), there is a junction to a finite chain which terminates at the ring oxygen O(5). The linkage oxygen O(1) is included by means of a weak *intramolecular* component of a three-center bond. The second ring oxygen O(5′) is excluded.

13. *TURANS*01. Infinite and finite chains are linked by the weak component of a three-center bond from O(2′)H. The finite chain terminates at O(2)H which is *not* involved as a hydrogen-bond donor. A second hydroxyl O(4)H is also almost non-bonding with a very weak interaction with the ring oxygen O(6′). Both these non- or weakly-bonded hydroxyl groups were identified with sharp solid-state IR absorption spectra at 3610 and 3550 cm^{-1}. The ring oxygens O(5) and O(6′) each receive two weak three-center bonds. The linkage oxygen O(1) is also bonded by a weak three-center bond.

14. *GAPRHM*10. An infinite chain and a finite chain. The finite chain terminates at the ring oxygen O(5). The linkage oxygen O(1) is included by means of a weak *intramolecular* component of a three-center bond.

15. *BDGPGL*. An infinite chain and a spiral, linked through a three-center bond from O(1′)H. The ring oxygen O(5) is omitted.

The hydrogen-bond lengths and angles

Since this data set contains the results of X-ray analyses of varying quality, there is a larger variation due to experimental error than in the previous analysis of Ceccarelli, Jeffrey & Taylor (1981), which was based on neutron diffraction data only.

The mean values for the alditols, the four main hydrogen-bonding schemes of the monosaccharides and the disaccharides, are shown in Table 4. With the exception of the values for type II, there is a definite trend towards longer and less linear bonds, as the bonding patterns include other arrangements than the infinite chains. The shortening of ~0·05 Å due to cooperativity is consistent with estimates of the magnitude of this effect based on theoretical calculations (Tse & Newton, 1977; Newton, 1983). These data did not permit any distinction between the hydrogen-bond lengths from anomeric and non-anomeric hydroxyl groups, which was also predicted and was observed from the neutron data.

Table 4. *Mean values for* H···O *bond lengths* (Å) *and* O–H···O *angles* (°) *for two-center bonds*

O–H bond lengths are normalized to 0·97 Å.

	Number of bonds	⟨H···O⟩	⟨O–H···O⟩
Alditols	46	1·807	166·5
Monosaccharides			
Type I	29	1·813	165·5
Type II	36	1·864	153·8
Type III	29	1·823	164·9
Type IV	34	1·830	165·9
Disaccharides	86	1·863	161·2
Neutron diffraction values*	74	1·818	167·1

* From Ceccarelli, Jeffrey & Taylor (1981).

The longer mean-bond-length value for type II systems is, in part, due to long bonds to acetal oxygens at the termination of the finite chains. When these are excluded, the mean value is 1·84 Å. The explanation for this value is that this data set contains *no* infinite chains, as do types I, III and IV.

The bonds in the disaccharide structures are significantly longer and less linear, owing undoubtedly to the more complex packing of these dumbbell-shaped molecules and the more common occurrence of hydrates (Jeffrey, Gress & Takagi, 1977), which put further constraints on the ability of the molecules to pack into favorable patterns for optimum hydrogen bonding.

There are 63 three-center bonds observed out of a total of 277 hydrogen bonds (excluding the sugar alcohols); this corresponds to the almost 25% proportion noted in the previous analysis. These three-center bonds range from the symmetrical to the unsymmetrical, as exemplified by the neutron diffraction data given in Table 2 of Ceccarelli *et al.* (1981). The geometry from the X-ray analyses is less reliable than for the two-center bonds, owing to a tendency for the analysts to bias the hydrogen positions in favor of two-center hydrogen bonds.

Conclusions

It is clearly an oversimplification in the crystal structures of carbohydrates to presume that the crystal structure will be that which allows the maximum number of two-center or linear hydrogen bonds possible. Two additional factors have to be considered. One is the importance of hydrogen-bond cooperativity. The second is that three-center (bifurcated) hydrogen bonds are energetically comparable to two-center bonds, despite the longer H···O bond distances. As a result, there is no way of predicting, from a structural survey alone, which of the four general patterns described will be adopted by a particular carbohydrate molecule. While it is reasonably certain that any new mono- or disaccharide crystal structure will have a hydrogen-bonding pattern that will fall within this classification, carbohydrates with functional groups other than hydroxyls will comply with different priorities. It should also be recognized that by selecting 'good quality' structure analyses, there is a bias towards those carbohydrates that form good crystals. Carbohydrates that crystallize poorly, and are excluded from this survey, may do so because molecular arrangements consistent with these rules are not possible.

This work is supported by grant GM-24526 from the National Institutes of Health.

References

ALLEN, F.H., BELLARD, S., BRICE, M.D., CARTWRIGHT, B.A., DOUBLEDAY, A., HIGGS, H., HUMMELINK, T., HUMMELINK-PETERS, B. G., KENNARD, O., MOTHERWELL, W. D. S., RODGERS, J. R. & WATSON, D. G. (1979). *Acta Cryst.* B**35**, 2331–2339.

ALLINGER, N. L. (1976). *Adv. Phys. Org. Chem.* **13**, 1–82.

AZARNIA, N., JEFFREY, G. A. & SHEN, M. S. (1972). *Acta Cryst.* B**28**, 1007–1013.

CECCARELLI, C., JEFFREY, G. A. & TAYLOR, R. (1981). *J. Mol. Struct.* **70**, 255–271.

HAMILTON, W. C. & IBERS, J. A. (1968). *Hydrogen Bonding in Solids*. New York: Benjamin.

JEFFREY, G. A. (1969). *Acc. Chem. Res.* **2**, 344–352.

JEFFREY, G. A., GRESS, M. E. & TAKAGI, S. (1977). *J. Am. Chem. Soc.* **99**, 609–611.

JEFFREY, G. A. & KIM, H. S. (1970). *Carbohydr. Res.* **14**, 207–216.

JEFFREY, G. A. & MALUSZYNSKA, H. (1982). *Int. J. Biol. Macromol.* **4**, 173–185.

JEFFREY, G. A. & SHIONO, R. (1977). *Acta Cryst.* B**33**, 2700–2701.

JEFFREY, G. A. & WOOD, R. A. (1982). *Carbohydr. Res.* **108**, 205–211.

KOLLMAN, P. A. & ALLEN, L. C. (1969). *J. Chem. Phys.* **51**, 3286–3293.

LINDNER, K. & SAENGER, W. (1982). *Acta Cryst.* B**38**, 203–210.

MALUSZYNSKA, H., KINOSHITA, Y. & JEFFREY, G. A. (1982). *Carbohydr. Res.* **100**, 17–28.

MALUSZYNSKA, H., RUBLE, J. R. & JEFFREY, G. A. (1981). *Carbohydr. Res.* **97**, 199–204.

NEWTON, M. D. (1983). *Acta Cryst.* B**39**, 104–113.

NEWTON, M. D., JEFFREY, G. A. & TAKAGI, S. (1979). *J. Am. Chem. Soc.* **101**, 1997–2002.

NORRESTAM, R., BOCK, K. & PEDERSEN, C. (1981). *Acta Cryst.* B**37**, 1265–1269.

PICKERING, S. U. (1893). *J. Chem. Soc.* **63**, 141–195.

PIMENTEL, G. C. & MCCLELLAN, A. L. (1960). *The Hydrogen Bond*, p. 100. San Francisco: Freeman.

TAKAGI, S. & JEFFREY, G. A. (1978*a*). *Acta Cryst.* B**34**, 1591–1596.

TAKAGI, S. & JEFFREY, G. A. (1978*b*). *Acta Cryst.* B**34**, 3104–3107.

TAKAGI, S. & JEFFREY, G. A. (1978*c*). *Acta Cryst.* B**34**, 2551–2555.

TAKAGI, S. & JEFFREY, G. A. (1979*a*). *Acta Cryst.* B**35**, 902–906.

TAKAGI, S. & JEFFREY, G. A. (1979*b*). *Acta Cryst.* B**35**, 1482–1486.

TSE, Y.-C. & NEWTON, M. D. (1977). *J. Am. Chem. Soc.* **99**, 611–613.

Acta Cryst. (1983). B**39**, 480–490

The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and Its Complexes with Inhibitors

BY MARKUS MARQUART, JOCHEN WALTER, JOHANN DEISENHOFER, WOLFRAM BODE AND ROBERT HUBER

Max-Planck Institut für Biochemie, D-8033 *Martinsried bei München, Federal Republic of Germany*

(*Received* 26 *August* 1982; *accepted* 8 *December* 1982)

Abstract

Sixteen protein crystal structures of the system bovine trypsin, trypsinogen, bovine basic pancreatic trypsin inhibitor (PTI) and Kazal inhibitor (porcine secretory inhibitor and Japanese quail ovomucoid), which had been refined by energy-restrained crystallographic refinement (EREF) at high resolution (1·4 to 1·9 Å), were analysed with respect to the geometries of the active site and the peptide groups in general. It was found that the conformation of Asp 102–His 57–Ser 195 is well conserved, irrespective of the different functional states of the enzyme. The His 57–Ser 195 hydrogen bond, however, improves considerably in the complexes as regards bond length and angle. The inhibitor binding at the main chain and side chain of P1 is also well conserved, including solvent molecules involved in the network of hydrogen bonds. The small-molecule inhibitor APPA (amidinophenyl pyruvate) bound to trypsin was also studied at high resolution. It binds by forming a tetrahedral adduct. The relation between the distance of the Ser 195 O_γ to the carbonyl C of the inhibitors and the pyramidalization at the carbonyl C obeys the Bürgi–Dunitz–Shefter relation for nucleophile–electrophile interaction. Analysis of the peptide groups suggests improvements of the standard geometry of several interbond angles. It also shows that peptide groups may be substantially non-planar and this non-planarity can be reliably analysed under the conditions of this study.

1. Introduction

Refinement of crystal structures of proteins requires geometrical constraints or restraints as the number of observed diffraction intensities is insufficient to determine atomic positions and 'thermal' parameters, even at 1·5 Å resolution.

The crystalline order of protein crystals often limits resolution to less than 2 Å. Consequently, geometrical restraints become even more important. With geo-

0108-7681/83/040480-11$01.50