Crystal structure of cellulose polymorphs by potential energy calculations: 1. Most probable models for mercerized cellulose

A. J. Pertsin, O. K. Nugmanov, G. N. Marchenko and A. I. Kitaigorodsky
Institute of Organo-Element Compounds, USSR Academy of Sciences, Moscow 117334, Vavilov 28, USSR

An attempt is made to establish the crystal structure of mercerized cellulose through combined optimization of the crystallographic discrepancy factor and potential energy of the system. The potential energy is calculated using the semi-empirical atom–atom potential method. Several low-energy structures are found, all consistent with the available X-ray diffraction data. The most preferable model possesses a mobile hydrogen bonding network capable of facile rearrangement through thermal migration of protons.

INTRODUCTION

Despite a great number of crystallographic studies of cellulose, which have spanned a period of more than 50 years, the detailed crystal structure of this important polymer is still a subject of debate. This also applies to all known polymorphs of cellulose, including the most extensively studied polymorphs: cellulose I (native) and cellulose II (mercerized or regenerated). Inspection of the models suggested by various authors for the crystal structure of these polymorphs[1-10] shows lack of agreement even in such fundamental features as the hydrogen bonding network and the mutual polarity of the chains, in addition to fine details of the crystal structure.

The difficulties with cellulose are not surprising in view of a comparatively poor quality of the diffraction patterns from oriented cellulose samples. Typical X-ray diagrams of celluloses I and II contain only few tens reflections, which is clearly insufficient to refine all atomic parameters with standard crystallographic methods.

To increase the ratio of observations to refinable parameters, use is ordinarily made of various sterochemical and packing constraints, by keeping, in the course of refinement, the bond angles and lengths at the corresponding standard values and checking that the crystal model has no shortened non-bonded contacts[5,6]. However, even in this case the available diffraction data prove to be too few to make a unique choice between some competing crystal models. This is illustrated later by presenting several models for cellulose II, which are different in their structure but are nevertheless statistically indistinguishable.

Some supplementary information about the crystal structure of cellulose is afforded by spectroscopic measurements. Most valuable seem to be the measurements of the i.r. dichroism of the OH stretching vibrations[11], which yield the orientation of the hydrogen bonds with respect to the chain axis. Here again, however, a variety of distinct crystal models may in principle be constructed, all consistent with the experimental evidence.

The ambiguity involved in interpretation of the diffraction data leads one to resort to non-crystallographic criteria for assessing the reasonableness of a crystal model. The most obvious criterion is the potential energy of the crystal, which should be lowest at the actual crystal configuration. To be certain, such a complicated system as cellulose can only be treated with very simple potential models. An example is the potential model used by Sarko and coworkers[1,2,9], which represents a sum of penalty functions of the form:

$$f(x_i) = W_i(x_i - x_i^0)^2 \qquad (1)$$

In this equation x_i is a conformational parameter of the macromolecule or the separation distance between two non-bonded atoms, x_i^0 is the corresponding standard or equilibrium value, w_i is a weighting factor.

It is clear that this potential model is too oversimplified to be capable of giving quantitative estimates for the actual potential energy. A much more realistic model may be constructed on the basis of the atom–atom potential method[12] which has shown itself as a reasonably reliable tool for predicting the structure and properties of organic crystals.

In this paper the atom-atom potential method is used to find the most probable models for mercerized cellulose. The other cellulose polymorphs will be considered in subsequent papers of this series.

METHOD OF CALCULATION

The geometrical model adopted for cellulose II represents an array of deformable cellulose chains packed together in accordance with the observed crystal symmetry. On generating the chains the conformational parameters of the glucose rings are fixed at their 'average' values, as reported by Arnott and Scott[13]. All hydrogen atoms attached to the rings are generated at the expected positions, using standard bond angles and lengths. Standard values are also used for the bond angles and lengths
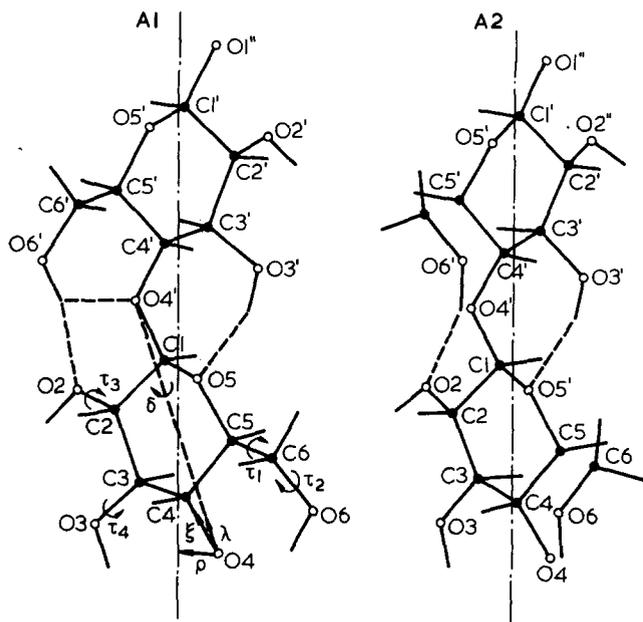
Figure 1 Conformers A1 and A2 of cellulose macromolecule

of the side groups. The only parameters allowed to vary in the monomer residues are the torsion angles τ_i (see *Figure 1*) describing rotations of the hydroxymethyl ($i = 1$) and three hydroxyl ($i = 2, 3, 4$) groups. The particular bond sequencies used in the definitions of τ_i are: $\tau_1 = C4-C5-C6-O6$, $\tau_2 = C5-C6-O6-H$, $\tau_3 = C1-C2-O2-H$, $\tau_4 = C2-C3-O3-H$. Each angle is zero when the respective bond sequence, A-B-C-D, is *cis*. Counter-clockwise rotation of the C-D bond when looking down the B-C bond represents positive rotation. Note that the conventional gg, gt, and tg positions used in describing the orientation of the CH_2OH group[1-11] correspond to $\tau_1 = -60°, 180°$ and $60°$, respectively.

The monomer residues are linked into the chain with the variable virtual bond method[14,15], assuming the chain symmetry to be 2_1. This requires two further parameters to be introduced to specify the chain conformation. These are chosen to be the projection h of the virtual bond O4–O4' on the chain axis (half the chain period) and an angle, δ, describing rotation of the monomer residue about the virtual bond. The angle δ is defined by three vectors, $\bar{\rho}, \bar{\xi}$ and $\bar{\lambda}$, emanating from the chain origin, O4 (see *Figure 1*). The radius vector $\bar{\rho}$ is perpendicular to the chain axis, $\bar{\xi}$ is along the O4–C4 bond, and $\bar{\lambda}$ along the virtual bond. The angle δ is introduced as the dihedral angle between the plane defined by $\bar{\lambda}$ and $\bar{\xi}$, and the plane defined by $\bar{\lambda}$ and $\bar{\rho}$; δ is zero when $\bar{\rho}, \bar{\lambda}$ and $\bar{\xi}$ are coplanar and δ is positive when $\bar{\rho}, \bar{\lambda}$ and $\bar{\xi}$ form a right-handed set.

It is noteworthy that with the variable virtual bond method the conformational parameters at the junction of two successive residues are not independent variables of the model, but are implicit functions of h, δ and the residue conformation. In the subsequent discussion one such parameter is referred to as — the glycosidic bond angle, β.

The chains generated with the variable virtual bond method are positioned in the unit cell according to the P2$_1$ two-chain model commonly adopted for cellulose II[5,6,9,16,17]. One chain is positioned at the origin of the unit cell, and the other at the centre of the *ab* plane, with the chain axes parallel to *c*. (The unit cell dimensions are fixed at the values reported by Kolpak *et al.*[6] for mercerized cellulose). At this step, five further parameters,

hereafter designated as ϕ^1, ϕ^2, p^1, p^2, and s, are introduced to specify the orientation, direction and relative shift of the chains. The angles ϕ^1 and ϕ^2 describe rotations of the origin and centre chains about their axes and are defined as the angles formed by the vectors $-\bar{p}$ (see *Figure 1*) and the unit cell axis $a^* = a \cos(\gamma - \pi/2)$. The chain position with O4 at (O, $-y$, O) for the origin chain and (1/2, 1/2 − y, z) for the centre chain corresponds to $\psi1 = \psi2 = -\tau_i/2$. Rotation in the direction from *a* to *b* represents positive rotation.

The chain direction is defined as positive ($p = 1$) when $z_{C1} > z_{C4}$ and negative ($p = -1$) otherwise. The glycosidic oxygen O4 of the origin chain is always kept at $z = O$, while that of the centre chain is at $z = s$.

Altogether, this model of cellulose II involves 15 variable parameters: $\tau_i^k, \delta^k, \phi^k, p^k$, and s, with $i = 1,...,4$ and k referring either to the origin or centre chain. (The parameter k needed to generate the chains is fixed at $c/2$). There is also one 'non-geometrical' parameter – the average isotropic temperature factor B appearing in the structure amplitudes (see equation (6)).

The applicability of a trial set of model parameters is assessed by computing the objective function:

$$\phi = U + WR'' \tag{2}$$

where U is the potential energy of the system, R'' the crystallographic discrepancy factor, and W a weighting factor.

The potential energy, U, includes both the intramolecular (intrachain) and intermolecular (interchain) contributions:

$$U = U_{intra} + U_{inter} \tag{3}$$

The intermolecular energy is computed with the atom–atom potential method as a sum of non-bonded interactions between the atoms belonging to different chains. The intramolecular energy is evaluated as:

$$U_{intra} = U_{mon} + U_{junct} \tag{4}$$

In this equation U_{mon} is the conformational energy of the monomer residues, which includes the torsional and non-bonded contributions from the dihedral angles and atom–atom contacts which are affected by variation of the monomer parameters τ_i^k; U_{junct} involves bond angle bending contributions from the glycosidic bond angles, torsional contributions from the dihedral angles at the junctions, and also all non-bonded atom–atom interactions between two successive residues along each of the two crystallographically distinct chains.

The potential functions used in computing the energy of the system are presented in *Table 1*. Note that different non-bonded potentials for the same atom–atom species are used depending on what interaction, intramolecular or intermolecular is involved. This is more appropriate than using the same potentials for both the intra- and intermolecular interactions because the non-bonded potentials fitted to intramolecular properties (as are the conformational non-bonded potentials[18] used here) are strongly correlated with the other constituents of the conformational energy model, and are generally unsuitable for dealing with the intermolecular energy. Similarly, non-bonded potentials derived from purely 'intermolecular' properties (as are the potentials from ref. 19, 20 derived from crystal structure and heat-of-sublimation data) are generally inappropriate in treating the confor-

Table 1 Potential functions and parameter sets used in energy calculations. (The units correspond to energies in kcal mol^{-1} angles in degrees and distances in nm)

Potential function						Reference
Glycosidic bond angle bending potential	$= 0.01\ (\beta - 90°)^2$					18
Torsional potential	$= \dfrac{U_0}{2}(1 + \cos 3\tau)$, $U_0 = 0.332$ for C–C $U_0 = 0.336$ for C–O					18
Hydrogen bond potential	$= 4.0\ [e^{-6\ (r_H \ldots O\ -1.8)} - 2e^{-3\ (r_H \ldots O^{-1.8})}]$					18
Non-bonded atom– atom potential	$= -Ar^{-6} + Be^{-Cr}$					
	Intramolecular interactions	$A/10^2$	$B/10^4$	C		
	H...H	0.40	2.86	5.200		
	H...C	1.21	3.28	4.130		
	H...O	1.22	5.75	4.727		18
	C...C	4.76	3.77	3.513		
	C...O	4.41	6.37	3.881		
	O...O	3.46	9.65	4.333		
	Intermolecular interactions					
	H...H	0.29	0.49	4.29		
	H...C	1.18	1.86	3.94		
	H...O	0.87	1.95	4.24		19, 20
	C...C	4.21	7.16	3.68		
	C...O	3.31	7.46	3.47		
	O...O	2.59	7.77	4.18		

mational energy. The use of different potential functions for the conformational and intermolecular energies gives, therefore, a more reliable model for the total potential energy.

The R''-factor in equation (2) is defined by[21]:

$$R'' = \left(\sum_{m=1}^{M} w_m |F_m^{\text{obs}} - F_m^{\text{calc}}|^2 \Big/ \sum_{m=1}^{M} w_m |F_m^{\text{obs}}|^2 \right)^{1/2} \quad (5)$$

where F_m^{obs} and F_m^{calc} are the observed and calculated structure factor amplitudes, w_m is the weight applied to the m-th reflection, M is the number of observed reflections. Each F_m^{calc} is a function of the 16 parameters of the model and is computed from[5]:

$$F_m^{\text{calc}} = K \left\{ \sum_{hkl} [F_{hkl}^{\text{calc}} \exp(-B\rho_m^2/4)]^2 \right\}^{1/2} \quad (6)$$

The summation in this equation is over all planes hkl contributing to the m-th reflection; ρ_m is the reciprocal d spacing; K is the scale factor.

The numerical values for F_m^{obs} and w_m for 29 observed and 11 non-observed reflections were taken from the X-ray diffraction study by Kolpak et al.[6] on mercerized cellulose. Unlike the cited work, however, the scale factor K was not treated as a variable parameter of the model, but was determined for each running set of F_m^{calc} so as to minimize R''.

As seen from equation (2), the calculated R''-factor is given a weight W. The numerical value for W was chosen so as to make small changes in R'' and U equally 'significant' for the objective function. The considerations used in selecting W are as follows. Consider two alternative models, A and B, with the R''-factors R_A'' and R_B'' ($R_B'' > R_A''$). Evaluate what difference in R'' is statistically significant. Each model is described by 16 parameters, six of which, i.e. τ_2^k, τ_3^k and τ_4^k ($k = 1, 2$), have a negligible effect

on the R''-factor. Also, as shown later, two more parameters, p^1 and p^2, may be formally reduced to a single parameter, p, assuming three distinct values for one antiparallel ($p = 0$) and two parallel ($p = \pm 1$) variants of the chain packing. Thus, the number of parameters 'important' for the R''-factor is $16 - 7 = 9$. The number of observations (here, reflections) is 29. The hypothesis to be tested is: the 9 parameters corresponding to the poorer model B correctly describe the crystal structure. The dimension of the hypothesis is 9. The number of degrees of freedom is $29 - 9 = 20$. For the most usual significance level of 1 per cent it is apparent, from Hamilton's tables[21], that the hypothesis cannot be rejected if $R_B''/R_A'' < 1.6$. Assuming the better model A to have an R''-factor of ≈ 0.12 (a value typical of the best models of cellulose II), it is evident that a difference in R'' less than 0.07 is insignificant.

However, the accuracy offered by the atom–atom potential method in potential energy calculations is typically of the order of 1 kcal mol^{-1}. In comparisons of different modifications of the same crystal the actual accuracy may be better, but for the present it is assumed that any two crystal models are energetically equivalent if their energies differ by less than 1 kcal mol^{-1}. Thus, to make the objective function equally sensitive to small changes in U and R'', the R''-factor in equation (2) should be given a weight of $1/0.07 \approx 15$ (kcal mol^{-1}).

Formally, the search for the optimum crystal model represents a search of the global minimum of the objective function ϕ in 16-dimensional parameter space. In a general case this is a significant problem. Fortunately, in this particular case the problem may be substantially simplified by using simple packing and symmetry considerations. Thus, based on the chain shape in the projection down the c-axis it is easy to see that a reasonable (non-overlapping) packing of the chains may
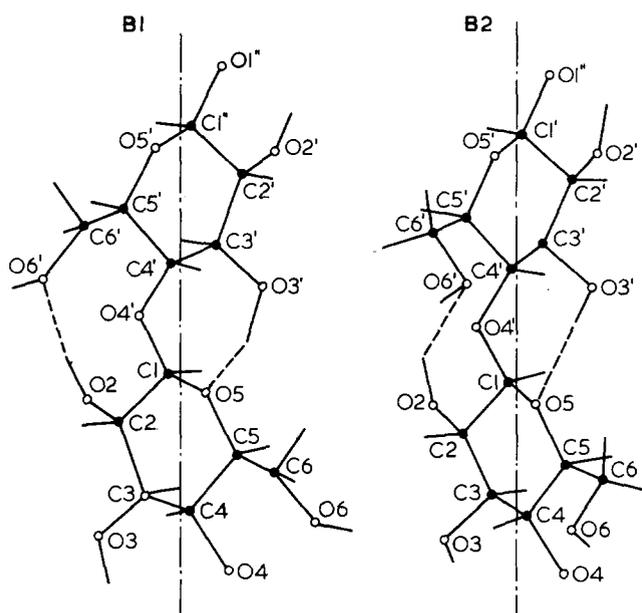
Figure 2 Conformers B1 and B2 of cellulose macromolecule

only result when the glucose rings are disposed nearly parallel to the *ac* plane. This corresponds to variation of within relatively narrow intervals around $\psi^k \approx \pm\tau_i/2$ for $p^k = 1$ or $\psi^k \approx 0$ and τ_i for $p^k = -1$.

Further, the parameter space to be scanned in the search for the global minimum may be reduced by resorting to the obvious symmetry relations:

$$\phi(\eta^1, \phi^1, p^1; \eta^2, \phi^2, p^2, s)$$
$$= \phi(\eta^2, \phi^2, p^2; \eta^2, \phi^2, p^2; \eta^1, \phi^1, p^1, -s)$$
$$= \phi(\eta^1, \phi^1 + \tau_i, p^1; \eta^2, \phi^2 + \tau_i, p^2, s)$$
$$= \phi(\eta^1, \phi^1, p^1; \eta^2, \phi^2 + \tau_i, p^2, s \pm 1/2)$$
$$= \phi(\eta^1, \phi^1 + \tau_i, p^1; \eta^2, \phi^2, p^2, s \pm 1/2) \qquad (7)$$

where η^k denotes all conformational parameters of the k-th chain, $\eta^k = \{\tau_i^k, \delta^k\}$. Based on these symmetry relations, the subspace of parameters ϕ^k, p^k and s may be subdivided into three crystallographically distinct regions chosen as follows:

(a) $p^1 = 1$, $p^2 = -1$, $-1/2 < s \leqslant 1/2$, $\phi^1 \approx -\tau_i/2$, $\phi^2 \approx 0$;

(b) $p^1 = p^2 = 1$, $-1/2 < s \leqslant 1/2$, $\phi^1 \approx \phi^2 \approx -\tau_i/2$;

(c) $p^1 = p^2 = -1$, $-1/2 < s \leqslant 1/2$, $\phi^1 \approx \phi^2 \approx 0$.

Instead of the direction parameters p^1 and p^2 a polarity parameter, $p = (p^1 + p^2)/2$, may now be introduced which assumes the values 0 $(p^1 = 1, p^2 = -1)$, $1(p^1 = p^2 = 1)$, and $-1(p^1 = p^2 = -1)$. The value $p = 0$ corresponds to the antiparallel arrangement of the chains, while $p = \pm 1$ to the parallel variants.

Further reduction of the computational work associated with the global search is possible due to the fact that reasonable values for δ^k prove to be within a rather narrow range, $\delta^k \approx 45 \pm 10°$. Variation of δ^k out this interval leads either to shortened O2...C6' contacts or to unrealistic glycosidic bond angles[14].

Finally, the optimum value for the average isotropic temperature factor is approximately $32A^2$ and dependent only slightly on the other parameters of the model.

Thus, the model chosen for calculation involved only nine 'global' parameters $(\tau_i^k$ and $s)$, i.e. the parameters which had to be scanned throughout the entire range of their variation. These nine parameters were sampled from the corresponding parameter subspace using uniform random grids.

Each trial set of model parameters was refined by local minimization of the objective function using Powell's quasi-Newton algorithm[22]. The most favourable regions of parameter space were also explored by varying, in turn, one of the nine 'global' parameters, while allowing the other 15 parameters of the model to adjust to minimum ϕ.

RESULTS AND DISCUSSION

Before proceeding with the calculations of the optimum crystal structure, it appeared reasonable to see what conformations are most favourable for an isolated cellulose chain of 2_1 symmetry. In seeking the low-energy conformations all intrachain conformational parameters, τ_i and δ, were allowed to vary, as well as the chain half-period h. There were six most favourable conformers found in the isolated chain calculations. These are depicted in *Figures 1-3*, and their conformational and energetic parameters are listed in *Table 2*. In all the conformers there is an O5... HO3' hydrogen bond, with an energy of ≈ 3.9 kcal mol^{-1}. In the lowest energy conformer, labelled A1, there are also three-centre (bifurcated) hydrogen bonds, O6'H $\because$ $\begin{smallmatrix}O4\\O2\end{smallmatrix}$, which involve one donor and two acceptor oxygens. The bonds of this kind will hereafter be referred to as two-acceptor bonds, to distinguish them from two-donor bonds of the type O $\because$ $\begin{smallmatrix}HO\\HO\end{smallmatrix}$.

The energetic preferrence of two-acceptor OH $\because$ $\begin{smallmatrix}O\\O\end{smallmatrix}$ bonds over linear OH...O ones has been demonstrated recently by Newton *et al.*[23] through *ab initio* 4-31G calculations on the water trimer. It was found that the appearance of the second acceptor of proton made the system 1-2 kcal mol^{-1} more stable as compared to the equilibrium linear case. For the A1 conformer, the energy difference between the O6'H $\because$ $\begin{smallmatrix}O4'\\O2\end{smallmatrix}$ bond and the equilibrium linear bond was 2.6 kcal mol^{-1}, which is similar to the *ab initio* result.
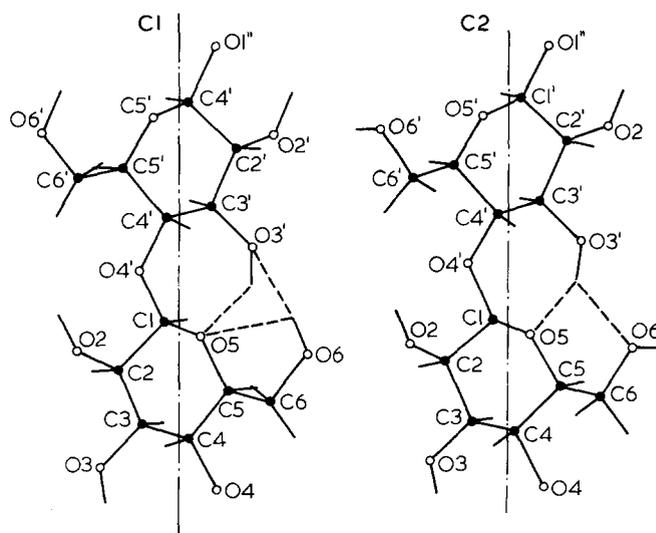
Figure 3 Conformers C1 and C2 of cellulose macromolecule

Table 2 Conformational and energetic parameters of six stable conformers of the isolated cellulose macromolecule (the units are degrees, nm and kcal mol^{-1})

Conformer	τ_1	τ_2	τ_3	τ_4	δ	h	β	v
A1	64	-52	174	170	44	5.16	117	-4.8
A2	-20	79	179	180	36	5.17	116	3.1
B1	68	167	-53	171	44	5.16	118	0.3
B2	-17	184	38	180	36	5.17	116	-0.5
C1	-174	47	-56	-122	53	5.11	117	-0.1
C2	-168	179	-56	156	51	5.13	118	2.6

Table 3 The structural and energetic parameters for different models of mercerized cellulose (The units are degrees and kilocalories per mole)

Model	τ_1^1	τ_2^1	τ_3^1	τ_4^1	δ^1	φ^1	τ_1^2	τ_2^2	τ_3^2	τ_4^2	δ^2	φ^2	s	R''	U	ϕ
a_1	66	-51	-151	174	43	-82	69	162	-53	163	46	-3	0.106	0.179	-21.4	-18.7
a_2	68	-52	-152	176	42	-83	71	103	-57	163	45	-2	0.102	0.183	-21.2	-18.5
a_3	65	57	-79	177	42	-83	71	166	-52	175	46	-2	0.122	0.178	-21.2	-18.5
a_4	-63	173	-161	182	37	-90	69	163	-52	174	45	-2	0.125	0.164	-20.1	-17.6
a_5	67	-50	-152	-55	42	-83	68	102	-56	161	47	-6	0.100	0.186	-20.4	-17.6
a_6	178	164	-97	174	44	-81	72	161	-56	167	42	1	0.226	0.124	-19.4	-17.5
a_7	67	-52	-153	176	42	-86	70	34	-54	72	45	1	0.116	0.208	-20.5	-17.4
p_1	64	64	-70	172	45	-74	170	175	-77	169	45	-79	0.136	0.168	-19.4	-16.9
a_8	174	166	-123	177	42	-82	64	-65	177	166	43	-1	0.217	0.126	-18.3	-16.4
a_9	177	162	-109	176	43	-82	76	172	183	167	43	0	0.226	0.121	-17.6	-15.8
a_{10}	-178	168	88	174	41	-83	-164	-62	-35	167	44	0	0.245	0.133	-17.1	-15.1
a_{11}	77	-48	-156	173	57	-61	102	162	-52	160	44	-7	0.107	0.111	-9.7	-8.0
a_{12}	154	164	-81	165	43	-80	157	80	-49	159	61	-18	0.234	0.108	9.4	-7.8

One further argument to support the possibility of formation of two-acceptor hydrogen bonds in cellulose is the wide occurrence of such bonds in carbohydrate crystals. An analysis of the crystal structure of the carbohydrates for which accurate neutron diffraction data are available shows[24] that of 100 hydrogen bonds observed in the crystals, 25 are two-acceptor ones.

Table 2 shows that the CH_2OH group of the A1 conformer is close to the *tg* position. Rotation through $\approx 80°$ makes the group closer to the *gg* position and yields a new stable conformer, A2. In this conformer the O6'H...O4' bond breaks down, so that the two-acceptor bond transforms to a usual linear O6'H...O2 bond with an energy of 3.8 kcal mol^{-1}.

The O6'...O2 hydrogen bond can also be formed through the hydrogen at O2, with O2 serving as proton donor and O6' as proton acceptor. This is characteristic of the conformers of type B (see *Figure 2*). Here again, two distinct orientations of the CH_2OH group are possible. In the B1 conformer the CH_2OH group is close to the *tg* position, while in B2 it becomes closer to *gg*. The energy of the O2H...O6' hydrogen bond is found to be 3.9 and 3.6 kcal mol^{-1} for B1 and B2, respectively.

As seen from *Table 2* and *Figure 3*, there are also two stable conformers, labelled C1 and C2, the CH_2OH groups of which are close to the *gt* position. In the C1

conformer there is a 'cyclic' $\overset{O3'\cdots}{\underset{O5\cdots}{H\cdots HO6}}$ bond with a total

energy of 8.5 kcal mol^{-1}. The C2 conformer possesses a O3'H $\overset{O6}{\underset{O5}{\cdots}}$ two-acceptor bond with an energy of ≈ 6 kcal mol^{-1}.

It is essential that all the six stable conformers of the isolated chain have an equilibrium period close to that observed experimentally for cellulose (≈ 10.3 Å). Also, in all the conformers there are free sites capable of participat-

ing in intermolecular hydrogen bonds. For this reason, each of the conformers may, in principle, exist in the crystal state.

In the search for the optimum crystal structure of cellulose II a great number of local minima of the objective function were found. There were three deepest minima, labelled a_1, a_2 and a_3, with an objective function of approximately -18.6 kcal mol^{-1}. Next were four minima, a_4-a_7, ≈ 1 kcal mol^{-1} less deep. All these minima corresponded to the antiparallel packing of the chains. The deepest minimum found for the parallel packing, p_1, had an objective function of -16.9 kcal mol^{-1}. The parameters of these minima are listed in *Table 3*. Also included in this Table are five more minima which were not so satisfactory by the objective function, but had a low R''-factor (minima a_8-a_{12}).

Table 3 shows that the best parallel model is markedly inferior to the best antiparallel ones. The difference in the objective function between the best parallel and antiparallel variants of the chain packing appears quite convincing to reject the hypothesis that cellulose II may have parallel chains[10].

Another conclusion which can be made from *Table 3* is that it is indeed hardly possible to determine the structure of mercerized cellulose on the basis of X-ray diffraction data alone. For all the models presented in *Table 3*, excepting model a_7, the ratio R_i''/R_j'' ($R_i''>R_j''$) does not exceed 1.6. This means that only the a_7 model can be rejected at a significance level of greater than 1 per cent, while the other models are all statistically indistinguishable.

A good illustration for the fact that low R''-factor values cannot be given much weight is provided by the a_{11} and a_{12} models: these models possess very low R''-factors but are quite unsatisfactory from the energetic viewpoint.

Of the best antiparallel models, model a_6 is noteworthy. By its parameters, this model is close to the model of
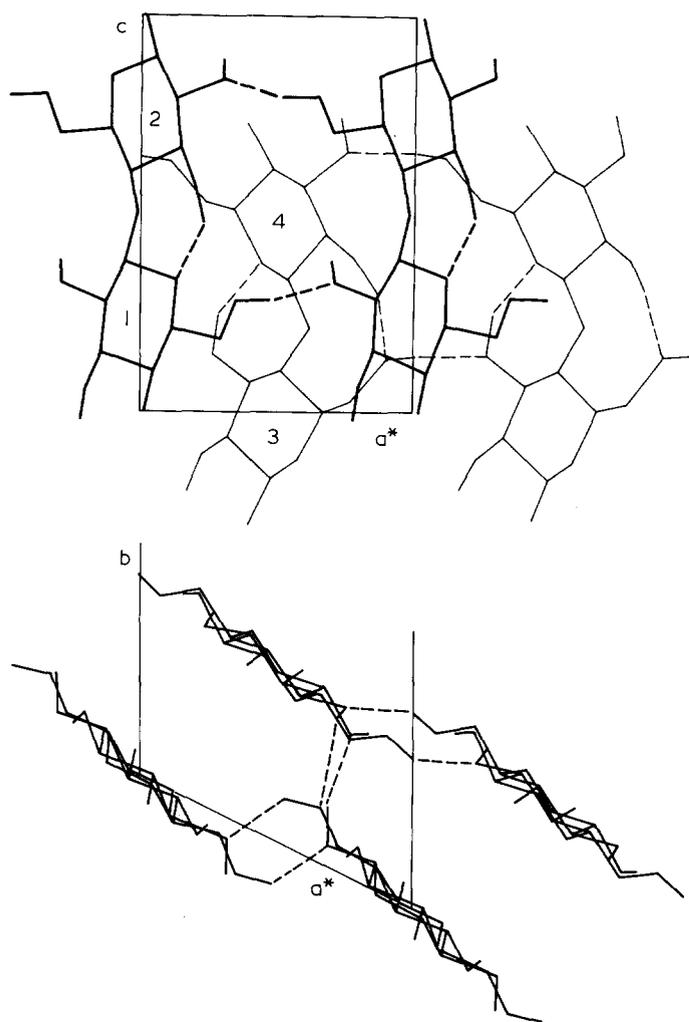
Figure 4 Projections of the cellulose chains down the *b*- and *c*-axes for model a_6 (only carbon, oxygen and hydroxyl hydrogen atoms are drawn in, for clarity; hydrogen bonds are indicated by dashed lines)

cellulose II derived by Kolpak and Blackwell[5] through rigid-body refinement of X-ray diffraction data. A similar model was also obtained by Stipanovic and Sarko[9] by a combined X-ray diffraction analysis and stereochemical packing refinement based on penalty functions of the form of equation (1). Transformation of the model parameters reported in refs. 5 and 9 to the conventions used here yields $\tau_1^1 = 176°$, $\tau_1^2 = 80°$, $\phi^1 = -91°$, $\phi^2 = 2°$, $s = 0.211$ for the model in ref. 9, and $\tau_1^1 = 186°$, $\tau_1^2 = 70°$, $\phi^1 = -90°$, $\phi^2 = 6°$, $s = 0.216$ for the model in ref. 5, which is similar to the a_6 model parameters.

Using the notation in ref. 9, the a_6 model may be clasified into $gt + tg$ models, with the symbols gt and tg indicating the position of the CH_2OH group in the origin and centre chain, respectively. The projections of the a_6 model structure down the c and b axes are shown in *Figure 4*. The structure represents and array of alternating sheets parallel to the ac plane, each formed by translationally equivalent chains. The centre chains, which form the (020) sheet, are in a conformation similar to conformation B1 of the isolated chain (see *Table 2* and *Figure 2*). The sheet possesses two intramolecular hydrogen bonds, $O5_4...HO3_3$ (3.66) and $O2_4H...O6_3$ (3.98), and an intermolecular $O6_3H...O3_3^a$ (3.98) bond*.

The chain conformation in the (010) sheet is not typical of the isolated chain: here there is only one intramolecular

hydrogen bond, $O5_1...HO3_2$ (3.96), while the CH_2OH group is involved in a strong intermolecular intrasheet bond, $O6H...O2_1^a$ (3.81).

The neighbouring sheets of the a_6 model are linked together through a strong $O2_4...HO2_1^a$ (3.26) bond and weak $O3_3H...O6_1^b$ (1.65) bond. Thus the proton at $O3_3$ proves to be involved in a two-acceptor $O3_3H$ $\begin{smallmatrix}\cdots O6_1^b\\ \cdots O5_4\end{smallmatrix}$ bond.

As compared to the model in ref. 5, model a_6 possesses an additional intersheet hydrogen bond, $O3_3H...O6_1^b$. This bond occurs also in the model in ref. 9. The latter, however, has one more intersheet bond, $O6_3H...O3_1^a$, which enters a two-acceptor $O6_3H$ $\begin{smallmatrix}\cdots O3_1^a\\ \cdots O3_3^a\end{smallmatrix}$ bond. An attempt to find a structure where such a bond would occur along with the other hydrogen bonds was made but was unsuccessful; all trial structures with the hydrogen bonding network of the model in ref. 9 were unstable and transformed, after refinement, to the a_6 model.

Contrary to the finding of Stipanovic and Sarco[9], who reported that the $gt + tg$ models were significantly lower in packing energy over any possible O6 rotationally 'mixed' models (i.e., the models whose origin and centre chains had their CH_2OH groups at different positions), here, a model of the type $gg + tg$ has been found, possessing a low potential energy and an acceptable R''-factor. This is the model designated in *Table 3* as model a_4. The structure of a_4 is characterized by a total of three intramolecular hydrogen bonds, $O5_1...HO3_2$ (3.87), $O5_4...HO3_3$ (3.98) and $O2_4H...O6_3$ (3.99), four intermolecular intrasheet bonds, $O6_1H...O2_1^a$ (1.15), $O6_1H...O3_1^a$ (2.17), $O6_1...HO2_1^a$ (1.88), $O6_3H...O3_3^a$ (3.81), and one intersheet bond, $O2_1^aH...O2_4$ (3.87), with the protons at $O6_1$ and $O2_1^a$ involved in two-acceptor $O6_1H$ $\begin{smallmatrix}\cdots O2_1^a\\ \cdots O3_1^a\end{smallmatrix}$ and $O2_1^aH$ $\begin{smallmatrix}\cdots O2_4\\ \cdots O6_1\end{smallmatrix}$ bonds.

In agreement with the results of Stipanovic and Sarco[9] no model of the type $gt + gt$, which would have an acceptable packing energy, was found. The best model of this type is a_{10}. It is noteworthy, however, that unlike $gt + gt$ models examined by Stipanovic and Sarco[9], the a_{10} model possesses a low R''-factor, similar to that of the a_6 model.

An unexpected result was obtained with models a_5 and a_7 (both being of the type $tg + tg$). A characteristic feature of these models is the absence of the intramolecular O5...HO3' bond in one of the sheets. Usually, such structures are *a priori* excluded from the possible structure of cellulose, arguing that the breakdown of the O5...HO3' bond leads necessarily to a too high energy loss. The calculation shows, however, that the loss of the O5...HO3' bond may be compensated for by formation of new intermolecular bonds. (The net result of such a compensation is shown in *Table 3*: the a_5 and a_7 models are ≈ 1 kcal mol^{-1} more stable than the a_6 model where both the O5...HO3' bonds occur). In the a_5 model the liberated proton at O3' is involved in an intermolecular intersheet

---
* The symbols used in labelling hydrogen bonds are the following: atom subscript refers to one of the four basis residues numbered in *Figure 4*, atom superscript indicates translation applied to the basis residue; the figure in parentheses is the hydrogen bond energy in kcal mol^{-1}. For each model only crystallographically distinct hydrogen bonds are presented. In selecting such bonds use was made of the obvious symmetry relations of the type: $O_1H...O_1^{-a} = O_1^aH...O_1$, $O_3...HO_2^{-c} = O_4^{-a}...HO_1$, etc.
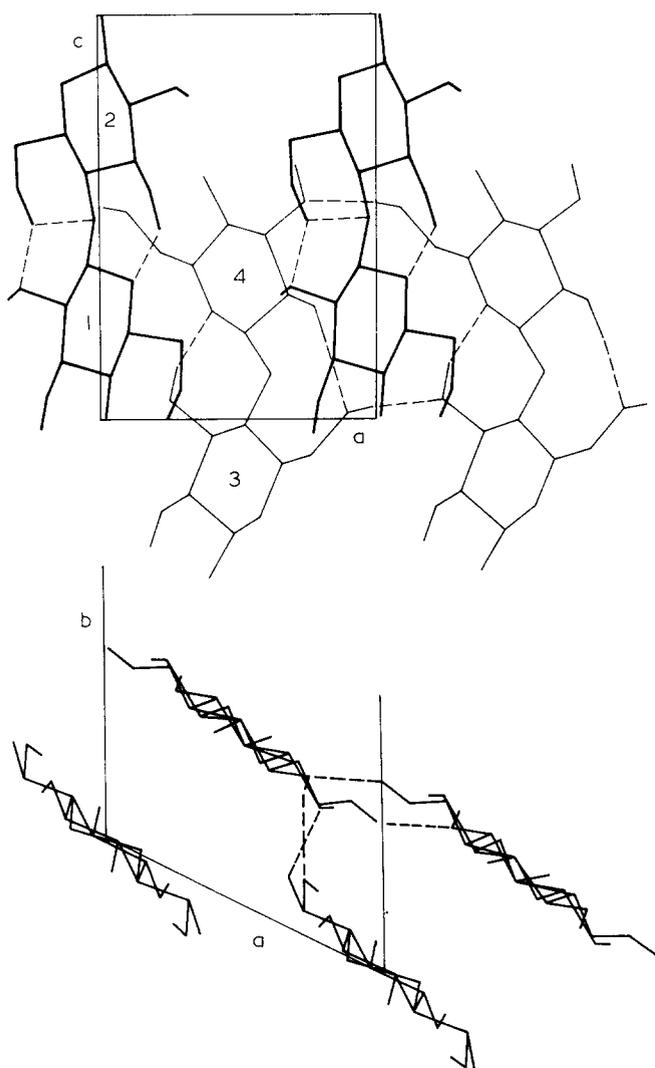
Figure 5 Projections of the cellulose chains down the *b*- and *c*-axes for model a_1

$O3_1^a H...O6_3$ (2.39) bond. Aside from this bond, the a_5 model possesses also the following bonds:

$$O6_2H \overset{\cdot\cdot\; O4_2 \; (2.63)}{\underset{\cdot\cdot\; O2_1 \; (3.84)}{\cdot}}, \; O2_4H...O6_3 \; (3.97),$$

$$O3_3H \overset{\cdot\cdot\; O5_4 \; (3.67)}{\underset{\cdot\cdot\; O6_1^b \; (1.80)}{\cdot\cdot}}, \; O6_3H...O3_3^a \; (3.57),$$
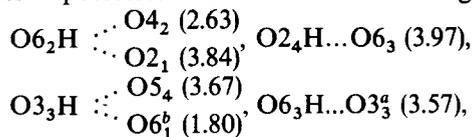
and $O2_1^a H...O2_4$ (3.49).

In the a_7 model the loss of the $O5...HO3'$ bond is compensated for by formation of an intermolecular bond which serves as a binding unit of the following chain:

$$O6_1^{a+b} \overset{\cdot\cdot\; HO6_3 \; (1.27)}{\underset{\cdot\cdot\; HO2_4 \; (3.96)}{\underset{O6_3 \; \overset{\cdot\cdot\; HO3_3^a \; (3.55+3.44)}{\cdot}}{\cdot}}}$$
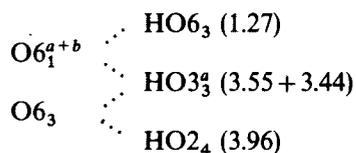
Also present in the a_7 model are $O5_1...HO3_2$ (3.44), $O6_2H \overset{\cdot\cdot\; O4_2 \; (2.58)}{\underset{\cdot\cdot\; O2_1 \; (3.95)}{\cdot}}$, and $O2_1^a H...O2_4$ (3.39) bonds.

Model a_1 corresponding to the global minimum of the objective function may be classified with $tg + tg$ models by the positions of the CH_2OH groups. The projections of the a_1 model structure down the *b* and *c* axes are shown in *Figure 5*. As seen from the drawing, the (020) sheet of the

a_1 model is constructed similarly to the (020) sheet of model a_6. The chain conformation in the sheet is close to conformation B1 of the isolated chain. The sheet possesses two intramolecular hydrogen bonds, $O3_3H...O5_4$ (3.65) and $O2_4H...O6_3$ (3.96) and one intermolecular bond, $O6_3H...O3_3^a$ (3.81).

The chains in the (010) sheet are in conformation close to A1 with its characteristic hydrogen bonds: $O5_1...HO3_2$ (3.94) and $O6_2H \overset{\cdot\cdot\; O4_2 \; (2.65)}{\underset{\cdot\cdot\; O2_1 \; (3.89)}{\cdot}}$. There are no intermolecular hydrogen bonds in this sheet.

The intersheet variety of hydrogen bonds in model a_1 is represented by $O2_1^a H...O2_4$ (3.29) and $O3_3H...O6_1^b$ (1.57) bonds, the proton at $O3_3$ being involved in a two-acceptor $O3_3H \overset{\cdot\cdot\; O5_4}{\underset{\cdot\cdot\; O6_1^b}{\cdot}}$ bond.

Note that the only difference in the hydrogen bonding network topography between models a_1 and a_6 is the occurrence of the intramolecular $O6_2H \overset{\cdot\cdot\; O4_2}{\underset{\cdot\cdot\; O2_1}{\cdot}}$ bond in the former, instead of the intermolecular $O6_1H...O2_1^a$ bond in the latter.

Table 3 shows that model a_1 differs from the next two models, a_2 and a_3, mainly in orientation of the HO6 and HO2 hydroxyls. Thus, a_2 may be derived from a_1 by rotation of $HO6_2$ through $\approx 60°$ (see parameter τ_2^2 in *Table 3*), and a_3 by rotation of $HO6_1$ and $HO2_1$ through 108 and 70°, respectively (see parameters τ_2^1 and τ_3^1). In model a_2 the topography of the hydrogen bonding network remains unaltered: the only change is the displacement of the proton of the $O6_2H...O3_2^a$ bond to a new position relative to the line joining the acceptor and donor oxygens. In model a_3 there are marked changes in the hydrogen bonding network. The resulting structure possesses the following hydrogen bonds:

$$O5_1...HO3_2 \; (3.94), \; O2_1H \overset{\cdot\cdot\; O6_2 \; (3.95)}{\underset{\cdot\cdot\; O2_4^{-a} \; (1.30)}{\cdot}},$$

$$O5_4...HO3_3 \; (3.99),$$

$$O3_3^a \overset{\cdot\cdot\; HO6_3 \; (3.90)}{\underset{\cdot\cdot\; HO6_1^{a+b} \; (3.37)}{\cdot}}, \; \text{and} \; O2_4H...O6_3 \; (3.91).$$
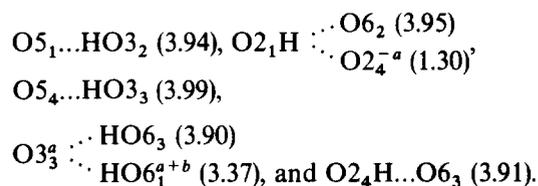
The rearrangement of the hydrogen bonding network, occurring on the $a_1 \rightarrow a_3$ transition may be depicted as:
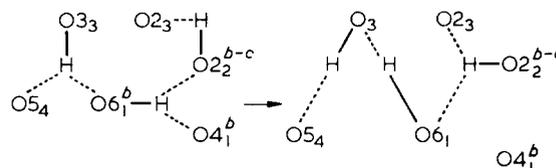
It is essential that all the three models, a_1, a_2 and a_3, are almost equivalent by energy (see *Table 3*). A calculation of the barriers to the $a_1 \rightarrow a_2$ and $a_1 \rightarrow a_3$ transitions, carried out by varying, in turn, the transition parameters τ_2^2, τ_2^1 and τ_3^1, while allowing the other 15 parameters of the model to adjust to minimum energy, yields energy barriers as small as 0.2 and 1.0 kcal mol^{-1}, respectively (see *Figure 6*). It appears, therefore, that all the three structures may occur in the crystal at one time and transform to one another through thermal migration of the protons; i.e. models a_1, a_2 and a_3 may be regarded as a single model (hereafter, model a_0) with a mobile hydrogen bonding network.
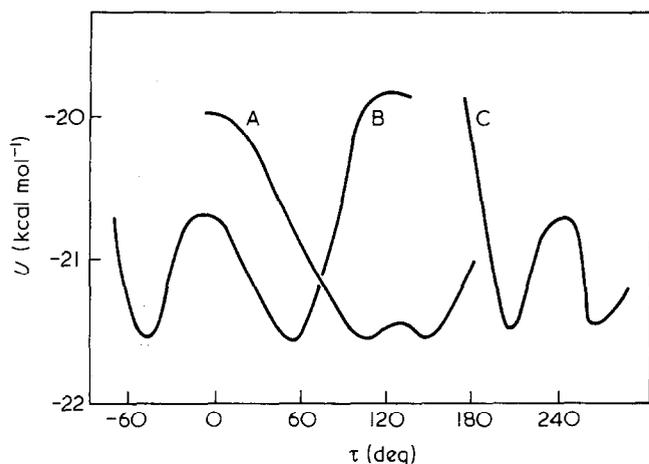
Figure 6 The energy barriers to $a_1 \rightarrow a_2$ and $a_1 \rightarrow a_3$ transitions. Curves A, B and C, correspond to variation of τ_2^2; τ_2^1; τ_3^1; respectively

The occurrence of 'soft' degrees of freedom and distinct conformational states makes model a_0 advantageous from the entropic point of view.

A comparison of model a_0 with model a_6, an analogue of the models in refs. 5 and 9, shows that a_0, while possessing an acceptable R''-factor, is approximately 2 kcal mol^{-1} more stable. In principle, an energy of the order of 1–2 kcal mol^{-1} is within the accuracy ensured by the atom–atom potential method in predictions of the absolute values of the lattice energy. Nevertheless, the 2 kcal mol^{-1} difference between a_0 and a_6 appears to be quite significant. It is apparent that the errors introduced to the lattice energy by assumptions and approximations of the atom–atom potential method are generally systematic and, as a consequence, greatly cancel out in computing the energy difference between different modifications of the same crystal. This has been demonstrated, for instance, by Bernstein and Hagler[23] for polymorphic modifications of N-(p-chlorobenzylidene)-p-chloroaniline.

One further argument in favour of model a_0 over a_6 is provided by comparison of the structure amplitudes for 002 and 004 meridional reflections. Due to the known difficulties in applying Lorentz and polarization corrections to the meridional reflections, their intensities have not been estimated quantitatively[6] and have not been included in the R''-factor. Visually, these reflections are observed as moderately weak and strong, respectively, which qualitatively agrees with model a_0($|F_{002}|/|F_{004}| = 1.5/26.3$) and conflicts with a_6 ($|F_{002}|/|F_{004}| = 27.8/21.8$).

CONCLUSIONS

The detailed scanning of the configurational space of a 16-parameter model of mercerized cellulose has revealed a variety of structures consistent with the available X-ray diffraction data. The use of an additional, energetic criterion, based on the atom–atom approximation to the potential energy, has allowed one antiparallel model (a_0) to be selected as the most probable. The remarkable feature of this model is the ability of its hydrogen bonding network to undergo facile rearrangement through thermal migration of protons. Although the available experimental data are too few to substantiate the optimum model in full detail, the success of the atom–atom potential method in describing the structure and properties of organic crystals supports the predicted structure.

REFERENCES

1 Sarko, A. and Muggli, R. *Macromolecules* 1974, 7, 486
2 Woodcock, C. and Sarko, A. *Macromolecules* 1980, 13, 1183
3 Gardner, K. H. and Blackwell, J. *Biopolymers* 1974, 13, 1975
4 French, A. D. *Carbohydr. Res.* 1978, 61, 67
5 Kolpak, F. J. and Blackwell, J. *Macromolecules* 1976, 9, 273
6 Kolpak, F. J., Weih, M. and Blackwell, J. *Polymer* 1978, 19, 123
7 Jones, D. W. *J. Polym. Sci.* 1960, 42, 173
8 Watanabe, S. and Hayashi, J. *Kogyo Kagaku Zasshi* 1970, 73, 1890
9 Stipanovic, A. and Sarko, A. *Macromolecules* 1976, 9, 851
10 Hunter, R. E. and Dweltz, N. E. *J. Appl. Polym. Sci.* 1979, 23, 249
11 Blackwell, J. and Marchessault, R. H. in 'Cellulose and Cellulose Derivatives' (Eds. Bikales, N. M. and Segal, L.) Part IV, Wiley, N.Y., London, Sydney, Toronto, 1971
12 Kitaigorodsky, A. I. 'Molecular Crystals and Molecules', Academic Press, N.Y., London, 1973
13 Arnott, S. and Scott, W. E. *J. Chem. Soc., Perkin Trans. II* 1972, 2, 324
14 Pertsin, A. J., Nugmanov, O. K., Sopin, V. F., Marchenko, G. N. and Kitaigorodsky, A. I. *Vysokomol. soed.* 1981, 23(A), 2147
15 Zugenmaier, P. and Sarko, A. in 'Fiber Diffraction Methods', Am. Chem. Soc. Symp. Ser. No. 141, 1980, 226
16 Legrand, C. *Acta Crystallog.* 1952, 5, 800
17 Norman, N. *Textile Res. J.* 1963, 33, 711
18 Dashevsky, V. G. 'Konformatsii organicheskikh molekul' (in Russian), Khimia, Moscow, 1974
19 Mirskaya, K. V., Kozlova, I. E. and Bereznitskaya, V. F. *Phys. Stat. Sol.* 1974, 62, 291
20 Kitaigorodsky, A. I., Mirskaya, K. V. and Nauchitel', V. V. *Kristallografia* 1969, 14, 900
21 Hamilton, W. C. *Acta Crystallogr.* 1965, 18, 502
22 Powell, M. J. D. *Computer J.* 1964, 7, 155
23 Newton, M. D., Jeffrey, G. A. and Takagi, S. *J. Am. Chem. Soc.* 1979, 101, 1997
24 Jeffrey, G. A. and Maluszynska, H. *Int. J. Quantum. Chem, Quantum Biol. Symp.* 1981, 8, 231
25 Bernstein, J. and Hagler, A. T. *J. Am. Chem. Soc.* 1978, 100, 673