Crystal structure of cellulose polymorphs by potential energy calculations: 2. Regenerated and native celluloses

A. J. Pertsin, O. K. Nugmanov and G. N. Marchenko
Institute of Elemento-Organic Compounds, USSR Academy of Sciences, Moscow 117334, Vavilov Str. 28, USSR
(Received 21 February 1985; revised 21 May 1985)

Most probable crystal structure models of regenerated and native celluloses are derived through combined optimization of the crystallographic R'' factor and potential energy of the system. The most preferable models of regenerated cellulose are found to be essentially identical to those found previously for mercerized cellulose. The best parallel and antiparallel models of native cellulose are very close in energy and possess statistically equivalent R'' factors. In both models there are intramolecular O(3)H–O(5′) and O(2)H–O(6′) hydrogen bonds and intermolecular O(6)H–O(3″) bonds in the *ac* plane.

(Keywords: cellulose; crystal structure refinement; potential energy calculation)

INTRODUCTION

In the first paper of this series[1] an attempt was described to solve the crystal structure of mercerized cellulose through combined optimization of the crystallographic R'' factor and potential energy of the system. The potential energy was calculated semiempirically, using the molecular mechanics method[2] for the conformational energy and the atom–atom potential method[3] for the intermolecular interaction energy. In this work similar calculations are reported for regenerated cellulose and native cellulose of the perennial plant, ramie (*Boehmeria nivea*).

METHOD OF CALCULATION

The method and strategy of crystal structure determination have previously been described in detail[1]. On generating trial crystal models the conformational parameters of the glucose rings were kept fixed at their standard values, as reported by Arnott and Scott[4]. The monomer residues were linked into the chains with the variable virtual bond method[5,6], assuming the chain symmetry to be 2_1. The chains were positioned in the unit cell according to a $P2_1$ two-chain model, with one chain placed at the origin of the unit cell and the other at the centre of the *ab* plane.

The variable parameters of the crystal model were: the torsional angles τ_i^k describing rotations of the hydroxymethyl group ($i=1$) and the three hydroxyls ($i=2,3,4$) in the origin ($k=1$) and centre ($k=2$) chains; the two angles δ^k describing rotations of the monomer residues about the virtual bonds O(4)–O(4′) in the two crystallographically distinct chains; the angles ϕ^k specifying rotations of the chains about their axes; the relative shift of the chains, s; the polarity parameter p that assumes the value 0 for the antiparallel arrangement of the chains, $+1$ for the parallel 'up' variant and -1 for the parallel 'down' variant; and the average isotropic temperature factor B needed to calculate the structural amplitudes.

The angles τ_i in both chains are defined by the following bond sequences: τ_1 = C(4)–C(5)–C(6)–O(6), τ_2 = C(5)–C(6)–O(6)–H, τ_3 = C(1)–C(2)–O(2)–H, τ_4 = C(2)–C(3)–O(3)–H. Each angle is zero when the corresponding bond sequence, A–B–C–D, is *cis*. Anticlockwise rotation of the C–D bond when looking down the B–C bond represents positive rotation.

To define the angle δ, three vectors, ρ, ξ and λ, are used, all emanating from the chain origin, O(4). The radius vector ρ is chosen to be perpendicular to the chain axis, ξ to be along the O(4)–C(4) bond and λ along the virtual bond O(4)–O(4′). The angle δ is introduced as the dihedral angle between the plane defined by λ and ξ, and the plane defined by λ and ρ; δ is positive when ρ, λ and ξ form a right-handed set.

The angle ϕ is defined as the angle formed by the vector $-\rho$ and the unit cell axis a^*. The chain position with O(4) at $(0, -y, 0)$ for the origin chain and $(\frac{1}{2}, \frac{1}{2}-y, z)$ for the centre chain corresponds to $\phi^1 = \phi^2 = -\pi/2$.

The 'up' direction of a chain corresponds to $z_{C(1)} > z_{C(4)}$. The glycosidic oxygen O(4) of the origin chain is always kept at $z=0$, while that of the centre chain is at $z=s$.

The quality of a trial crystal model was assessed in terms of the objective function

$$\Phi = U + WR'' \qquad (1)$$

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 intra- and intermolecular contributions. Both contributions involve non-bonded atom–atom potentials of the six-exponential type and also Morse potentials to describe hydrogen bonds. The intramolecular energy includes, in addition, torsional contributions and a glycosidic bond angle bending potential. The parameters of the potential

0032-3861/86/040597-05$03.00

functions used are the same as in the previous work[1].

For regenerated cellulose the R'' factor was calculated using the diffraction data of Kolpak and Blackwell[7]. For ramie cellulose the data of Mann *et al.*[8] were used, as refined by French[9]. These latter data included the intensities of 0 0 2 and 0 0 4 meridional reflections, and for this reason they were preferred to more recent data of Woodcock and Sarko[10].

The weighting factor W in (1) was chosen on the grounds that the smallest change in WR'', which is statistically significant for a 1% significance level, will be about the same as the accuracy of calculating U. In the previous work[1] the uncertainties in R'' and U were adopted to be roughly 0.07 and 1 kcal mol^{-1}, respectively, which resulted in $W \approx 15$ kcal mol^{-1}.

For the celluloses studied in this work the weighting factor would strictly speaking have to be corrected since the number of observed reflections in regenerated and ramie celluloses is different from that in mercerized cellulose. However, considering that the above choice of weighting factor is very approximate, it appeared more reasonable to leave W unchanged. The advantage of using the same W was that the values of the objective function were compatible in all three celluloses.

RESULTS AND DISCUSSION

Regenerated cellulose

X-ray diffraction diagrams of regenerated and mercerized celluloses are very similar, so that these celluloses are commonly thought to be representatives of the same cellulose polymorph, cellulose II. This view is supported by the calculation results summarized in *Table 1*. Given in this table are the structural and energetic parameters for the eight deepest minima found in seeking the global minimum of Φ. A comparison of the optimized model parameters in *Table 1* with those found previously for mercerized cellulose (see Table 3 in ref. 1) shows that the best crystal models of the two celluloses are essentially identical. A typical difference in the model parameters is 3–4° for the angles and 0.01–0.02 for the shift s.

Inspection of the hydrogen bonding systems in the best models of regenerated cellulose also shows no significant changes as compared to the corresponding results for mercerized cellulose[1]. The only marked difference between the present and previous results is in the average isotropic temperature factor B: while in mercerized cellulose it was about 33 Å^2, in the best models of regenerated cellulose it ranges from 19 to 21 Å^2. The difference in B hardly reflects the actual difference in thermal motion and most likely stems from different crystallinities of the two celluloses.

As with mercerized cellulose[1], there are three best models of regenerated cellulose, a_1, a_2 and a_3, which are nearly equivalent in energy and separated from the other models by an energy gap of about 1 kcal mol^{-1}. *Table 1* shows that model a_1 may be transformed to a_2 by rotation of the HO(6$_1$) hydroxyls (see parameters τ_2^1 and τ_3^1) and to a_3 by rotation of HO(6$_2$) (see parameter τ_2^2). These transformations affect only the hydrogen bonding network, while the basic structure remains practically unchanged. A calculation of the energy barriers to the $a_1 \rightarrow a_2$ and $a_1 \rightarrow a_3$ transformations yields values of 0.9 and 0.2 kcal mol^{-1}, respectively. All this allows one to regard the best three models of regenerated cellulose as a single model with a mobile hydrogen bonding network.

Native cellulose

Native cellulose, usually known as cellulose I, possesses a higher degree of crystallinity than cellulose II and gives more resolved diffraction patterns. Typical X-ray diagrams of cellulose I from ramie allow the measurement of 30–40 reflection intensities. As with cellulose II, however, such diffraction data prove to be too few to determine the crystal structure unambiguously, even with the use of stereochemical and packing constraints. The difficulties with determination of the crystal structure of cellulose I can be appreciated by comparing the most probable models suggested for cellulose I by Woodcock and Sarko[10] and by Gardner and Blackwell[11], on the one hand, with that by French[9], on the other. Even though they are derived from similar diffraction data, these models differ in such a fundamental characteristic as the chain polarity, let alone the other details of the structure.

The ambiguity involved in interpretation of the X-ray diffraction data for cellulose I is seen well from *Table 2*, which presents the parameters of 13 local minima found in the search for the global minimum of the objective function Φ. The minima are labelled as a_j, p_j and $\bar{p}_j$ depending on the chain polarity: a_j corresponds to the antiparallel packing, p_j and $\bar{p}_j$ to the parallel 'up' and 'down' variants, respectively. Analysis of the R'' factors in *Table 2* with the use of Hamilton's tests[12] shows that at the usual significance level of 1% most of the structural models presented are statistically indistinguishable. As with cellulose II[1], the calculation reveals several models ($\bar{p}_7$, p_5 and a_{13}) that have a very low R'' factor, while being completely unsatisfactory from the energetic point of view. The latter result indicates once again the inadequacy of the R'' factor as a criterion for unambiguous solution of the structural problem for cellulose.

Among the parallel models in *Table 2* there is a model, p_1, that is close in its parameters and hydrogen bonding network to the model suggested by Gardner and Blackwell[11] and, more recently, by Woodcock and Sarko[10]. Thus, transformation of the model parameters reported in ref. 11 to the conventions used here yields

Table 1 The structural and energetic parameters for eight models of regenerated cellulose (the units are degrees and kcal mol^{-1})

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	−50	−151	176	42	−84	69	162	−53	163	47	−5	0.117	0.216	−21.5	−18.3
a_2	64	60	−86	177	40	−83	70	170	−51	170	47	−3	0.144	0.212	−21.8	−18.6
a_3	67	−51	−152	176	41	−86	69	101	−55	162	45	−2	0.116	0.216	−21.5	−18.3
a_4	−63	171	−158	178	39	−89	67	162	−52	172	46	−3	0.125	0.191	−20.3	−17.4
a_5	67	−52	−154	−54	41	−85	66	101	−55	160	47	−5	0.111	0.219	−20.5	−17.3
a_6	−181	162	−97	176	42	−83	70	162	−56	168	43	−1	0.216	0.161	−19.7	−17.3
a_7	67	−152	−153	176	41	−88	69	37	−53	69	46	0	0.128	0.232	−20.9	−17.3
p_1	62	68	−70	171	45	−75	−168	171	−75	165	48	79	0.133	0.183	−19.6	−16.9

Table 2 The structural and energetic parameters for different models of native cellulose (the units are degrees and kcal mol^{-1})[a]

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
$\bar{p}_1$	70	164	−54	176	45	27	71	164	−54	177	44	−150	0.218	0.165	−19.9	−17.4
$\bar{p}_2$	70	163	−56	175	43	27	70	76	−56	174	42	−143	0.202	0.187	−19.7	−16.9
a_1	68	163	−52	174	43	−39	67	163	−53	173	46	25	0.155	0.213	−19.7	−16.5
a_2	67	166	−51	174	44	−41	68	165	−53	169	48	23	0.645	0.155	−18.6	−16.1
a_3	75	163	−54	180	38	−49	73	68	−56	177	39	35	0.084	0.240	−19.7	−16.0
a_4	74	163	−55	180	38	−48	75	61	−56	62	38	37	0.083	0.240	−19.7	−16.0
a_5	61	−52	92	163	48	−34	70	162	−56	176	43	26	0.210	0.218	−18.9	−15.6
a_6	69	161	−53	175	43	−38	65	−56	−151	169	47	21	0.173	0.208	−18.5	−15.4
p_1	70	164	−54	176	43	−43	67	167	−52	172	46	−43	0.302	0.167	−17.5	−15.0
p_2	67	167	−58	173	44	63	77	162	59	187	33	39	0.270	0.196	−17.5	−14.6
$\bar{p}_7$	74	136	−131	108	60	10	45	172	−191	88	65	163	0.209	0.113	4.7	6.4
p_5	51	148	−170	119	59	−33	65	147	−152	102	65	−23	0.283	0.127	11.8	13.7
a_{13}	50	151	−191	120	65	−26	63	155	−158	88	60	196	0.143	0.109	14.9	16.5

[a] The numbering of models in each given variant of the chain packing follows the order of increasing Φ. For reasons of space, some of the models found are not included in the table

$\tau_1^1 = \tau_1^2 \approx 80°$, $\phi^1 = \phi^2 \approx -52°$, $s = 0.266$, $R'' = 0.210$, which is similar to the p_1 model parameters.

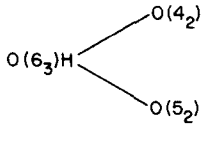
The p_1 model structure represents an array of alternating sheets parallel to the *ac* plane, each formed by translationally equivalent chains. The chains are linked into the sheets with two crystallographically distinct intermolecular hydrogen bonds of the type O(6)H–O(3″). These are O(6$_1$)H–O(3$_1^a$) and O(6$_3$)H–O(3$_3^a$) bonds*. The chain conformation in the (0 1 0) sheet (origin chains) is similar to that in the (0 2 0) sheet (centre chains), both being close to conformation B1 of an isolated cellulose chain[1,5] with its inherent intramolecular hydrogen bonds O(3)H–O(5′) and O(2)H–O(6′).

As seen from *Table 2*, model p_1 is the best of the parallel 'up' models but it is markedly inferior to the best antiparallel and parallel 'down' models.

Of the antiparallel models in *Table 2*, a_2 is noteworthy, being similar to the antiparallel model suggested by French[9] ($\tau_1^1 = \tau_1^2 \approx 63°$, $\phi^1 = -29°$, $\phi^2 = 29°$, $s = -0.360$, $R'' = 0.158$). The chain conformation and the system of hydrogen bonds in model a_2 are identical to those in p_1. In both the origin and centre chains there are two intramolecular hydrogen bonds, O(3)H–O(5′) and O(2)H–O(6′), with energies of about 4 kcal mol^{-1}. The neighbouring chains along *a* are linked into sheets with intermolecular hydrogen bonds, O(6$_1$)H–O(3$_1^a$) and O(6$_3$)H–O(3$_3^a$) (≈ 3.9 kcal mol^{-1}).

The global minimum of Φ occurs within the parallel 'down' packings of the chains (see model $\bar{p}_1$ in *Table 2*). *Figure 1* shows the projections of the $\bar{p}_1$ model structure down the *c* and *b* axes and *Table 3* lists the fractional unit cell coordinates of the corresponding asymmetric unit. The hydrogen bonding system in model $\bar{p}_1$ is completely identical to those in p_1 and a_2. The basic differences from p_1 are the chain direction and the relative shift of the chains. These differences leave the R'' factor almost unaffected but make the $\bar{p}_1$ model 3 kcal mol^{-1} more preferable in energy.

Table 2 shows that model $\bar{p}_1$ can be transformed to the next model, $\bar{p}_2$ by rotation of the O(6$_3$)H hydroxyl (see parameter τ_2^2), with the other structural parameters changed only slightly. The $\bar{p}_1$ to $\bar{p}_2$ transition leads to breakdown of the intermolecular intrasheet O(6$_3$)H–O(3$_3^a$) bond and formation of an intersheet two-acceptor bond,



Models $\bar{p}_1$ and $\bar{p}_2$ are very close in energy and separated by a low energy barrier (see *Figure 2*). Thus, as with cellulose II[1], one may suppose that models $\bar{p}_1$ and $\bar{p}_2$ may coexist in the crystal and transform to one another through thermal migration of the protons at O(6$_3$).

The best antiparallel model is a_1. It differs from a_2, an analogue of the French model[9], mainly in the shift parameter s (by about $\frac{1}{2}c$). The chain conformation and the hydrogen bonding system in both models are very similar but the difference in s makes a_1 about 1 kcal mol^{-1} more stable. Two projections of the a_1

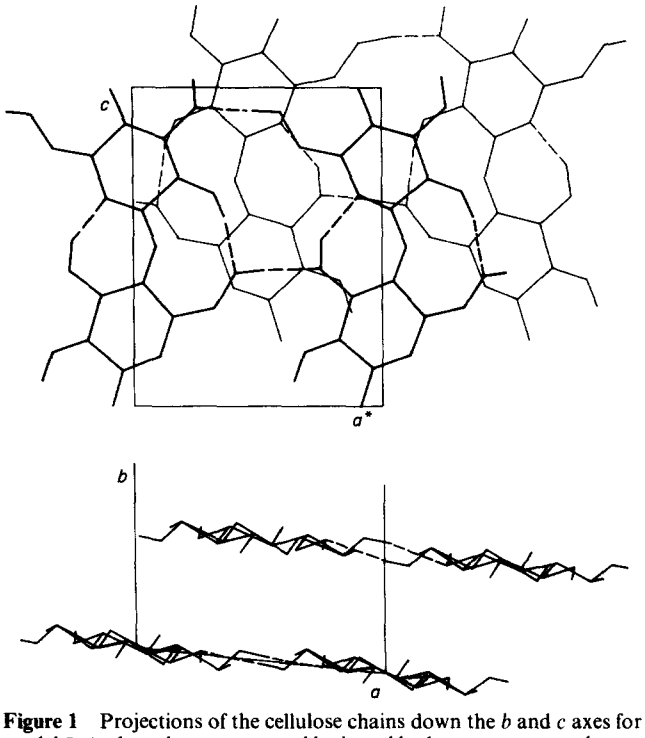


Figure 1 Projections of the cellulose chains down the *b* and *c* axes for model $\bar{p}_1$ (only carbon oxygen and hydroxyl hydrogen atoms are drawn, for clarity; hydrogen bonds are indicated by broken lines)

*The symbols used in labelling hydrogen bonds are the following: subscript refers to one of the four monomer residues in the unit cell (1 and 2 are in the origin chain, while 3 and 4 in the centre chain); superscript indicates translation applied to the basis residue.

Table 3 Fractional unit cell atomic coordinates for the best parallel and antiparallel models of cellulose I ($\times 10^3$)

	Origin chain			Centre chain		
Atom	*X*	*Y*	*Z*	*X*	*Y*	*Z*
Model $\bar{p}_1$						
O(1)	89	57	0	413	438	218
O(2)	−324	52	−329	827	464	−111
O(3)	−257	7	−62	758	519	156
O(4)	−89	−57	500	−413	−438	718
O(5)	100	−53	340	397	548	−121
O(6)	406	29	−84	93	450	133
C(1)	−43	20	−383	543	483	−165
C(2)	−182	−22	−283	681	531	−68
C(3)	−130	43	−153	632	463	65
C(4)	31	−21	−144	468	519	105
C(5)	160	18	−219	340	474	0
C(6)	319	−56	−190	177	541	29
C(1)H	−15	157	−391	521	345	−173
C(2)H	−214	−159	−283	708	668	−80
C(3)H	−106	182	−155	615	324	45
C(4)H	9	−158	−99	484	656	105
C(5)H	188	155	−228	318	337	−27
C(6)H(1)	293	−191	−169	198	676	38
C(6)H(2)	397	−47	−275	100	530	−72
O(2)H	−368	27	−416	870	490	−214
O(3)H	−247	31	27	748	484	228
O(6)H	522	15	−70	−24	459	103
Model a_1						
O(1)	78	−55	0	592	552	161
O(2)	−334	−62	330	178	562	−167
O(3)	−257	−1	2	244	498	99
O(4)	−78	55	500	−592	−552	661
O(5)	111	56	339	598	445	−180
O(6)	401	−17	84	903	513	77
C(1)	−46	−21	383	458	523	−222
C(2)	−178	17	287	318	483	−125
C(3)	−138	−48	153	372	546	9
C(4)	35	22	114	531	477	47
C(5)	157	−14	218	660	514	−58
C(6)	331	66	189	817	434	−31
C(1)	−44	−158	391	489	660	−229
C(2)H	−185	153	283	282	347	−122
C(3)H	−141	−186	155	399	684	8
C(4)H	38	158	99	506	340	60
C(5)H	160	−150	228	692	651	−67
C(6)H(1)	329	200	167	788	299	−12
C(6)H(2)	407	59	274	896	446	−115
O(2)H	−372	−47	417	135	540	−254
O(3)H	−255	−41	−26	254	541	188
O(6)H	520	0	70	19	498	91

model structure are shown in *Figure 3* and the coordinates of its asymmetric unit are listed in *Table 3*.

Scanning the energy surface near the local minima corresponding to models a_1, a_3 and a_4 shows that a_1 is separated from a_3 and a_4 by energy barriers of 0.4 and 1.1 kcal mol^{-1}, respectively. In this case, however, the transitions $a_1 \rightarrow a_3$ and $a_1 \rightarrow a_4$ seem to be very unlikely in view of an appreciable difference (≈ 0.7 Å) in the relative shift of the chains between the models. The realization of such a shift in the crystal involves a cooperative move of the atoms within the whole sheet, which is indeed a highly improbable event. (Formally, this means that the pre-exponential factors in the expressions for the probability of the a_1 to a_3 and a_1 to a_4 transitions are negligible).

Unlike cellulose II, whose best models were separated from the others by a considerable energy gap (see *Table 1*), the energy distribution of the best models of cellulose I is more or less uniform. What is more, the best parallel and antiparallel models of cellulose I are close in energy and possess statistically equivalent R'' factors ($R''(a_1)/R''(\bar{p}_1) \approx 1.3$). This does not allow one to discriminate, with certainty, between the two variants of the chain packing.

A weighty argument in favour of the antiparallel structure is provided by the experimental fact that mercerization, leading to the cellulose I→cellulose II transition, involves only slight swelling of the cellulose fibres. This cannot produce drastic structural changes (e.g. in the chain polarity) and hence implies a definite similarity in structure between the two celluloses. A comparison of the structural parameters in a_1 models of celluloses I and II shows that these models differ mainly in orientation of the O(6)H and O(2)H hydroxyls. As to the orientation of the hydroxymethyl group and the relative shift of the chains, these are very similar in both a_1 models.

In conclusion we return to *Tables 1* and *2* and note that the best crystal models of cellulose II are about 1.5 kcal mol^{-1} lower in energy than the best models of cellulose I. This is in agreement with the commonly accepted view that cellulose II is a more stable polymorph of cellulose.

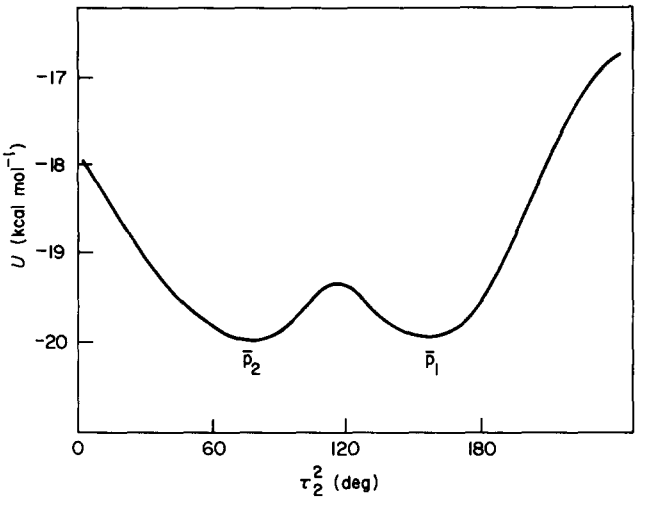


Figure 2 Energy barrier to $\bar{p}_1 \rightarrow \bar{p}_2$ transformation

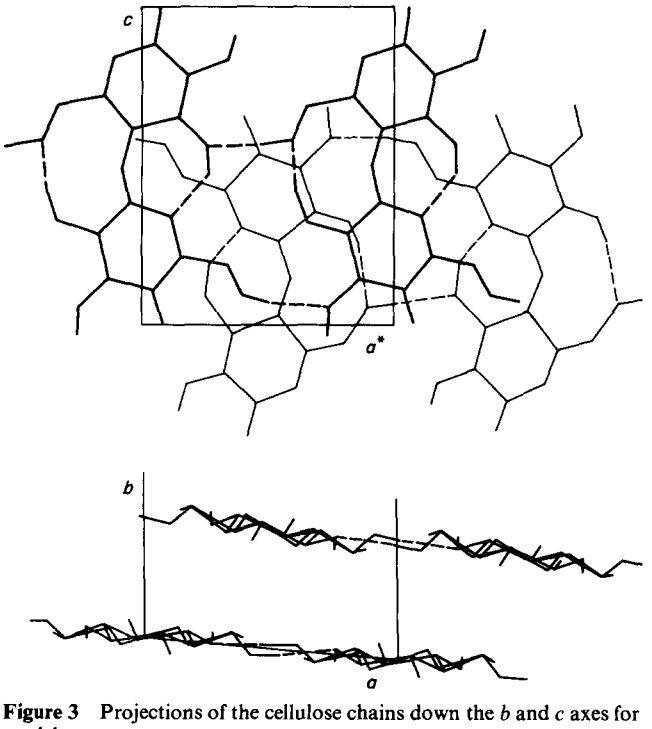


Figure 3 Projections of the cellulose chains down the *b* and *c* axes for model a_1

REFERENCES

1 Pertsin, A. J., Nugmanov, O. K., Marchenko, G. N. and Kitaigorodsky, A. I. *Polymer* 1984, **25**, 107
2 Bunkert, U. and Allinger, N. L. 'Molecular Mechanics', ACS Monograph No. 177, American Chemical Society, Washington DC, 1982
3 Kitaigorodsky, A. I. 'Molecular Crystals and Molecules', Academic Press, New York and London, 1973
4 Arnott, S. and Scott, W. E. *J. Chem. Soc., Perkin Trans. II* 1972, **2**, 324
5 Pertsin, A. J., Nubmanov, O. K., Sopin, V. F., Marchenko, G. N. and Kitaigorodsky, A. I. *Vysokomol. Soed. A* 1981, **23**, 2147
6 Zugenmaier, P. and Sarko, A. in 'Fiber Diffraction Methods', ACS Symp. Ser. No. 141, American Chemical Society, Washington DC, 1980, p. 226
7 Kolpak, F. J. and Blackwell, J. *Macromolecules* 1976, **9**, 273
8 Mann, J., Roldan-Gonzales, L. and Wellard, H. J. *J. Polym. Sci.* 1960, **42**, 165
9 French, A. D. *Carbohydr. Res.* 1978, **61**, 67
10 Woodcock, C. and Sarko, A. *Macromolecules* 1980, **13**, 1183
11 Gardner, K. H. and Blackwell, J. *Biopolymers* 1974, **13**, 1975
12 Hamilton, W. C. *Acta Crystallogr.* 1965, **18**, 502