Accurate *X*-ray Diffraction Analysis of Fibrous Polysaccharides containing Pyranose Rings. Part I. The Linked-atom Approach

By **Struther Arnott** * and **W. E. Scott**, Department of Biological Sciences, Purdue University, Lafayette, Indiana 47907, U.S.A.

X-Ray diffraction data from polymers are, by themselves, rarely sufficient to allow determination of atomic positions within a few tenths of an Ångstrom. When this accuracy is required, the polymer diffraction data have to be supplemented by stereochemical data from accurate analyses of relevant monomers. The linked-atom least-squares approach is a logical and convenient procedure for achieving the necessary synthesis of different types of data. The procedure has been used extensively in analysing polynucleotides and polypeptides but not, so far, polysaccharides. Its adaptation for these is described, with particular attention being paid to the precise conformations of rings that dominate the geometry of these polymers.

Since the determination of accurate structures of polysaccharides depends both on diffraction data from the polymers themselves and on additional data such as bond-lengths and bond-angles, we present a survey of the values of these latter quantities as determined by *X*-ray diffraction analysis of single crystals of appropriate small molecules.

As examples of the application of linked-atom least-squares methods we present and discuss the preparation of standard α- and β-D-glucose rings and preliminary molecular models of chitin and ι-carrageenan.

To correlate molecular geometry with, say, biological function, it will usually be sufficient that relative atomic positions are known to a precision of *ca.* 0·1 Å. This accuracy is necessary for the more intimate non-bonded and hydrogen-bonded contacts to be identified with certainty. For fibrous systems, the required accuracy is not generally achievable by analysis of *X*-ray diffraction intensities alone. This is mainly a consequence of the relatively low resolving power of *X*-ray data, even from crystalline fibres. James [1] has calculated that, if $d_{min.}$ is the minimum periodicity for which Fourier terms are included in a synthesis of electron density, detail on a scale less than $0{\cdot}715\ d_{min.}$ cannot be resolved. Since the perfect conditions appropriate to this calculation normally do not exist, $d_{min.}$ rather than $0{\cdot}715\ d_{min.}$ may be a better estimate of the resolving power of a set of diffraction amplitudes. In any event, for the systems we are considering $d_{min.} > 2{\cdot}0$ Å, usually, and therefore independent variation of atomic co-ordinates is certain not to be a meaningful exercise.

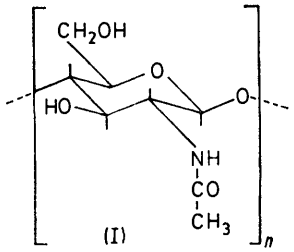
A molecular structure is as completely described by sets of bond-lengths, bond-angles, and conformation-angles as by a set of atomic co-ordinates. Bond-lengths in different structures do not usually differ to an extent that would affect the accuracy we seek. The same can be asserted (but less strongly) for bond-angles. Certainly conformation angles vary to a much greater extent from structure to structure. This hierarchy of molecular structural parameters prompted Arnott and Wonacott [2] to introduce, in 1966, a linked-atom description for polymer chains in which bond-lengths, bond-angles, and conformation-angles were explicit parameters. In their fibre diffraction analyses the bond-lengths had fixed, assigned values (determined by surveying accurate single-crystal *X*-ray analyses of relevant monomers), the bond-angles usually had fixed but might have variable values, and the conformation-angles had variable values determined by analysis of the polymer diffraction pattern. Such linked-atom descriptions of molecular chains, and the determination by least-squares methods of best values for the structural variables, has been applied very successfully in *X*-ray analyses of fibrous polypeptides [3,4] and polynucleotides.[5] Analogous procedures have been proposed and found highly productive for refining globular protein conformations.[6]

To justify our strategy and aid our analyses we have made a survey of the available, accurately determined structures of mono- and small oligosaccharides. We present below a discussion of their variability in the context of which polymer parameters have values that can be assigned initially and which have values that need to be determined from experiments with the molecular species in question. We then introduce the linked-atom method and the associated least-squares procedure for obtaining optimized parameter values. In the process we produce, as examples, standard pyranose rings and preliminary models for the molecular chains of chitin (the 1,4-linked polymer of 2-acetamido-2-deoxy-β-D-glucopyranose) (I) and of ι-carrageenan (II) [the 1,3-linked polymer of *O*-4-sulphate-β-D-galactopyranosyl(1 ⟶ 4)-*O*-3,6-anhydro-α-D-galactopyranose].

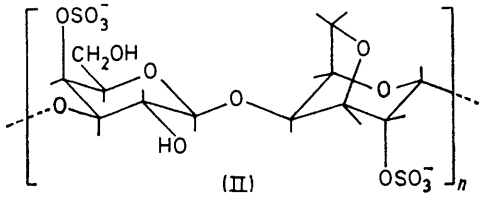
Survey of Sugar-residue Geometry.—In quantitative studies of polysaccharides conducted hitherto the residue geometries used have varied from investigator to investigator. Most workers have selected the results of one, or a few, single-crystal structure determinations on which to base their model-building efforts. We prefer to derive a standard sugar residue with bond-lengths, bond-angles, and ring conformation-angles that are averages from a large number of relevant carbohydrate structures. A

[1] R. W. James, ' The Optical Principles of the Diffraction of *X*-Rays,' Cornell University Press, Ithaca, 1965, p. 400.
[2] S. Arnott and A. Wonacott, *Polymer*, 1966, **7**, 157.
[3] S. Arnott and A. Wonacott, *J. Mol. Biol.*, 1966, **21**, 371.
[4] S. Arnott and S. D. Dover, *J. Mol. Biol.*, 1967, **30**, 209; S. Arnott, S. D. Dover, and A. Elliott, *J. Mol. Biol.*, 1967, **30**, 201; S. Arnott and S. D. Dover, *Acta Cryst.*, 1968, **B24**, 599; S. Arnott, in ' Symposium on Fibrous Proteins, Australia 1967,' ed. W. G. Crewther, Butterworths, Australia, 1968, p. 26.
[5] S. Arnott, S. D. Dover, and A. Wonacott, *Acta Cryst.*, 1969, **B25**, 2192.
[6] R. Diamond, *Acta Cryst.*, 1966, **21**, 253; R. Diamond, *Acta Cryst.*, 1969, **A25**, S189.

previous standard pyranose residue was derived [7] by averaging the co-ordinates of a number of structures (after they had been transformed to a common set of axes). However, our need is for an explicit investigation



(I)



(II)

of the variability of the different parameters of carbohydrate stereochemistry and to include the results of many recent structure determinations. Other surveys [8] have been made of some aspects of sugar-residue geometry, but none are comprehensive enough for our present purposes.

Included in this survey are the pyranose sugars in Table 1. By the criteria of published agreement indices and estimated standard deviations, their structures have been determined very accurately. We have therefore not thought it meaningful to weight different results differentially. Bond-lengths, bond-angles, and ring conformation-angles were computed for each of 27 pyranose residues. A histogram was prepared for each parameter (Figures 1—3). The mean, estimated standard deviation from the mean, and range are given in Table 2 for these. In certain instances—when the histograms show bimodal distributions—there are two entries.

The mean values of the lengths of particular C–C bonds are all close to the over-all mean of 1·523 Å. The mean length for C(5)–C(6) is noticeably shorter (1·514 Å) but since the estimated standard deviations are *ca.* 0·009 Å it is not clear that this difference is significant.

The mean C–O bond-lengths are all close to 1·426 Å except when either O(5) or O(1) is involved. The mean length for C(5)–O(5) is 1·436 Å. Bond-lengths in the O(5)–C(1)–O(1)–R system depend both on the configuration at C(1) and on the nature of R.[8] We find that when C(1)–O(1) is equatorial, its mean length is 1·389 Å and that of C(1)–O(5) is 1·429 Å, regardless of the nature of R. The same values are found when C(1)–O(1) is axial *and* R = H. However, for axial C(1)–O(1) and R ≠ H, the mean lengths of C(1)–O(1) and C(1)–O(5) are 1·415 and 1·414 Å respectively and essentially equal.

[7] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, in 'Aspects of Protein Structure,' ed. G. N. Ramachandran, Academic Press, New York, 1963, p. 121.

TABLE 1

Carbohydrate structures included in stereochemical survey

Compound	Residues used	Ref.
Methyl-β-maltoside	α-D- and β-D-Glucopyranosyl	*a*
Methyl-β-cellobioside, MeOH	β-D-Glucopyranosyl (2)	*b*
β-D-Cellobiose	β-D-Glucopyranosyl (2)	*c*
β-D-Glucose	β-D-Glucopyranosyl	*c*
Methyl-α-D-glucoside	α-D-Glucopyranosyl	*d*
α-D-Glucose	α-D-Glucopyranosyl	*e*
Raffinose	α-D-Galactopyranosyl α-D-Glucopyranosyl	*f*
α-D-Galactosamine-1-phosphate	α-D-Galactopyranosyl	*g*
β-D,L-Arabinose	β-D-Arabinopyranosyl	*h*
Methyl-β-D-xyloside	β-D-Xylopyranosyl	*i*
α-Lactose,H_2O	β-D-Galactopyranosyl α-D-Glucopyranosyl	*j*
Planteose	α-D-Glucopyranosyl α-D-Galactopyranosyl	*k*
(+)-Kestose	α-D-Glucopyranosyl	*l*
Trehalose	α-D-Glucopyranosyl (2)	*m*
Trehalose	α-D-Glucopyranosyl (2)	*n*
Methyl-α-D-altropyranoside	α-D-Altropyranosyl	*o*
Methyl-α-D-mannopyranoside	α-D-Mannopyranosyl	*p*
Methyl-α-D-galactoside, H_2O	α-D-Galactopyranosyl	*q*
Sucrose	α-D-Glucopyranosyl	*r*

[a] S. S. C. Chu and G. A. Jeffery, *Acta Cryst.*, 1967, **23**, 1038. [b] J. T. Ham and D. G. Williams, *Acta Cryst.*, 1970, **B26**, 1373. [c] S. S. C. Chu and G. A. Jeffery, *Acta Cryst.*, 1968, **B24**, 830. [d] H. M. Berman and S. H. Kim, *Acta Cryst.*, 1968, **B24**, 897. [e] G. M. Brown and H. A. Levy, *Science*, 1965, **147**, 1038. H. M. Berman, *Acta Cryst.*, 1970, **B26**, 290. [f] M. Sundaralingam and D. C. Fries, personal communication. [g] S. H. Kim and G. A. Jeffrey, *Acta Cryst.*, 1967, **22**, 537. [h] C. J. Brown, Sir G. Cox, and F. J. Llewellyn, *J. Chem. Soc.* (*A*), 1966, 922. [i] D. C. Fries, S. T. Rao, and M. Sundaralingam, *Acta Cryst.*, 1971, **B27**, 994. [k] D. C. Rohrer, personal communication. [l] G. A. Jeffrey and Y. J. Park, personal communication. [m] D. C. Rohrer, personal communication. [n] R. O. Gould, personal communication. [o] B. M. Gatehouse and B. J. Poppleton, *Acta Cryst.*, in the press. [p] B. M. Gatehouse and B. J. Poppleton, *Acta Cryst.*, 1970, **B26**, 1761. [q] B. M. Gatehouse and B. J. Poppleton, *Acta Cryst.*, 1971, **B27**, 654. [r] G. M. Brown and H. A. Levy *Science*, 1963, **141**, 921.

The ring bond-angles at carbon atoms are all very similar with mean values between 109·2° and 110·3°. The mean values of C(5)–O(5)–C(1) are larger [112·0° for C(1)–O(1) equatorial and 114·0° for C(1)–O(1) axial].

Bond-angles involving exocyclic atoms, with certain exceptions, fall into the surprisingly narrow range 108·4—110·8°. The angles involving C(6) have mean values outside this range (106·9°, 111·8°, and 112·7°). The angle O(5)–C(1)–O(1) is also exceptional in that it varies with configuration: when C(1)–O(1) is axial the mean value is 111·6°, when equatorial it is only 107·3°.

An important stereochemical feature in polysaccharide chains is the value of the glycosidic bond-angle. The values observed for this angle (Table 3) are correlated with the local substitution pattern. Therefore, although the angle may range from 111 to 122°, for similar substitutions the values lie in narrow ranges. For example,

[8] G. A. Jeffrey and R. D. Rosenstein, *Adv. in Carbohydrate Chem.*, 1964, **19**, 7; H. M. Berman, S. S. C. Chu, and G. A. Jeffrey, *Science*, 1967, **157**, 1576; M. Sundaralingam, *Biopolymers*, 1968, **6**, 189.

there are six entries between 115·7 and 117·6° for one of the most common substitution patterns listed in Table 3.

In contrast to the narrow ranges of values observed for all bond-angles, the *mean* values of the ring conformation-angles are distinctly different, varying from 53 to 62°. Moreover, the same angle in different structures may take any value in a range 16° wide.

that there is no basis for believing that the glycosidic bond-angles are inherently more variable between similar structures than the bond-angles within rings. On present evidence (Table 3) the contrary would appear to be true. For any given structure a value for the former would be predictable with a higher probability than for the latter.

TABLE 2

Results of stereochemical survey of pyranose sugar residues

	Parameter	Average value	Standard deviation	Range	Number of determinations
C–C Bond-lengths	C(1)–C(2)	1·523 (Å)	0·008 (Å)	1·510—1·538 (Å)	27
	C(2)–C(3)	1·521	0·007	1·508—1·536	27
	C(3)–C(4)	1·523	0·009	1·509—1·537	27
	C(4)–C(5)	1·525	0·008	1·511—1·539	27
	C(5)–C(6)	1·514	0·009	1·495—1·534	25
C–O Bond-lengths	C(1)–O(1)	1·389	0·011	1·376—1·397	12 [a]
		1·415	0·009	1·405—1·435	15 [b]
	C(2)–O(2)	1·423	0·008	1·411—1·440	26
	C(3)–O(3)	1·429	0·008	1·410—1·446	27
	C(4)–O(4)	1·426	0·010	1·409—1·446	27
	C(5)–O(5)	1·436	0·009	1·425—1·464	27
	C(6)–O(6)	1·427	0·008	1·415—1·442	24
	C(1)–O(5)	1·429	0·006	1·421—1·443	12 [a]
		1·414	0·009	1·392—1·428	15 [b]
Ring bond-angles	C(1)–C(2)–C(3)	110·5°	1·3°	108·3—113·9°	27
	C(2)–C(3)–C(4)	110·5	1·4	106·0—113·6	27
	C(3)–C(4)–C(5)	110·3	1·4	107·9—112·9	27
	C(4)–C(5)–O(5)	110·0	1·3	107·6—111·8	27
	C(5)–O(5)–C(1)	112·0	1·0	110·6—112·7	8 [c]
		114·0	0·4	113·2—114·7	19 [d]
	O(5)–C(1)–C(2)	109·2	1·1	107·4—112·3	26
Bond-angles involving pendant atoms	O(5)–C(1)–O(1)	107·3°	0·7°	105·9—108·3°	8 [c]
		111·6	0·9	109·8—112·7	19 [d]
	C(2)–C(1)–O(1)	108·4	1·9	106·2—110·4	27
	C(1)–C(2)–O(2)	109·3	1·8	105·1—112·1	27
	C(3)–C(2)–O(2)	110·8	1·8	106·4—113·2	27
	C(2)–C(3)–O(3)	109·6	1·8	105·0—112·9	27
	C(4)–C(3)–O(3)	109·7	1·7	106·5—112·5	27
	C(3)–C(4)–O(4)	110·4	1·3	107·8—112·4	27
	C(5)–C(4)–O(4)	108·6	1·9	105·8—112·2	27
	C(4)–C(5)–C(6)	112·7	1·4	109·8—115·3	25
	O(5)–C(5)–C(6)	106·9	0·5	106·8—107·9	25
	C(5)–C(6)–O(6)	111·8	1·2	109·4—113·8	23
Ring conformation-angles	O(5)–C(1)–C(2)–C(3) [e]	56·0°	4·0°	(47·1)—(63·5)°	23
	C(1)–C(2)–C(3)–C(4)	−53·2	3·3	(−43·9)—(−58·5)	23
	C(2)–C(3)–C(4)–C(5)	53·0	3·3	(48·0)—(60·5)	23
	C(3)–C(4)–C(5)–O(5)	−55·4	4·0	(−48·3)—(−61·4)	23
	C(4)–C(5)–O(5)–C(1)	61·1	3·8	(54·2)—(70·0)	23
	C(5)–O(5)–C(1)–C(2)	−62·2	3·3	(−56·0)—(−69·0)	23

[a] All structures except axial C(1)–O(1) glycosides. [b] Axial C(1)–O(1) glycosides. [c] Equatorial C(1)–O(1). [d] Axial C(1)–O(1). [e] Looking from atom 2 to atom 3, the clockwise rotation of bond 3-4 with reference to bond 2-1 is given.

There is a fortunate correspondence between the structural parameters observed to have very limited ranges and those parameters that have to be assigned fixed values in an *X*-ray analysis where data are of lowish resolution. The values of *bond-lengths* usually have an estimated standard deviation from the mean less than 0·01 Å. This is clear justification for assigning the mean bond-lengths as fixed values in polymers where the aim is a precision of 0·1 Å in atomic positions.

The greater variability of *bond-angles* is attested to by their estimated standard deviations from the mean (*ca.* 1·5°). While it might be preferable to have the polymer data define these angles, assigning the mean values in monomers as fixed values in a polymer chain is an acceptable first approximation. It should be noted

The broad ranges within which the *conformation-angles* fall, and the profound effect these parameters have on overall molecular geometry make it imperative that their values are determined for each polymeric structure, and therefore from experimental data derived from the polymer itself. Fortunately these are the parameters to which calculated *X*-ray intensities are sensitive, even for fibre diffraction data.

Standard Ring Parameters for α- and β-D-Glucose.—The linked-atom description of a molecular chain can be used to describe a ring system with, in particular, all conformation-angles as explicit parameters. As indicated schematically in Figure 4 combination of the mean bond-lengths, mean bond-angles, and mean conformation-angles of Table 2 does not necessarily lead to a closed

ring. We therefore consider generating a closed ring with standard bond-lengths and with bond-angles and conformation-angles deviating minimally (in a least-squares sense) from the mean values in monomers while conforming to the requirement of being compatible with

FIGURE 1

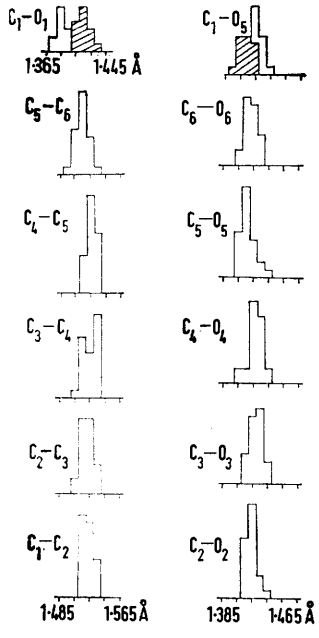


FIGURE 2

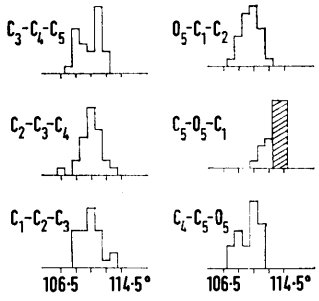


FIGURE 3

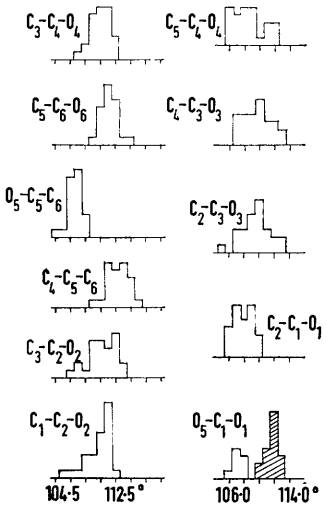


FIGURE 1 Histograms showing the distributions of values of bond-lengths in pyranosyl residues as determined by single-crystal *X*-ray diffraction analysis. In the cases of C(1)–O(1) and C(1)–O(5) the values are bimodally distributed in a manner correlated with whether C(1)–O–R is axial or equatorial. The shaded entries correspond to axial glycosides

FIGURE 2 Histograms of ring bond-angle values. The distributions have comparable widths, but that for C(5)–O(5)–C(1) can be sub-divided with the shaded entries corresponding to axial C(1)–O(1)

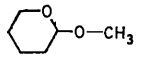
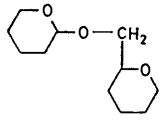
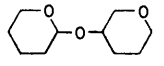
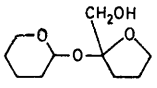
FIGURE 3 Histograms of exocyclic bond-angle values. For O(5)–C(1)–O(1) there is a bimodal distribution in one-to-one correspondence with axial (shaded) or equatorial configurations of C(1)–O(1)

a closed ring system. This same linked-atom least-squares approach with implicit, exact constraints is an essential ingredient of our strategy for accurate *X*-ray analyses. Its application to the relatively simple problem of obtaining standard, closed pyranose rings, to which are attached the immediately pendant atoms, is therefore of more general interest.

We start with the chain of nine atoms, O(5), C(1), · · ·, C(1′), C(2′) and five pendant atoms, O(1), O(2), O(3), O(4), and C(6) shown in Figure 4. The triangles formed by the first three and last three chain atoms will later be

TABLE 3

Glycosidic bond-angles

Type of glycosidic angle	Average value	Standard deviation	Range	Ref.*
	113·2°	0·1°	113·0—113·4°	*a, b, d, o, p, q*
	111·3	0·1	111·2—111·4	*f, k*
	116·5	0·7	115·7—117·6	*a, b, c, j, m, n*
	120·1	1·4	118·9—122·1	*f, k, l*

* All refer to footnotes to Table 1.

required to coincide. This way of achieving ring-closure also ensures that the six ring conformation-angles are explicit in our description.

To describe all the atomic positions of β-D-glucose in terms of some reference axial set and as functions of bond-lengths, bond-angles, and conformation-angles, we

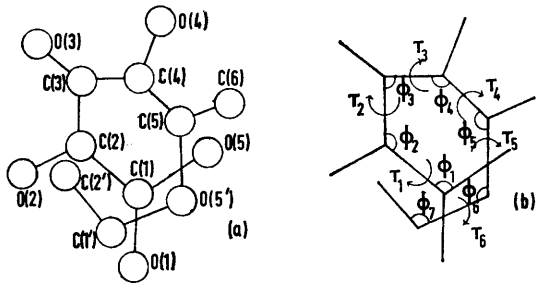


FIGURE 4 A linked-atom description of a pyranose ring with immediately substituent atoms. The triangle of atoms, C(2)C(1)O(5) is included twice so that all six ring conformation-angles (τ) are explicit. There is therefore a redundancy in bond-angles (ϕ) since $\phi_7 = \phi_1$. This is removed by the ring-closing conditions that cause triangles C(2′)C(1′)O(5′) and C(2)C(1)O(5) to be coincident

TABLE 4

Ring angles in standard β-D-glucose

Parameter	Average value	Model values	
ϕ[O(5)–C(1)–C(2)]	109·2	109·2 [a]	109·3 [b]
ϕ[C(1)–C(2)–C(3)]	110·5	110·5	110·5
ϕ[C(2)–C(3)–C(4)]	110·5	110·5	110·3
ϕ[C(3)–C(4)–C(5)]	110·3	110·3	110·2
ϕ[C(4)–C(5)–O(5)]	110·0	110·0	110·2
ϕ[C(5)–O(5)–C(1)]	112·0	112·0	112·3
* τ[O(5)–C(1)–C(2)–C(3)]	56·0	57·3	57·3
τ[C(1)–C(2)–C(3)–C(4)]	−53·2	−52·5	−53·2
τ[C(2)–C(3)–C(4)–C(5)]	53·0	52·6	52·2
τ[C(3)–C(4)–C(5)–O(5)]	−55·4	−57·6	−55·6
τ[C(4)–C(5)–O(5)–C(1)]	61·1	63·8	62·4
τ[C(5)–O(5)–C(1)–C(2)]	−62·2	−63·5	−62·8

* The definition of τ_{ijkl} is given at the foot of table 2.

[a] Bond-angles fixed. [b] Bond-angles varied.

consider that there exists at each atom, C(1) through C(1′), a right-handed, rectangular set of axes such that the X-axis at C(1) lies along C(1)–O(5) and the Y-axis lies in the plane O(5)–C(1)–C(2), within the obtuse angle; at C(2), the X-axis lies along C(2)–C(1) and the Y-axis in the plane C(3)–C(2)–C(1); and so on for the five remaining axial sets.

The co-ordinates of O(5) and O(1) in the axial set with C(1) as origin are (1·429, 0, 0) and (−0·4131, −0·6096, 1·1777). Corresponding co-ordinates referred to the axes at C(2) are obtained by the transformation

$$\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} = \begin{bmatrix} -\cos\phi_1 & \sin\phi_1\cos\tau_1 & \sin\phi_1\sin\tau_1 \\ -\sin\phi_1 & -\cos\phi_1\cos\tau_1 & -\cos\phi_1\sin\tau_1 \\ 0 & -\sin\tau_1 & \cos\tau_1 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} + \begin{bmatrix} 1{\cdot}523 \\ 0 \\ 0 \end{bmatrix} \quad (1)$$

say, $\boldsymbol{X}_2 = \boldsymbol{A}_2\boldsymbol{X}_1 + \boldsymbol{L}_2$.

As well as the atoms with co-ordinates that depend on ϕ_1 and τ_1 (defined in Figure 4b), C(1) and O(2) have constant co-ordinates in the C(2) axial set, *viz.* (1·523, 0, 0) and (−0·4703, −0·7158, −1·1364), and can be added to the list of known co-ordinates. Successive transformations of atomic co-ordinates and additions of new atoms to the list can be continued until the whole set is referred to the axes at C(1′). In this system the co-ordinates of C(1′) and O(5′) are simply (0, 0, 0) and (1·429, 0, 0). The other co-ordinates have varying degrees of dependence on the values of bond-lengths, bond-angles, and conformation-angles. For example, the co-ordinates of C(2′) depend only on the length of C(1′)–C(2′) and the values of ϕ_7, but the co-ordinates of O(5) depend on 7 bond-lengths, 7 bond-angles, and 6 conformation-angles.

When the mean values of the structural parameters for β-D-glucose (Table 2) are used, we find X[O(5′)] − X[O(5)] = 0·017 Å, Y[O(5′)] − Y[O(5)] = 0·003 Å, Z[O(5′)] − Z[O(5)] = −0·009 Å, X[C(1′)] − X[C(1)] = −0·010 Å, Y[C(1′)] − Y[C(1)] = 0·043 Å, Z[C(1′)] − Z[C(1)] = −0·000 Å, X[C(2′)] − X[C(2)] = −0·011 Å, Y[C(2′)] − Y[C(2)] = 0·026 Å, Z[C(2′)] − Z[C(2)] = 0·032 Å. Clearly, for a closed ring system we would want all of these nine differences to be zero. This requirement imposes nine constraints on the molecular parameters. When the bond-lengths and bond-angles in triangles O(5′)C(1′)C(2′) and O(5)C(1)C(2) are fixed, only six of the nine constraints are independent. For fixed bond-lengths but variable bond-angles, seven of the nine are needed.

A convenient solution to this problem is to find that set of molecular parameters $\{D_p\}$ that differs from the standard set $\{{}_oD_p\}$ in a way that minimizes

$$\Sigma k_p({}_oD_p - D_p)^2 + \Sigma\lambda_h G_h = \Omega + \Lambda, \text{ say} \quad (2)$$

where k_p is the inverse of the square of the estimated standard deviation of ${}_oD_p$, the G_h are the six or seven independent co-ordinate relationships that have all to be zero finally, and $\{\lambda_h\}$ are the (initially undetermined) Lagrange multipliers.

The variability in bond-lengths is so small that we always regard them as constants. As variable D_p we then might have the 13 values ϕ_1 through ϕ_7 and τ_1 through τ_6 (with the expectation that the τ will deviate more from standard values than the ϕ, since $k_\tau \sim 0{\cdot}1\ k_\phi$), or the six variables, τ_1 through τ_6, if bond-angles as well as bond-lengths are regarded as constants. In this situation the six variables are subject to six constraints and therefore there is no ' best ' set of D_p in the conventional least-squares sense, but a unique set near $\{{}_oD_p\}$ (from Table 2). This unique set is shown in Table 4 together with the least-squares, best set that minimizes $(\Omega + \Lambda)$ when all ϕ and τ are co-variable. In the latter the maximum difference in any bond-angle from the monomer mean is 0·3° for ϕ_6, and in conformation-angles 1·3° for τ_1 and τ_5. The Cartesian co-ordinates for the atoms in β- and α-D-glucose rings, developed in this fashion are in Tables 5 and 6.

TABLE 5

Co-ordinates [a] for a standard β-D-glucose residue obtained by an $(\Omega + \Lambda)$ minimization with both ring bond-angles and conformation-angles as variable parameters

Atom	X (Å)	Y (Å)	Z (Å)
C(1)	0·0000	0·0000	0·0000
C(2)	−0·5044	1·4371	0·0000
C(3)	0·0412	2·1939	−1·2012
C(4)	1·5585	2·0739	−1·2567
C(5)	1·9732	0·6083	−1·1815
C(6)	3·4751	0·4287	−1·1171
O(1)	−0·4131	−0·6096	1·1777
O(2)	−1·9274	1·4361	−0·0071
O(3)	−0·3301	3·5707	−1·1094
O(4)	2·0566	2·6257	−2·4736
O(5)	1·4290	0·0000	0·0000
H(1) [b]	−0·3672	−0·5180	−0·8982
H(2)	−0·1851	1·9382	0·9257
H(3)	−0·3936	1·7814	−2·1236
H(4)	2·0014	2·6246	−0·4137
H(5)	1·6051	0·0764	−2·0712
O(6) [c]	4·0335	1·0468	0·0415

[a] Referred to right-handed, rectangular axes with C(1) as origin, X-axis along C(1)O(5) and Y-axis in the O(5)C(1)C(2) plane (and within the obtuse angle). [b] The ring hydrogens were placed with carbon-hydrogen distances of 1·1 Å and angles with the two adjacent ring bonds of 109·5°. [c] The hydroxymethyl group was assigned the *gg* conformation.

Linked-atom Descriptions of Polysaccharide Chains.—We have chosen chitin and ι-carrageenan to illustrate a number of features that arise in the application of our linked-atom approach to polysaccharide structures. Chitin (I) is a β-1,4-linked homopolymer in which, for the moment, we will assume that a single 2-acetamido-2-deoxyglucopyranosyl residue is the asymmetric unit of the molecular chain. ι-Carrageenan (II) has two residues in its chain asymmetric unit. It is an alternating copolymer, $(-B-A-)_n$, where B = β-D-galactosyl-4-sulphate, A = 3,6-anhydro-α-D-galactosyl-2-sulphate. The B–A linkage is 1,4-, the A–B linkage 1,3.

Chitin.[9]—We assume initially that chitin has the

[9] D. Carlstrom, *J. Biophys. and Biochem. Cytol.*, 1957, **3**, 669; J. Blackwell, *Biopolymers*, 1969, **7**, 281.

fixed, standard glucopyranose ring derived earlier (Table 5), and that the orientations of the hydroxymethyl and acetamido-substituents relative to this ring are fixed. The simplest linked-atom description would then involve a chain of six atoms, O(4), C(4′), C(1), O(1), C(4), and C(3), with pendant atoms as in Figure 5a. (One might omit the side-chains until the backbone conformation is established.) As a consequence of the fixed ring geometry, the C(4′)–C(1) ' bond ' will have an assigned length like other bonds. For the same reason the ' bond-angles ' O(4)–C(4′)–C(1), C(4′)–C(1)–O(1), and O(1)–C(4)–C(3), and the conformation-angle τ_1, will also have fixed values. The remaining conformational parameters that might be variables are the bond-angle, ϕ_3, and conformation-angles τ_2 and τ_3.

TABLE 6

Co-ordinates [a] for a standard α-D-glucose residue obtained from an ($\Omega + \Lambda$) minimization with both ring bond-angles and conformation-angles as variable parameters

Atom	X (Å)	Y (Å)	Z (Å)
C(1)	0·0000	0·0000	0·0000
C(2)	−0·5011	1·4382	0·0000
C(3)	0·0594	2·1970	−1·1931
C(4)	1·5775	2·0814	−1·2312
C(5)	1·9958	0·6167	−1·1590
C(6)	3·4953	0·4423	−1·0444
O(1)	−0·5209	−0·6545	−1·1413
O(2)	−1·9239	1·4405	−0·0227
O(3)	−0·3169	3·5728	−1·1059
O(4)	2·0885	2·6414	−2·4390
O(5)	1·4140	0·0000	0·0000
H(1) [b]	−0·3672	−0·5168	0·8989
H(2)	−0·1900	1·9355	0·9305
H(3)	−0·3636	1·7857	−2·1215
H(4)	2·0089	2·6289	−0·3802
H(5)	1·6549	0·0925	−2·0640
O(6) [c]	4·0116	1·0560	0·1359

[a] Referred to right-handed, rectangular axes with C(1) as origin, X-axis alone C(1)O(5) and Y-axis in the O(5)C(1)C(2) plane (and within the obtuse angle). [b] The ring hydrogens were placed with carbon-hydrogen distances of 1·1 Å and angles with the two adjacent ring bonds of 109·5°. [c] The hydroxymethyl group was assigned the *gg* conformation.

When considering helical polymers, we need to introduce four further parameters that would define the relationship of the sugar residue to the helix axial set. Let us assume that the final, reference set of molecular axes is at C(4), and that the helix origin (O_h) is the projection of C(4) on the helix axis (Figure 5a). The (right-handed, rectangular) helix axial set is completely defined by having O_hZ_h coincide with the helix axis and O_hX_h with O_hC(4). The four additional parameters of interest are then the ' helix radius ', r[C(4)] = O_hC(4), and the three Eulerian angles, θ_x, θ_y, and θ_z, that define the orientation of the molecular axial set at C(4) with respect to the helix axial set.

When an X-ray diffraction diagram is available, the rotation per residue, $(2\pi/N)$, and the translation per residue, (c/N), would usually be determinable from preliminary analysis. It follows that the seven variable parameters r, θ_x, θ_y, θ_z, τ_2, τ_3, ϕ_3 and the fixed parameters would have to conform to nine relationships between the cylindrical polar co-ordinates of O(4) and O(1), of C(4′) and C(4), and of C(3′) and C(3). As an example, the three relationships involving O(4) and O(1) would be: $G_1 = r$[O(4)] − r[O(1)] = 0, $G_2 = \theta$[O(4)] − θ[O(1)] $-2\pi/N = 0$, $G_3 = Z$[O(4)] − Z[O(1)] − $c/N = 0$. Since the triangles O(4), C(4′), C(3′) and O(1), C(4), C(3) are congruent, only six of the nine relationships are independent. In the helix residue there therefore would be one residual degree of freedom.

Although most economical in terms of number of parameters, this description of the chitin chain suffers from a serious disadvantage: the glycosidic bond-angle has been arbitrarily selected as a variable in unjustifiable preference to the ring bond-angles, variation of which would also permit variation of the ring conformation-angles. Therefore, for backbone model building purposes, we have preferred to set up a more elaborate model for chitin in which the ring bond-angles and conformation-angles, the glycosidic bond-angle, and glycosidic conformation-angles are all considered to be potentially variable. Figure 5b shows how this description can be achieved in a continuous linked-atom system. The main chain now has 12 atoms, C(3″), C(2′), C(1′), · · ·, C(4), C(3) and the other atoms are considered to be pendant. As before, several atoms are included twice in order to formulate constraints in terms of their co-ordinates.

Seven independent atomic co-ordinate constraints relating C(1′), C(2′), C(3″) to C(1), C(2), C(3′) are necessary for ring closure because bond-angles are being varied. Seven further constraints relating the co-ordinates of triangles O(4)C(4′)C(3′) and O(1)C(4)C(3) and the helix constants are required. Finally, since there is a redundancy in using both τ_1 and τ_7, a further constraint must relate these two parameters, giving a total of 15 constraints.

The 23 variables ($9\tau + 10\phi + 3\theta_E + 1r$) would therefore be subject to 15 exact, but implicit relationships, G_h, that would reduce the number of degrees of freedom in the system to 8. We would need to know preferred values for only 9 of the variables for a constrained, least-squares best set of parameters to be obtainable. Normally there would be preferred values for 19 parameters (ring bond-angles and conformation-angles, glycosidic bond-angle, and glycosidic conformation-angles, ϕ and ψ).

Where (as in this case and most others) there have been insufficient relevant experiments with oligomers to allow a survey to provide preferred values and variances for the two conformation-angles at the glycosidic linkage the following procedure might be used. The preferred values of ϕ and ψ would correspond to a minimum energy position calculated with fixed ring geometry but without the constraints of prescribed helix dimensions. The values of ϕ and ψ would normally have wide ranges (say, $\pm 16°$) within which there would be no significant increase in conformational energy. Inspection of Table 2 shows that the *limits* of variability of bond-angles and conformation-angles are *ca.* $\pm 2\sigma$ (*i.e.* twice the estimated standard deviation) from the average value. One might

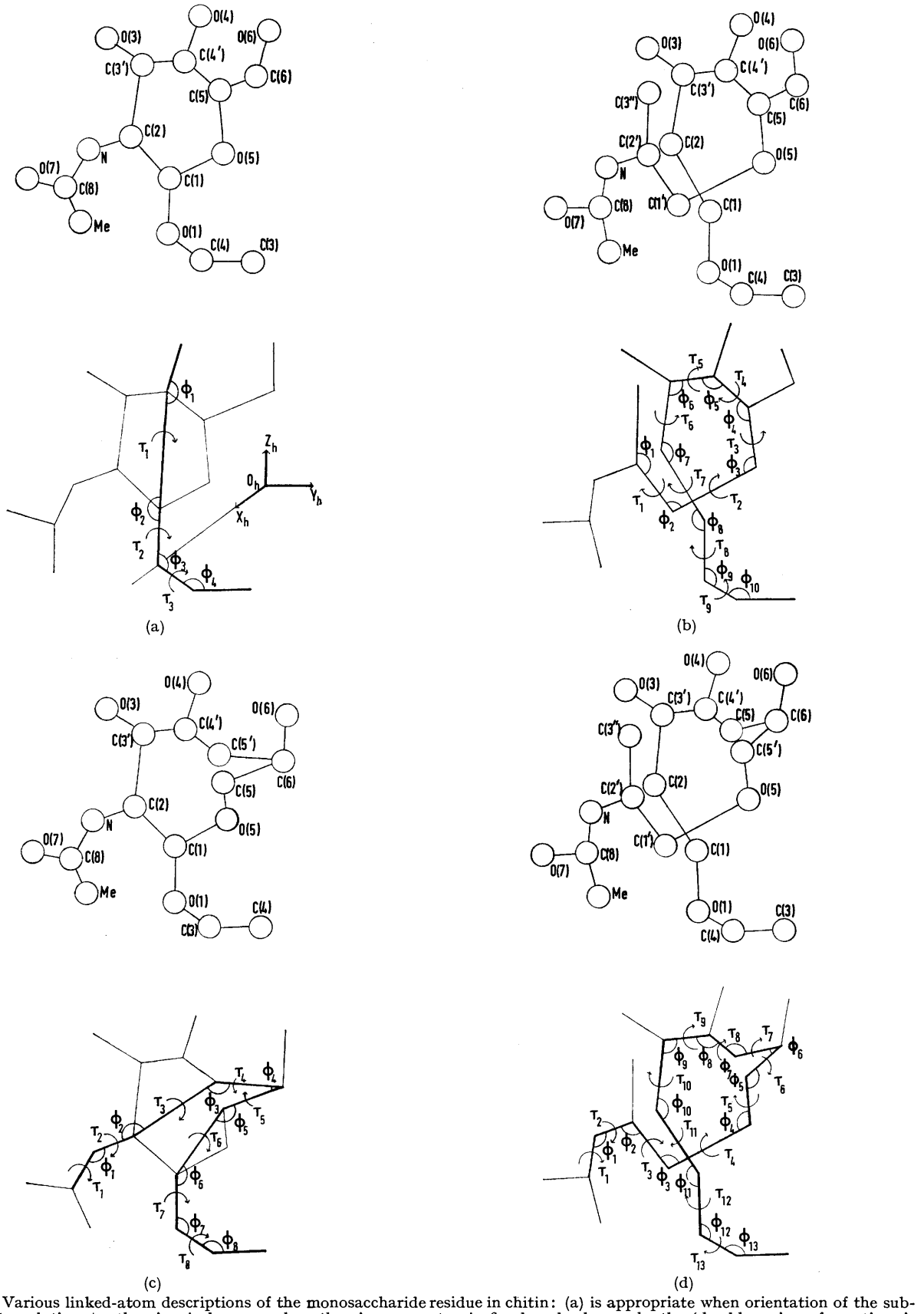


FIGURE 5 Various linked-atom descriptions of the monosaccharide residue in chitin: (a) is appropriate when orientation of the substituents relative to the ring is known, when the ring geometry is fixed and when only the 'backbone' conformation is being considered; (b) can be used when variation of the conformation of the ring is considered necessary; in (c) the side chain orientations are explicit parameters; (d) shows explicit parameterisation of all ring and side-chain conformation-angles. Continuous linked-atom descriptions frequently require the introduction of redundant parameters such as τ_4 and τ_5 in (c), τ_6 and τ_7, τ_8 and τ_{11} in (d)

therefore consider for both ϕ and ψ that 16° would be an adequate estimate of 2σ so that the appropriate weight for the corresponding term in equation (2) would be $(\frac{1}{3})^2$, *i.e.* between four and six times smaller than for the terms involving ring conformation-angles.

In the example considered by us, an ($\Omega + \Lambda$) minimization was carried out with 23 variables, 19 terms in Ω and 15 constraints in Λ corresponding to $N = 2$ (*i.e.* a two-fold helix for the chitin molecule) and $c/N = 5{\cdot}14$ Å. The preferred values for the sugar ring parameters and the glycosidic bond-angle were derived from the stereochemical survey. We chose the values (−40°, 140°) recently reported [10] for the Hermans bent chain conformation for cellulose. The fixed geometry of the *trans*, planar acetamido-group is the same as in 2-acetamido-2-deoxy-α-D-glucose.[11] Its orientation corresponds to bond N–C(8) being 180° away from the *gauche–gauche* conformation relative to ring bonds C(1)–C(2) and C(2)–C(3). The hydroxymethyl side-chain was assumed to be in the *gauche–gauche* conformation. Cylindrical polar co-ordinates for the complete model are in Table 7.

TABLE 7

Cylindrical polar co-ordinates of a chitin model

Atom	r (Å)	θ (°)	Z (Å)
C(1)	0·478	171·8	−2·391
C(2)	1·504	145·1	−3·447
C(3)	1·103	169·2	−4·813
C(4)	0·390	0·0	−5·140
C(5)	1·311	−34·2	−4·000
C(6)	2·723	−18·9	−4·211
O(1)	0·933	124·6	−1·193
O(3)	1·981	148·0	−5·799
O(4)	0·933	−55·7	−6·333
O(5)	0·902	−1·5	−2·767
O(6) *	3·234	7·0	−4·291
N	2·827	160·9	−3·099
C(7)	3·953	147·3	−3·203
O(7)	4·218	131·3	−3·571
C(8)	5·188	156·7	−2·820

* O(6) Is in the *gauche–gauche* position.

ι-Carrageenan.[12]—As with chitin, if attention is to be concentrated on the arrangement of the backbone, it is simplest first to consider a rigid-ring system with the substituent chains in fixed orientations relative to these rings. This would require a nine-atom main chain (Figure 6b) including O(1), C(1′), C(4), and O(4) of the anhydrogalactosyl moiety, followed by C(1), C(3), O(3) of the galactosyl moiety, and C(1) and C(2) of the anhydrogalactosyl in the next residue. The conformations about C(4)–O(4), O(4)–C(1), C(3)–O(3), and O(3)–C(1), the Eulerian angles and the helix radius would provide $4 + 3 + 1$ variables. Six helix-making constraints would reduce the number of degrees of freedom to 2, but this would allow the backbone conformation of the disaccharide residue still to be continuously variable without the need for any bond-angle to be added to the parameter list. The shape and position of this disaccharide residue would therefore have $4 + 1 + 3 - 6 = 2$ residual degrees of freedom without resorting to bond-angles as variable parameters.

Both from the point of view of stereochemical interactions and from the need, in an *X*-ray analysis, accurately to site the relatively electron-dense sulphate groups, it is desirable that explicit variation of their orientation be included at an early stage in any model. A 21 atom chain with 7 (Figure 6c) variable conformation-angles (τ_2, τ_5, τ_6, τ_{12}, τ_{15}, τ_{17}, and τ_{18}) would achieve this. The redundancy in the description (τ_{12}, τ_{15}) would be removed by adding one constraint to the six helix-making relationships so that this system would have $7 + 3 + 1 - 6 - 1 = 4$ degrees of freedom, *i.e.* the desired 2 in the sulphate groups in addition to the 2 of the first chain description. (Note that no constraint need be added to remove the redundancies of τ_9 and τ_{10} and of τ_{13} and τ_{14} while these conformation-angles are not being treated as variable parameters.) The cylindrical polar co-ordinates for a model derived in this manner are given in Table 8.

TABLE 8

Cylindrical polar co-ordinates for an ι-carrageenan model

Atom	r (Å)	θ (°)	Z (Å)
Galactose residue			
C(1)	1·912	51·9	4·296
C(2)	2·183	52·6	2·798
C(3)	2·417	17·6	2·198
C(4)	3·700	11·8	2·960
C(5)	3·421	10·0	4·455
C(6)	4·689	8·9	5·279
O(2)	1·630	88·2	2·177
O(3)	2·740	22·5	0·824
O(4)	4·736	25·0	2·747
O(5)	2·937	33·1	4·919
O(6) *	4·395	7·0	6·667
S	6·103	19·8	2·068
O(I)	7·068	29·3	1·985
O(II)	5·875	14·8	0·735
O(III)	6·819	11·2	2·867
3,6-Anhydrogalactose residue			
C(1)	2·863	120·0	8·667
C(2)	3·650	128·9	7·452
C(3)	3·590	112·6	6·315
C(4)	2·229	100·0	5·965
C(5)	2·372	77·8	7·201
C(6)	3·880	80·2	7·321
O(1)	2·740	142·5	9·490
O(2)	3·478	150·6	6·968
O(3)	4·421	97·4	6·804
O(4)	2·550	81·3	4·819
O(5)	1·759	97·7	8·309
S	4·931	159·8	6·896
O(I)	5·102	175·1	6·368
O(II)	5·509	159·8	8·226
O(III)	5·787	151·3	6·031

* O(6) Is in the *gauche–gauche* position.

For this model the structure parameters for 3,6-anhydro-α-D-galactose were taken from the single-crystal *X*-ray determination.[13] The 1,3-linked galactose residue was derived by inverting the configuration at C(4) in the standard β-D-glucose residue.

All the angles in the sulphate groups were assumed to

[10] D. A. Rees and R. J. Skerrett, *Carbohydrate Res.*, 1968, **7**, 334; P. R. Sundararajan and V. S. R. Rao, *Biopolymers*, 1969, **8**, 305; D. A. Rees and W. E. Scott, *J. Chem. Soc.* (*B*), 1971, 469.

[11] L. N. Johnson, *Acta Cryst.*, 1966, **21**, 885.

[12] N. S. Anderson, J. W. B. Campbell, M. M. Harding, D. A. Rees, and J. W. B. Samuel, *J. Mol. Biol.*, 1969, **45**, 85.

[13] J. W. B. Campbell, personal communication.

be tetrahedral. The S–O ester bond was assigned the value of 1·60 Å, and the other S–O bonds were fixed at 1·45 Å. The conformations of both sulphate groups have the C–H and O–S bonds eclipsed about bond C–O, and the C–O and one S–O bond *trans* to bond O–S.

The preferred values for τ_5, τ_6, τ_{17}, τ_{18} were from Model H of Anderson *et al.*[12]

An elaborate linked-atom description of ι-carrageenan that also has the ring conformation-angles as explicit variables (and therefore necessarily would have the ring-bond angles as variables) is possible in a way analogous to chitin (Figure 5d). Such a description would require a more inflated parameter list than is likely to be necessary for the structural accuracy we seek.

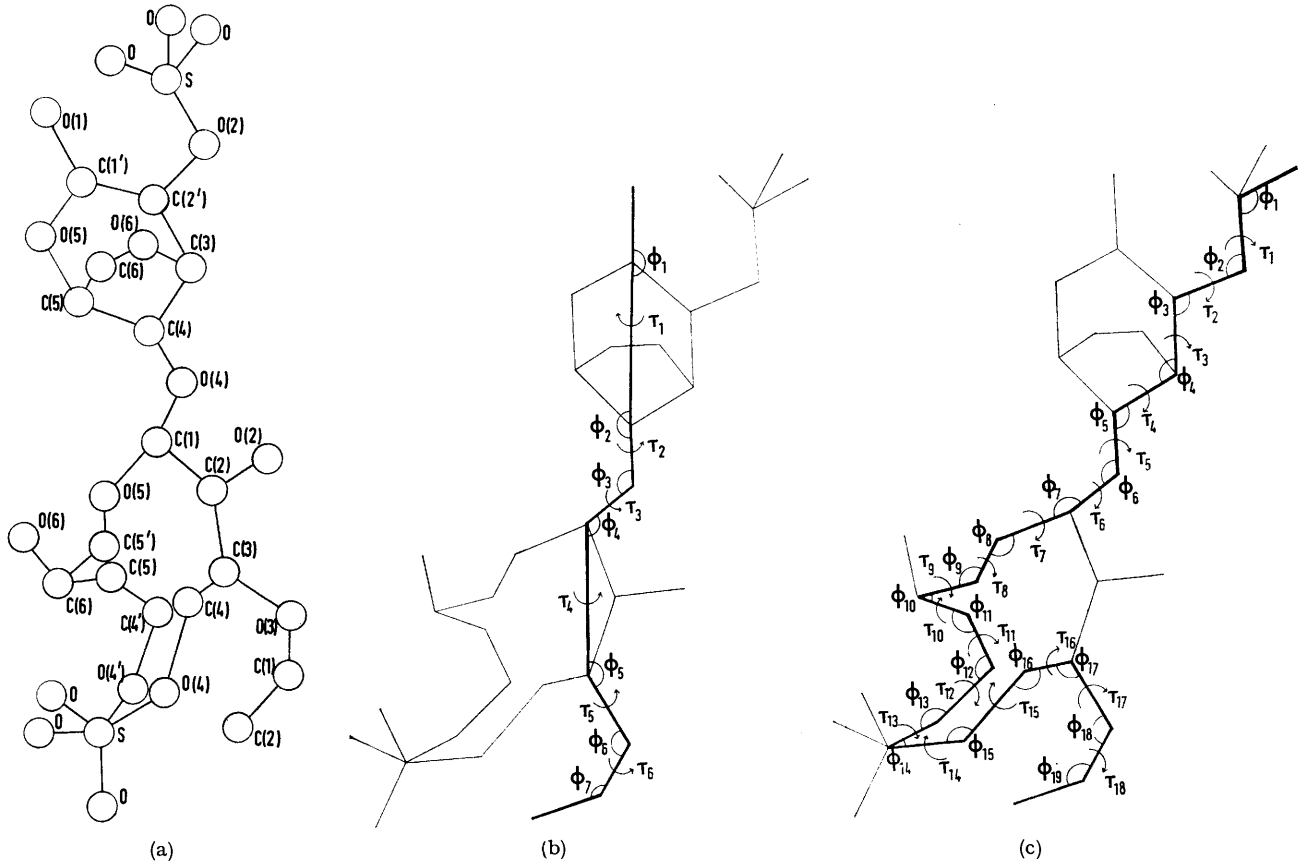


FIGURE 6 Two linked-atom descriptions of the disaccharide residue of ι carrageenan: (a) atom names; (b) a scheme that concentrates on parameters defining the ' backbone;' (c) a more elaborate scheme that has the sulphate and hydroxymethyl group orientations as explicit parameters

Strategy of X-Ray Analysis.—The observed diffraction pattern depends not only on the conformation of the diffracting molecules but also on the way they are packed together. Both packing and conformation need to be determined in any quantitative analysis of the *X*-ray data. In many cases a detailed, and essentially correct, *molecular* model can be developed from steric considerations alone and discovery of the packing scheme is the main structural problem. This was the case for α-poly-L-alanine fibres where the α-helix polypeptide model of Pauling and Corey[14] was a good solution of the molecular structure problem, but a successful fitting of the *X*-ray data and refinement of the complete structure[3] followed only when the packing was understood to be an hexagonally packed array of helices with each molecular site occupied by either an ' up-pointing ' or ' down-pointing ' molecule.[3,15] (This form of disorder is common when linear macromolecules with sensed chains crystallize). Nevertheless, at this time we will be concerned only with the choice of *molecular* structural variables.

We will speak of *assigning* values to parameters where it is doubtful whether they can be determined by analysing the *X*-ray data, or where their values are virtually invariant because of steric considerations and known with high precision. Bond-lengths are the structural variables that most obviously have to be assigned values for both these reasons. Other parameters may be *determinable* and *refinable*.

In the *X*-ray patterns obtained by us for both α-chitin

[14] L. Pauling and R. B. Corey, *Proc. Nat. Acad. Sci.*, 1951, **37**, 235.

[15] A. Elliott and B. R. Malcolm, *Proc. Roy. Soc.*, 1959, *A* **249**, 30.

and ι-carrageenan $d_{\text{min.}} \sim 2$ Å and there are about 30 reflections in each. The unit-cell dimensions have been determined within a few hundredths of an Ångstrom so that, in particular, the axial periodicity of each molecule is accurately known. Systematically absent reflections clearly indicate a three-fold screw axis for ι-carrageenan and possibly a two-fold axis for chitin. The nine constraints that would fit each molecular asymmetric unit appropriately into its helical chain can therefore be defined exactly.

We now consider the molecular parameters whose values the *X*-ray data might be capable of defining and those parameters that have to be assigned values from other considerations.

(1) The group parameters r, θ_x, θ_y, θ_z are peculiar to each structure and have to be determined and refined in each case. Since they are parameters that affect the whole chain unit, the calculated *X*-ray structure factors are relatively sensitive to changes in them. Their values are therefore determinable with high accuracy from the observed intensities.

(2) The conformation-angles of the glycosidic bonds are major structure parameters that data of about 2 Å resolving power should also be capable of defining accurately.

(3) The conformation-angles in the side chains would have to be assigned initially, except where their variation would change the position of a large group with a substantial electron content. A rotation about the C(5)–C(6) bond of an hydroxymethyl group would vary the position of only the hydroxy-group. If the position of this group could not be defined accurately by the *X*-ray intensities, then its value would have to be assigned after considering the need for a conformation of low energy and the possibility of making hydrogen-bonds. In contrast, rotations about a C–O bond leading to a sulphate group would vary the position of this dimensionally well-defined and electron-dense group and might therefore confidently be expected to be a refinable parameter. Rotation about the C–N bond of the acetamido-group in chitin would also be of this kind, but precisely how sensitive the *X*-ray data would be to its value would have to be determined by experiment.

(4) Changes in ring conformation-angles would certainly have subtle but important effects on the detailed geometry of the chain. Since there are only six, and the ring-closing condition requires at least six constraints, conformation-angles are not refinable unless the ring bond-angles are refined also.

(5) Refinement of the ring bond-angles may be possible with a data set where $d_{\text{min.}} \sim 2$ Å, but whether meaningful refinement to values near those in our survey would be possible can be determined only by experiment with the data set in question.

(6) Refinement of the glycosidic bond-angle would be justifiable only in the same circumstances that refinement of the ring bond-angles would also be undertaken (since refinement of the ring conformation-angles would then become possible). Otherwise, like them, it would be assigned a fixed value with (as we have indicated) no less certainty than any other bond-angle.

(7) The remaining structural variables are the bond-lengths and there is no question but that they have to be assigned values.

In fibre diffraction analyses it has not usually been found possible to solve the *X*-ray phase problem experimentally and thence the molecular structure, in contrast to the isomorphous replacement approach to globular protein structures. More usually, general stereochemical information is mustered in a preliminary model or a small number of different models, and a process of refinement and adjudication follows. (When different models have to be considered, parallel refinements have to be undertaken until considerations of stereochemistry, or the fit with the *X*-ray data, or, hopefully, both permit elimination of all but one model.)

The first problem is to reconcile the backbone of the model with the precise helical characteristics determined from the *X*-ray data. At this stage side-chains are irrelevant and for ι-carrageenan an appropriate model is shown in Figure 6b, with assigned values for all parameters except r, θ_x, θ_y, θ_z, τ_2, τ_3, τ_5, τ_6. The molecular chain would then have 8 variable parameters subject to 6 helix-making constraints, leaving 2 degrees of structural freedom. Four observational equations could be set up to retain the values of τ_2, τ_3, τ_5, τ_6 near some initial set, and values of all 8 variables obtained by an $(\Omega + \Lambda)$ minimization. It would not be important for the standard values of τ_2, τ_3, τ_5, and τ_6 to be known accurately since we would expect the subsequent refinement against *X*-ray intensities to define them more precisely. The main purpose of this initial model building exercise is to obtain a set of molecular parameters more likely to lead to smooth convergence during the *X*-ray refinement. If it turned out that a number of sets of (τ_2, τ_3, τ_5, τ_6) would appear suitable starting points, it would be necessary to continue parallel consideration of several models.

On completion of the backbone a complete molecular model (as in Figure 6c) would then be constructed with which to begin refinement that would take into account the agreement with the observed *X*-ray *intensities*.* In the Figure 6c description, the arguments in (1), (2), and (3) above would indicate that r, θ_x, θ_y, θ_z, τ_2, τ_5, τ_6, τ_{12} (and τ_{15}), τ_{17}, τ_{18} should be variables from the beginning. These 11 parameters would be made subject to six helix-making constraints and a seventh relationship connecting τ_{12} and τ_{15}. These would give rise to a seven-component Lagrange multiplier expression, Λ, and we would wish to minimize in least-squares fashion $\Theta = (\Phi + \Lambda)$ where

$$\Phi = \sum_1^M \omega_m\{{}_oF_m - (1/K)F_m \exp(-B\rho^2/4)\}^2 \quad (3)$$

* We wish in this paper to concentrate on the problem of refining *molecular* conformations and assume that the molecular packing parameters, on which the magnitudes of *X*-ray intensities also depend, are known. In some structures the determination of the packing arrangement may present the greater difficulty, but discussion of this is best postponed until particular structure determinations are reported.

M is the number of observed structure amplitudes, ${}_{o}F_{m}$, for which F_m is the corresponding calculated quantity; K is a scale factor; B is a parameter correlating the observed and calculated average variation of radial intensity; ω_m is a measure of the reliability of ${}_{o}F_{m}$. Arnott *et al.* have previously described the mathematical details [2,4,5] and the successful application of this approach to fibrous polyesters,[2] polypeptides,[3,4] and polynucleotides.[5] Since, in the model described here, there are only 4 residual degrees of freedom, we would expect accurate values of the conformation-angles (with perhaps estimated standard deviations of about 3°) to be obtained.

Refinement of the ring conformation-angles would require co-refinement with the ring bond-angles too. In this case it would be difficult to justify exclusion of the glycosidic bridge-angle. Whether or not meaningful refinement of the bond-angles is possible in a Θ minimization would have to be tested. If this is not possible, but further refinement of the molecular conformation through variation of the ring shape is deemed necessary, a ($\Psi = \Phi + \Omega + \Lambda$) minimization might be appropriate. In Ω would be included those variables, D_p, such as the bond-angles, that we would want elastically constrained to some standard value, ${}_{o}D_p$. Appropriate choice of values for the elastic constants, k_p, would ensure that all the variables included in Ω would not deviate more than, say, one estimated standard deviation from pre-determined standard values. However, for ι-carrageenan this would reduce the ratio (data)/(degrees of freedom) from $30/[11 - 6 - 1] = 30/4$ to $[(30 + 33)]/[(11 - 6 - 1) + (33 - 18)] = 63/19$ and whether Ψ minimization would lead to a demonstrably better model than Θ minimization with the smaller set of variables is a proposition that would have to be tested.

For chitin in particular, and for any other helical polysaccharide with only one sugar residue in the molecular asymmetric unit, there is an interesting strategic dilemma. The major structural parameters are the four in category (1) and the two (τ_2 and τ_3 in Figure 5a) in category (2), *i.e.* six in all. There are six equations connecting these parameters arising from the helix-making constraints so that there is an exact solution, or alternative sets of exact solutions, or possibly no solution in certain circumstances where the bond-length, bond-angle, and ring conformation-angle assignments would be imperfectly compatible with the helical dimensions. In this case we would want to augment our variable parameter list before a Θ or Ψ minimization and seek a least-squares minimization of $(\Omega + \Lambda)$ to parameter values near the desired ones. In the description in Figure 5b, 19 standard values of conformation-angles and bond-angles, 7 helix-making constraints and 7 ring-closing constraints and one constraint relating two τ values would mean that there would be $23 - 15 = 8$ degrees of freedom in such a model and 19 observational equations, making the least-squares approach appropriate. The weights, k_p, in Ω would be inversely proportional to the variance from the mean calculated in our survey so that the bond-angle observations would have weights [typically $1/(1{\cdot}5)^2$] an order of magnitude greater than conformational-angle weights [typically $1/(4{\cdot}5)^2$].

Once this initial model, compatible both with the desired stereochemistry and the helical constraints, had been prepared, the only molecular parameters that might obviously be varied in a Φ minimization would be the orientation of the acetamido-group and, possibly, the orientation of the hydroxymethyl group and for this the description in Figure 5c would be appropriate.

The next stage of refinement, a Ψ refinement (of bond-angles and ring conformation-angles) would add 8 degrees of freedom, but 15 terms in Ω, and might therefore be more appropriate than in the ι-carrageenan case. This would be the most elaborate model we might consider and is shown in Figure 5d.

CONCLUSION

This linked-atom, constrained-least-squares strategy has been devised in the context of determining polysaccharide conformations as far as possible from data supplied by the polymers themselves. These data are primarily from X-ray diffraction patterns that provide relatively precise information on chain periodicities and rather less precise information on the exact positions of atoms within the chain unit. This imprecision that results largely from the inadequate resolving power of the data, can, we think, be remedied to a large extent by synthesizing X-ray periodicites and X-ray intensities from the polymer, and appropriate, accurate stereochemical information from monomer derivatives. This accurate information will usually consist of the most probable values and variances of bond-lengths, bond-angles, and conformation-angles. The variance of a bond-length or a bond-angle will usually be small enough for the most probable values to be treated as constants. The variance of a conformation-angle will most often be large enough for it to be desirable to determine the polymer value independently of any presumed value. Sometimes a normally fixed value (*e.g.* of a bond-angle) may be needed to be stiffly elastic near some standard value and this too can be accommodated in our strategy.

This strategy has been rather intensively and successfully tested in studies on some other long-chain molecules. This has encouraged us to consider extending it to polysaccharides. There, the greater importance of ring structures make it more important that we consider limited variation of bond-angles in order to be able to vary ring conformations, whereas in most of the previous structures (with α-poly-L-alanine a noted exception [4]) bond-angles were usually kept fixed.

We are not the first to suggest linked-atom methods for chain molecules: Eyring [16] used an analogous approach in his dipole moment studies in 1932. Lagrange multipliers are a mathematical commonplace but were not

[16] H. Eyring, *Phys. Rev.*, 1932, **39**, 746.

commonly used in diffraction analyses until Arnott *et al.* made use of them to ensure chain continuity during refinement of helical molecules. Least-squares refinements for fitting *X*-ray data are of course ubiquitous in conventional single-crystal *X*-ray analyses of smaller molecules. (In these, however, atomic Cartesian coordinates are usually varied rather than the parameters suggested here.)

What is new is our argument that this combination of these three well-known procedures is both a rational and useful way in which to achieve accurate polysaccharide structures.

This research was sponsored by the U.S. National Science Foundation.

[1/1522 *Received, August 23rd,* 1971]