Translated from *Uspekhi Khimii*, **56**, 1339-1359 (1987) U.D.C. 543.422

The Molecular-crystal Structure of Cellulose

O.K.Nugmanov, A.I.Pertsin, L.V.Zabelin, and G.N.Marchenko

The structural models of the principal polymorphic varieties of cellulose proposed by different workers during the last two decades are analysed, the results obtained by the main structural methods are considered, and the reliability of a particular structural model is estimated. The ambiguity of the solution of the structural problems for cellulose in the cases of the existing diffraction data, even when stereochemical and packing limitations are invoked, is demonstrated. The results of semiempirical energy calculations for the stable structure of cellulose are discussed.
The bibliography includes 106 references.

CONTENTS

I. Introduction 764
II. General information about the structure of cellulose 764
III. Structural models of the main polymorphic varieties of cellulose 769
IV. Conclusion 774

I. INTRODUCTION

The processes occurring in the industrial processing of cellulose are as a rule heterogeneous and take place under diffusional conditions, which determines the dependence of the degree and kinetics of the reactions occurring on the structure of the solid polymer phase.

Several hierarchical levels of structural organisation can be differentiated in cellulose. These are the supermolecular structure with a characteristic distance ranging from several tens of angstroms and above and two lowest levels—crystalline and molecular, with characteristic distance of one angstrom. The higher structural levels are readily susceptible to study by electron and optical microscopic methods and low angle X-ray diffraction and have been fairly fully investigated.[1-8] The study of the molecular-crystal structure of cellulose constitutes the main problem.

Cellulose was one of the first polymers subjected to structural analysis. However, despite the large number of crystallographic studies of cellulose, the detailed structure of the crystalline region still remains the subject of extensive controversy.[9-11] This applies to all the known polymorphic varieties of cellulose, including the most thoroughly investigated varieties—cellulose I (CI, native) and cellulose II (CII, regenerated and mercerised). A large number of structural models, which conflict as regards fundamental characteristics such as the symmetry of the crystal lattice, the relative orientation of the chains, and the system of hydrogen bonds let alone even finer structural details, have been proposed.

The present review analyses the literature data on the molecular and crystal structures of CI and CII, the principal results of the structural investigations are considered, and the reliability of the existing concepts and structural models is estimated.

II. GENERAL INFORMATION ABOUT THE STRUCTURE OF CELLULOSE

Cellulose (C) consists of poly-(1 → 4)-β-D-glucopyranose, $(C_6H_{10}O_5)_n(C_6H_{12}O_6)_2$, where n (the degree of polymerisation) varies in the range 800–8000 depending on the nature of the specimen.[9-13] The structural formula of cellulose is presented in Fig.1, which also indicates the numbering of the atoms adopted in the stereochemistry of sugars and carbohydrates.[14-21]

1. The Symmetry and Parameters of the Unit Cell

The principal source of information about the molecular-crystal structure of cellulose is X-ray diffraction of highly oriented fibres. Typical X-ray diffraction patterns of cellulose contain darkened regions distributed continuously along layer lines which are caused by the one-dimensional diffraction and characterise the degree of ordering of the centres of gravity of the scattering molecules. The discrete spots of different intensity are due to three-dimensional diffraction. The number of independent spots of this type does not usually exceed 30–40.

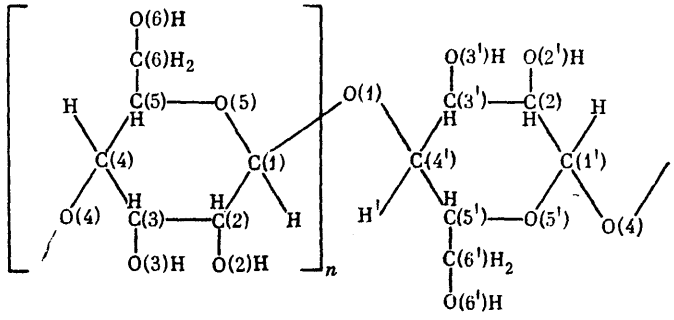


Figure 1. The structural formula of cellulose.

The paucity of the diffraction pattern is to some extent characteristic of the X-ray diffraction of polymers,[22-24] but, compared with the majority of synthetic polymers, the X-ray diffraction patterns of cellulose exhibit a steeper decrease of

intensity with increase of the Bragg angle, which additionally limits the region of the possible reflections.[25] Such decrease of intensity can be caused either by the intense thermal movement of the structural elements of cellulose or by statistical disorder.

The only symmetry element of the crystal structure of cellulose is a twofold screw axis parallel to the axis of the fibre (space group $P2_1$).[25] The presence of this axis is indicated by the systematic extinction of the $00l$ reflection $l = 2n + 1$ on the X-ray diffraction patterns of cellulose.

Among the observed reflections, the strongest are the $hk0$ reflections, the most intense being those for which $h + k = 2n$ (i.e. 200, 110, etc.). This shows that the centres of the scattering molecules are separated by a distance $R(x, y, z)$ with $x = a/2$ and $y = b/2$. Consequently, two chains pass through the unit cell of cellulose, one of which is the "angular" chain with the coordinates (0, 0, z), and the other is the "central" chain with the coordinates (1/2, 1/2, z). Two versions of the arrangement of chains are then possible. In the first the chains occupy common positions and are linked by a screw axis in the (1/4, 1/4, z) position. In this case the symmetrically independent unit of the crystal and the repeat unit of the polymer chain consists of a cellobiose fragment, the chains have the same direction (are parallel), and are displaced relative to one another by half the translation along the c axis. In the second version, the chain structures are located on the 2_1 screw axes, are crystallographically non-equivalent, and can be both parallel and anti-parallel. The symmetrically independent unit of the crystal consists of a pair of glucose residues (one in each chain); the repeat unit is also a glucose residue.

The choice between the above two versions of the structure is made on the basis of the intensity of the meridional 002 and 004 reflections. In the first version of the packing, the centres of the glucose units are displaced by $c/2$ along the z axis. For this version, the 002 reflection should be more intense than the 004 reflection, which conflicts with experiment. In the second version of the structure, the displacement of the glucose rings of the neighbouring chains is arbitrary and there are no limitations as regards the ratio of the intensities of the 002 and 004 reflections.

Thus the unit cell of cellulose, constructed on the basis of X-ray diffraction data, has the space group $P2_1$ (monoclinic system) and contains two crystallographically non-equivalent chains intersecting the ab plane at the origin of coordinates and at the centre. The symmetrically independent unit of each cell is the glucopyranose residue. In order to account for the observed intensity ratio of the even meridional reflections, it is usually assumed that the central chain in the two-chain unit cell is displaced by approximately $c/4$ relative to the angular chain.

The repeat period along the axis of the cellulose fibre, calculated from the distance between the meridional reflections on the X-ray diffraction pattern, is approximately 10.3 Å. The remaining unit cell parameters are found by indexing the diffraction patterns. Wellard[26,27] first drew attention to the fact that, even within the framework of a single polymorphic modification of cellulose, the unit cell parameters are within a fairly wide range. For example, the parameter a in CI varies within the limits 8.17–8.35 Å, the parameter b within the limits 7.78–7.90 Å, and the monoclinic angle γ in the range 96.0–96.5°. In CII the ranges of variation of the parameters are even greater: 7.83–8.14 Å for a, 8.99–9.38 Å for b, and 116.6–118° for γ. It is believed[3] that celluloses of different origin do indeed have different unit cells. The unit cell parameters depend on the content of impurities, the defectiveness of the crystallites, the heterounit nature of the polymer chain, and other factors, which can be determined to a large extent by the methods used to isolate and purify cellulose.

The model of the cellulose unit cell described above has been frequently criticised. Thus a number of workers[28–30] observed weak odd meridional $00l$ reflections, which cast doubt on the presence of the 2_1 axis in the structure. Special studies have shown, however, that these reflections are very weak and may be accounted for by the overlapping of neighbouring reflections.[31,32]

Norman[32] and Kitaigorodskii[34] drew attention to the fact that the appearance of odd meridional reflections may be a consequence of one-dimensional diffraction. However, one-dimensional diffraction calculations for the third layer line yielded an appreciably smaller intensity (by approximately 30%) than its observed value. According to Kitaigorodskii and Tsvankin,[33,34] cellulose nevertheless possesses a screw axis symmetry and the differences between the calculated and observed intensities can be associated with structural defects.

A simple model of the defects in the crystal structure of cellulose has been examined by Stipanovic and Sarko,[35] who suggested, as regards the calculation of the theoretical intensities, that the hydroxymethyl groups of the neighbouring residues in the chain are in different positions linked to one another by 120° rotation. However, the theoretical intensities of the 001 and 003 reflections, calculated for the given disordered model, proved to be much higher than the observed values, which permitted the conclusion that, even if orientational disorder of this kind is present in the cellulose structure, it involves only very few chains, for example, the chains located on the surfaces of crystallites.

Apart from diffraction data, NMR and IR spectroscopic data have been invoked for the discussion of the model of the cellulose unit cell described at the beginning of this Section. Thus the splitting of the signals due to the C(1) and C(4) atoms has been observed in the high resolution NMR spectrum of solid cellulose.[36,37] These two well resolved doublets indicated the presence in cellulose of symmetrically different pairs of C(1) and C(4) atoms and were interpreted as proof of the absence of the 2_1 screw axis from cellulose. An analogous situation obtained also in the study of the IR spectra,[38] where the bands with parallel dichroism at 3448 and 3488 cm^{-1} were attributed to two non-equivalent O(3)H...O(5′) hydrogen bonds alternating along the axis of the cellulose macromolecule.

The use of the above experimental facts as proof that the cellulose chain is not a helix with 2_1 symmetry constitutes in our view a misunderstanding. As noted above, the cellulose unit cell, constructed using the X-ray diffraction data, includes two crystallographically non-equivalent chains. In these chains the C(1) and C(4) atoms can have different local environments just as the O(3)H...O(5′) hydrogen bonds can have different structures. These differences are fully capable of giving rise to the difference in the NMR and IR spectra, so that their interpretation does not in any way make it necessary to "deprive" the cellulose chain of 2_1 symmetry.

In discussing the two-chain model of the cellulose unit cell with $P2_1$ symmetry, it is necessary to mention separately the structural studies of certain specific varieties of natural cellulose. Thus the X-ray diffraction patterns of seaweed[28,39] (*Valonia ventricosa*, *Chaetomorpha*) were found to contain three weak reflections (d = 8.61 and 4.42 Å along the first layer line and d = 3.29 Å along the third), which are absent from the X-ray diffraction patterns of cotton, ramie, flax, and wood cellulose. These reflections cannot be indexed on the basis of a two-chain unit cell model and require the doubling of the parameters a and b, which formally leads to models with eight crystallographically non-equivalent chains per unit cell. It may be that such non-equivalence of the

chains in the increased cell reduces to a simple shift of the neighbouring chains relative to one another by $c/4$ along the axis of the fibre.[40,41]

Yet another feature of the cellulose from seaweed was observed in an electron diffraction study of specimens with biaxial surface orientation.[29,30] It was found that the intensities of certain hkl and $\bar{h}kl$ reflections do not coincide, which requires the reduction of this symmetry to triclinic.

Despite the fact that the electron diffraction data[29,30] are not entirely convincing (owing to the strong dependence of the intensity of the reflections on the orientation of the reflecting planes), it should probably be recognised that cellulose from seaweed can indeed have an "individual" unit cell.

2. Conformation of the Monomer Unit

Whereas the symmetry and size of the cellulose unit cell have been determined by direct experimental studies, unfortunately only indirect data are available as regards the conformation of the monomer unit of cellulose. These are in the first place X-ray diffraction data for crystals of low-molecular-weight sugars and carbohydrates and also the results of theoretical conformational analysis.

It has been established experimentally and theoretically[18-21,42-47] that β-D-glucose, which is the elementary unit of the cellulose helix, has the "chair" form and is in the *Cl* conformation in terms of Reeves' nomenclature.[48-50] Three independent X-ray diffraction studies of β-cellobiose[51-53] showed that the conformation of its monomer units corresponds to the *Cl* form. Arnott and Scott[54] analysed the X-ray diffraction studies of the single crystals of low-molecular-weight sugars and carbohydrates. Statistical analysis of 27 model compounds, whose monomer unit is a pyranose ring (among these, 17 compounds contained a glucopyranose ring), showed that they all have the *Cl* conformation. The standard deviations of the geometrical parameters of the ring from the corresponding average values are extremely insignificant, which indicates a high conformational rigidity of the pyranose rings.

Conformational analysis of D-aldopyranoses, carried out by Rao and co-workers[55,56] and then in other investigations,[18,45-47] showed that, whatever the spatial disposition of the side groups, the *Cl* form is preferable on energy grounds and the ring geometry depends only slightly on the perturbing influence of the side groups. The valence angles at the carbon atoms remain close to the tetrahedral value and the angle at the oxygen ring atom is 112–114°. We may note that the ring conformation remains unaltered even in the presence of bulky substituents such as the nitrate groups.[57]

3. The Conformation of the Cellulose Helix

Bearing in mind the high conformational rigidity of the glucopyranose ring and also the constancy of the valence bonds and angles in the side groups, one may assume that the model of the cellulose chain which permits the variation of only the orientations of the side groups and of the parameters of the junction of the monomer units is fully acceptable. The valence angle β of the glucosidic oxygen atom (Fig.2) and also the torsional angles ϕ and ψ for the rotation of the monomer units about the C(1)–O(1) and O(4)–C(4′) bonds are usually employed to describe the geometry of the junction of the monomer units. The angles ϕ and ψ are determined by the sets of four atoms H(1)–C(1)–O(1)–C(4′) and C(1)–O(4)–C(4′)–H(4′) respectively.

The angles β, ϕ, and ψ are extremely convenient for the description of the geometry of the junction, since they make it possible to use standard computer programs for conformational analysis in order to generate the fragments of the cellulose macromolecule. At the same time, the parameters ϕ, ψ, and β suffer from a significant disadvantage, namely the cellulose chain does not meet the symmetry requirements for their arbitrary values. In other words, in generating the model of the cellulose chain with the aid of the above parameters, one cannot make use of the experimental fact that the cellulose chain is a helix with 2_1 symmetry.

Yet another experimental result, which cannot be used to describe the junction of the monomer units in terms of the angles ϕ, ψ, and β is the fact that the repeat period of the cellulose chain is approximately 10.3 Å and remains the same in all the polymorphic modifications of cellulose.[7-11]

It is easy to show that, for the specified conformation of the monomer unit and symmetry of the helix (2_1), only two parameters are needed to describe the conformation of the main chains. It has been suggested that the angle δ, determining the rotation of the monomer unit about the virtual O(4)...O(4′) bond, and the angle ε, determining the slope of this virtual bond relative to the axis of the helix (Fig.2), be used as such parameters.[58,59] The idea of switching to parameters describing the junction of not real but virtual bonds was developed in the so called variable virtual bond method.[60,61] The independent conformational parameters of the main chain are the angle δ and the half-period $c/2$ (or h), while the conformational parameters of the junction (ϕ, ψ, and β) are dependent variables. This algorithm is applicable to molecules with any structure of the main chain provided that its symmetry is known.

Passing now to the determination of the parameters specifying the orientation of the side groups in the cellulose chain, we may note immediately that in the vast majority of structural

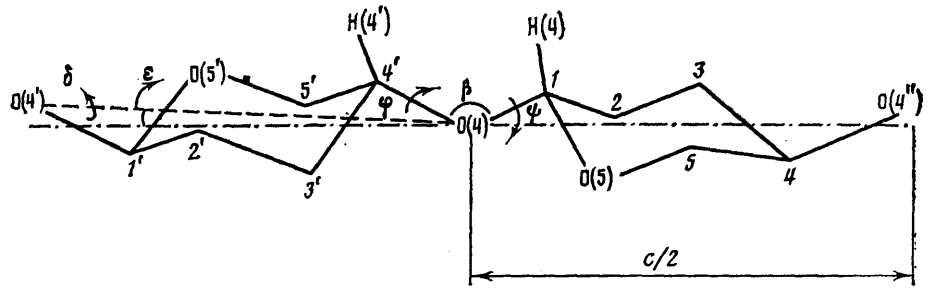


Figure 2. The structural parameters used to describe the conformation of the main chain of the macromolecule.

studies on cellulose hydrogen atoms are disregarded.[28,35,62-64] In this case it is sufficient to introduce only one angle, describing the rotation of the hydroxymethyl group about the C(5)–C(6) bond. The following designations have been adopted for the description of the position of the hydroxymethyl group:[65]

(a) *gt*—*gauche*-position of the C(6)–O(6) bond relative to the C(5)–O(5) bond and *trans*-position relative to the C(4)–C(5) bond;

(b) *tg*—*trans*-position of the C(6)–O(6) bond relative to the C(5)–O(5) bond and *gauche*-position relative to the C(4)–C(5) bond;

(c) *gg*—*gauche*-position relative to both bonds.

A number of investigators believe[36] that the *gt*-orientation of the CH_2OH group is most probable for cellulose chains, because this conformation is typical of crystalline monosaccharides[46,47,66] and certain oligosaccharides.[51-53,67,68] However, experimental data for cellotetraose[69,70] agree best with the hypothesis of the *tg*-conformation of the hydroxymethyl group.

In the entire set of parameters describing the conformation of the cellulose chain, the angle β at the glucosidic oxygen atom has been most reliably determined. According to the statistics of Arnott and Scott,[54] β varies in the range 115.7 to 117.6° in model compounds whose monomer unit is the glucopyranose ring. This range can be adopted as the most probable range of variation of the angle β.

In the early structural models of cellulose, proposed by Meyer and co-workers,[71,72] the angle β was assumed to be 109° and the torsional angles φ and ψ were assumed to be zero. (The conformation of the chain with $\phi=\psi=0°$ is usually referred to as the "straight" conformation, in which the plane of the glucosidic bonds is perpendicular to the plane of the glucopyranose ring). The hydroxymethyl group is close to the *tg*-position. For this set of parameters, there are no intermolecular hydrogen bonds in the system. The straight conformation can hardly be regarded as acceptable owing to the presence of certain short non-valence contacts between successive units [in particular the distance between the H(1) and H(4) atoms is 1.7 Å].

The so called "bent" conformation of Hermans et al.[73] is more realistic. This conformation can be obtained from the straight conformation by the rotation of the monomer units about the C(1)–O(1) and O(4)–C(4) bonds by angles of approximately 25°. In the bent conformation there are no shortened non-valence contacts and the distance between the O(3) and C(5′) atoms in successive units is close to 2.7 Å, which indicates the possibility of the formation of the O(3)H...O(5′) intramolecular hydrogen bond. We may note that in crystalline cellobiose[51-53] the relative positions of the glucopyranose rings correspond precisely to the "bent" conformation with its characteristic O(3)H...O(5′) hydrogen bond.

In the structural studies mentioned above,[58,59] the optimum conformation of the β-D-(1 → 4)-linked residues is also close to the bent conformation. It has been shown that reasonable values of δ (which do not lead to short non-valence contacts on either side of the oxygen bridge) are in a narrow range, where there is a possibility of the formation of the O(3)...O(5′) hydrogen bond.

Conformational analysis of the β-D-glucose residues linked in different ways to give a dimeric fragment (linkages of the 1 → 2, 1 → 3, and 1 → 4 types) has been carried out[56] within the framework of the semiempirical model of molecular mechanics.[21] The potential energy of the system, including the energy of the non-valence interactions and the energy of the hydrogen bonds, was calculated as a function of two angles, φ and ψ, for the fixed angle $\beta = 117.5°$. The calculations showed that the global minimum of the potential energy of the (1 → 4)-linked dimer is close to the conformation corresponding to the 2_1 symmetry. Furthermore, it has been established that only 4% of the space of the changing variables does not lead to short covalent contacts between the atoms in successive units. This agreed qualitatively with the results of the calculations by Jones[58] and confirmed the view of Rees and co-workers[74,75] that cellulose chains are highly rigid.

In the analogous conformation calculations by Sarko and Muggli,[28] the potential energy of the system was recorded as the sum of the squares of the deviations of the geometrical parameters of the structure from standard values. Analysis of the φ–ψ map showed that the conformation of the chain with the *tg*- and *gt*-orientations of the O(6) atom, corresponding to the 2_1 conformation and a repeat unit of 10.3 Å, is close to the potential energy minimum. In the case of the *tg*-orientation, there is a possibility of the formation of the O(2)H...O(6′) bond on the other side of the oxygen bridge in addition to the O(3)H...O(5′) hydrogen bond.

Panov et al.[76] included the energies of the non-valence and electrostatic interactions, the torsional potentials, and the energy of the hydrogen bonds in the conformational energy of the dimeric cellulose chain fragment. The angle β was assumed to be 116.0°. The conformation with the *tg*-orientation of the hydroxymethyl group and the parameters φ and ψ corresponding to the 2_1 symmetry proved to be most preferred as regards energy. Two hydrogen bonds, O(3)H...O(5′) and O(2)H...O(5′), were present in the system and according to the authors only these bonds can be formed simultaneously. The conformations with hydrogen bonds of the type O(6)H...O(2′) and O(6)...O(1′) are less favourable as regards energy.

A more detailed analysis of the conformational possibilities in the cellulose macromolecule has been carried out recently.[61] The method of the variable virtual bond is used in constructing the main chain, since the search for the optimum chain conformation was restricted *a priori* by the subset of conformations satisfying the helical 2_1 symmetry. The conformational energy included the energy of the atom–atom non-valence interactions, the torsional contributions, the deformation energy of the angle β, and the potentials of the hydrogen bonds. In order to discover the regions of the permissible conformations of the main chain, a preliminary calculation was carried out in which the pitch of the helix was assumed to be constant and the angle δ was varied from −180° to 180°. It is seen from Fig. 3 that two comparatively narrow regions correspond to reasonable values of the angle β ($114 \leq \beta \leq 118°$): $44° \leq \delta \leq 58°$ and $-70° \leq \delta \leq -56°$. The same Figure presents the dependence of the O(2)...C(6′) distance on the angle δ. Evidently the range $-70° \leq \delta \leq -56°$ can be eliminated from consideration with full justification owing to the presence of unduly short O(2)...C(6′) contacts. Subsequent calculations by Pertsin et al.[61] consisted in minimising the conformational energy in the space of six changing variables. The variables included four tortional angles $\tau_i (i = 1, \ldots, 4)$, confirming the orientation of the side groups, the angle δ, and the pitch of the helix *h*. The calculations revealed the six stable conformations of the isolated cellulose helix. These conformers are designated in Figs. 4–6 by *A*1, *A*2, *B*1, *B*2, *C*1, and *C*2. Their conformational energies were −4.8, 3.1, 0.3, −0.5, −0.1, and 2.6 kcal mol^{-1} respectively.

Evidently all the stable conformers of the isolated helix are characterised by the presence of the O(3)H...O(5′) bond. As mentioned above, the presence of bonds of this kind has been confirmed by numerous experimental findings. In the conformer *A*1, which is most preferred as regards energy, there is also the O(4)...O(6)H...O(2′) hydrogen bond, which includes one donor and two acceptor oxygen atoms. The possibility of the formation of such bonds has been recently

demonstrated by a non-empirical quantum-chemical calculation[77,78] and by a neutron diffraction study of carbohydrate crystals.[79,80] (Bonds of this type will henceforth be referred to as a biacceptor in contrast to the bidonor bonds of the type OH...O...HO).

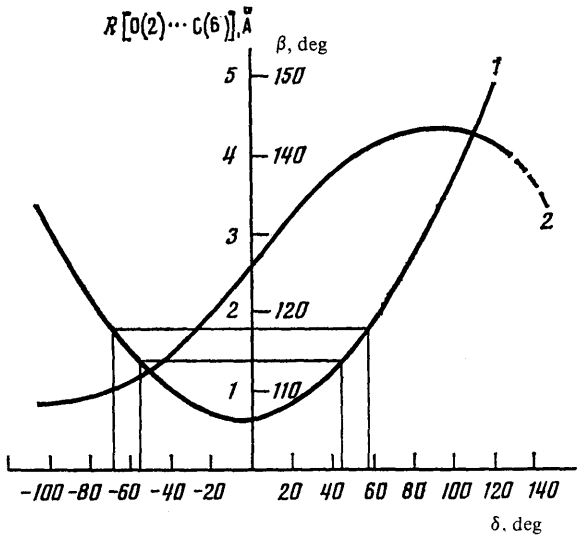


Figure 3. The permissible ranges of variation of the angles δ:[61] 1) dependence of β on δ; 2) dependence of the interatomic O(2)...C(6') distance on δ.

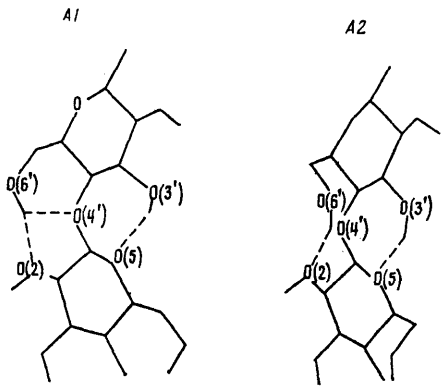


Figure 4. Type *A* conformers of the isolated cellulose helix. (Here and henceforth the hydrogen bonds are shown by dashed lines; for clarity, the ring hydrogen atoms are not shown).[61]

The hydroxymethyl group of the conformer *A*1 is close to the *tg*-position ($\tau_1 = 60°$). The rotation of this group by approximately 83° brings the hydroxymethyl group closer to the *gg*-position ($\tau_1 = -60°$) and affords the new stable conformer *A*2 (Fig.4). The bond O(6)H...O(4) is then broken and only the usual O(6)H...O(2') hydrogen bond remains. The O(6)...O(2') hydrogen bond can be formed also via the proton at the O(2) atom, which becomes the donor, and the O(6) atom—the proton acceptor. The situation described is characteristic of conformers of type *B* (Fig.5), which constitutes their principal difference from type *A* conformers. As for type *A* conformers, there is a possibility of two different positions of the hydroxymethyl group leading to the formation of the O(2)H...O(6') bond. In the conformer *B*1, the orientation of the O(6) atom is close to the *tg*-position, while in the conformer *B*2 it is close to the *gg*-position.

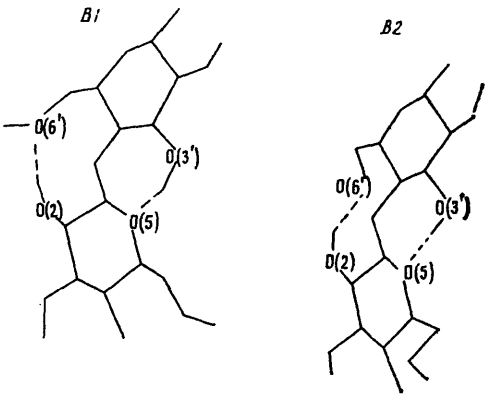


Figure 5. Type *B* conformers of the isolated cellulose helix.

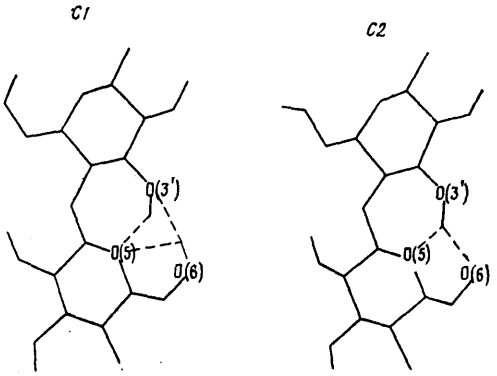


Figure 6. Type *C* conformers of the isolated cellulose helix.

It is seen from Fig.6 that there are also two stable *C*1 and *C*2 conformers whose hydroxymethyl groups are close to the *gt*-position ($\gamma_1 = 180°$). The biacceptor O(3')...O(6)H...O(5') bond and the bidonor O(6)H...O(5)...HO(3') bond are present in the conformer *C*1. In the conformer *C*2, whose energy is smaller than that of the conformer *C*1 by approximately 2.7 kcal mol^{-1}, only the biacceptor bond of the type O(5')...O(3)H...O(6') is present.

All six stable conformers of the isolated cellulose helix have an equilibrium period close to the observed repeat period (approximately 10.3 Å). The equilibrium values of the angle β are also close to those in carbohydrates and polysaccharides (115.7–117.6°). Furthermore in all the conformers there are

vacant sites for the formation of intermolecular hydrogen bonds. For this reason, each of the conformers found can in principle exist in the crystal.

The results of the conformational analysis reveal extremely extensive conformational possibilities in the cellulose macromolecule, indicating thereby the impossibility of an unambiguous prediction of its crystal structure from the consideration of an isolated chain. Evidently the packing of the macromolecules in the crystal and, in particular, the formation of intermolecular hydrogen bonds can affect significantly the relative stabilities of particular conformers.

III. STRUCTURAL MODELS OF THE MAIN POLYMORPHIC VARIETIES OF CELLULOSE

In describing the crystal structure of cellulose, the following packing parameters are usually employed: two angles of rotation of the chains about their bonds, the shift of the chains relative to one another, and two parameters specifying the direction of the chains. Here and henceforth we shall adhere to the following notation: the angles of rotation ϕ^k of the angular ($k = 1$) and central ($k = 2$) chains, the shift S of the O(4) atom of the central chain relative to the O(4) atom of the angular chain along the c axis (in fractions of the period c), and the directional parameter $p^k = \pm 1$ ($p^k = \pm 1$ if $Z_{O(5)} \gtrless Z_{C(5)}$ for the kth chain).

1. Cellulose I

As already mentioned, natural celluloses of different origin, for example, from cotton, flax, hemp, ramie, wood, etc., belong to this polymorphic modification.

The first X-ray diffraction studies on flax and hemp celluloses were carried out ten years before it was established that cellulose is a glucose polymer. At the end of the 1920's and 1930's, the familiar structural studies by Meyer and Mark,[71] Meyer and Misch,[72] Pierce,[83,84] etc. were published. These have played a decisive role in the development of ideas about the structure of crystalline regions of cellulose. Among the early structural investigations, we may also include the studies by Frey-Wyssling,[85] Marrinan and Maan,[86,87] Liang and Marchessault,[88,89] etc. carried out during the period between the 1940's and 1960's.

At the present time most of the early studies are of purely historical interest, because the structural models proposed in them conflict with later experimental data.

Thus it was assumed in Pierce's model[83,84] that the glucopyranose ring is in a planar conformation, which conflicts with experimental data considered in Section II. In Frey-Wyssling's model,[85] there is a free hydroxyl at C(2), which conflicts with IR spectroscopic data,[9,15,45,90] according to which all the hydroxy-groups of cellulose are involved in hydrogen bonds. Many early structural models, in particular the well known model of Meyer and Misch,[72] to not agree with data for the IR dichroism of the stretching vibrations of the OH groups. According to these data, in the region of such stretching vibrations[9,15,90] there are five bands, of which two exhibit perpendicular dichroism while the remaining exhibit parallel dichroism. This shows that there should be at least two OH groups in solid cellulose with preferential perpendicular orientation relative to the axis of the fibre and at least three OH groups with preferred parallel orientation. In Meyer and Misch's model mentioned above,[72] there are only intermolecular hydrogen bonds with preferential perpendicular orientation, i.e. the existence of three bands with parallel dichroism could not be accounted for in terms of this model.

Of the structural models belonging to the end of the 1960's, mention should be made of those of Maan and Marrinan[86] and Liang and Marchessault.[88] Maan and Marrinan's models were based mainly on the IR dichroism data mentioned above. X-Ray diffraction data were scarcely taken into account in these models (with the exception of the unit cell parameters). In the analysis of the structure of cellulose, Liang and Marchessault[88,89] combined spectroscopic data with the intensities of well resolved reflections. All the models proposed by the authors agreed satisfactorily only with the IR dichroism of the OH groups but conflicted with X-ray diffraction data.[58,59]

The first serious attempt to reconcile the stereochemical, IR spectroscopic, and X-ray diffraction data was undertaken by Jones.[58,59] Seven alternative CI models, differing in the relative polarities of the chains, the rotation of the monomer unit about the virtual O(4)...O(4′) bond, and the orientation of the hydroxymethyl group and the chains in the unit cell were considered. All the models considered satisfied the stereochemical criteria and agreed with IR spectroscopic data. The best agreement could be observed in X-ray reflection intensities (among the models examined) obtained for the antiparallel model, whose parameters are close to those of the modern structural models of CI.

The Jones model includes two intramolecular bonds, O(3)...O(5′) and O(2)...O(6′) [The orientation of the O(6) atom is close to *tg*] and the intermolecular O(6)...O(3″) bond located in the (200) plane. The chains are uncoiled around their axes in such a way that the planes of the glucopyranose rings are arranged almost parallel to the *ac* plane, forming a layered structure. The O(6)...O(3″) bond forms translationally identical chains, i.e. is located within the layer. The main difference between the Jones model described and the earlier structural models is the absence of interlamellar hydrogen bonds (between the angular and central chains).

As regards the structural model proposed by Jones for CI,[58] mention should be made that it has not been optimised in the sense that it does not correspond to the optimum of any specific criterion in the space of the changing variables. In essence, the Jones model is only the best among the seven discrete models examined by the author.

The search for the structural models of cellulose was first reduced to the optimisation problem in a study by Gardner and Blackwell.[91] The crystallographic divergence factor (R factor) was adopted as the target optimisation function:

$$R = \sum_{m=1}^{M} \Delta F_m / \sum_{m=1}^{M} |F_{m(\mathrm{obs})}| \qquad (1)$$

where $\Delta F_m = |F_{m(\mathrm{obs})} - F_{m(\mathrm{calc.})}|$ and the F_m are structural amplitudes. The summation in Eqn.(1) is carried out with respect to M observed reflections. The calculated structural amplitudes were obtained from the formula

$$F_{m(\mathrm{calc})} = K\left\{\sum_{hkl} [F^{hkl}_{m(\mathrm{calc})} \exp(-B_T d^2_m/4)]^2\right\}^{1/2} \qquad (2)$$

where K is the scaling factor, B_T the average isotropic temperature factor, and d_m the interplanar distance in the inverse lattice. The summation in Eqn.(2) is carried out with respect to all the hkl planes contributing to the mth reflection.

By comparing alternative models, Gardner and Blackwell[91] also calculated weighted R factors:

$$R' = \sum_{m=1}^{M} w_m^{1/2} \Delta F_m / \sum_{m=1}^{M} w_m^{1/2} |F_{m(\mathrm{obs})}| \qquad (3)$$

$$R'' = \left\{\sum_{m=1}^{M} w_m \Delta F^2_m / \sum_{m=1}^{M} w_m F^2_{m(\mathrm{obs})}\right\}^{1/2} \qquad (4)$$

where w_m are the weighting factors:

$$\Delta F_m^2 = |F_{m\,(obs)}^2 - F_{m\,(calc)}^2|.$$

In contrast to Eqn.(2), the summation in Eqns.(3) and (4) is carried out not only with respect to the observed reflections but also with respect to the postulated refelctions. The postulated reflections are included in the calculation of the R factor in order to avoid the appearance of intense theoretical reflections for planes which are not observed experimentally. The experimental set of values of $|F_{m(obs)}|$ was obtained on fibres from *Vallonia ventricosa*. The meridional 00*l* reflections were not included in the calculation of the R factor because of certain difficulties associated with the measurement of their intensities and the introduction of the Lorentz and polarisation corrections.

In the search for the most preferred structure of CI, the following initial trial models were employed;[91]

(1) P_1—parallel chains ($p^1 = p^2 = 1$), $S = {}^1/_4$;
(2) P_2—parallel chains ($p^1 = p^2 = -1$), $S = {}^1/_4$;
(3) a_1—antiparallel chains ($p^1 = 1$, $p^2 = -1$), $S = -{}^1/_4$;
(4) a_2—antiparallel chains, $S = {}^1/_4$.

In constructing the cellulose helix, the bond lengths and valence angles were set equal to their standard values.[54] The conformation of the main chain was assumed beforehand to be that corresponding to the helical symmetry 2_1 and the repeat period 10.38 Å and was not varied in the course of the calculations. The parameters of the junction of the monomer units were assumed to be equal to the values for which the O(3)...O(5′) hydrogen bond may be formed between the O(3) and O(5′) atoms ($\phi = 23°$, $\psi = -27.7°$, $\beta = 144.8°$).[56]

The optimum model was sought in the space of six changing variables.[91] These included the angle τ_1, describing the rotation of the hydroxymethyl groups (in both chains simultaneously), the three packing parameters ϕ^1, ϕ^2, and S, and also the parameters B_T and K which occur in the formulae for the calculation of the structural amplitudes [Eqns.(2)].

The model obtained from the P_1 approximation ($R = 0.254$, $R' = 0.233$, $R'' = 0.215$, and $B_T = 2.5$ Å^2) was found to be best. The shift in the optimised P_1 model was $S = 0.266$. In these models the intermolecular O(6)...O(3″) hydrogen bond is located in the (020) plane. In addition to the O(3)...O(5′) intramolecular bond, whose presence is predetermined by the initial chain conformation, the O(2)H...O(6′) bonds arises on the other side of the oxygen bridge. This bond is formed when the O(6) atom is located in a position differing from *tg* by 20°. The values of τ_1 corresponding exactly to *tg*, *gt*, and *gg* are found to be stereochemically unsuitable in the crystal and give rise to high values of the R'' factor (0.24, 0.29, and 0.35 respectively).[91]

The assumption by Gardner and Blackwell of the absolute rigidity of the main chain of the macromolecule can apparently be regarded as the main disadvantage of their treatment,[91] because the monomer units of cellulose exhibit a definite freedom of rotation about the virtual bonds. The neglect of this rotation can significantly impair the "packing possibilities" of the cellulose chains. In particular, the packing of rigid chains can lead to shortened contacts, which might have been avoided by means of small changes in the conformation of the main chain.

Among the disadvantages of the treatment of Gardner and Blackwell,[91] one should include also the fact that, in selecting the initial approximations, the authors, considered only two discrete values of the parameter $S(\pm{}^1/_4)$, leaving uninvestigated most of the range of variation of this parameter. Here it is appropriate to mention that the values $S = \pm{}^1/_4$ arise from the intensity of the 002 and 004 reflections only if we assume that the scattering centres of the cellulose chain are concentrated at the centres of the glucopyranose rings. The above approximation is fairly rough and, as will be shown below, the observed intensity ratios of the four meridional reflections can be obtained for values of S extremely remote from $\pm{}^1/_4$.

A more complex method for the refinement and selection of the optimum cellulose model was adopted by Sarko and co-workers.[28,93] The optimum model was determined by minimising a target function Φ of the following type:

$$\Phi = 100fR'' + (1-f)Y \qquad (5)$$

where $f = 0.8$ is the weighting factor and Y is the sum of the standard deviation functions taking into account the deviation of the geometrical parameters of the model from their standard values.[93] In the calculation of the R'' factor, the authors used in Eqn.(5) the X-ray diffraction data obtained using ramie fibres.

In order to test the hypothesis that the 2_1 axis is present in the crystal structure of cellulose, the authors[28,93] eliminated the limitations associated with the $P2_1$ symmetry and varied all the geometrical parameters of the model with the exception of their bond lengths (23 parameters in all). Apart from the geometrical parameters, the average temperature multiplier was also varied, being assumed to be anisotropic in the given instance and having three dependent components.

Two structural models were found to be most suitable as regards the value of Φ. These are a model with parallel packing of the chains ($R'' = 0.193$) and the antiparallel model ($R'' = 0.246$). In both cases the deviations from the symmetry of the 2_1 screw axis were found to be insignificant and did not exceed 1–2° for the valence and torsional angles in the ring and in the junction of the monomer units. In order to discover the best models, the authors resorted to Hamilton's statistical tests.[92] If a one-dimensional hypothesis is assumed (i.e. that either the parallel or the antiparallel model is possible), then the differences in the R' factor for the fifty observed reflections supporting the parallel model are found to be significant at a level of significance of 99.5%.

In terms of its structural parameters and especially the system of hydrogen bonds, the model proposed by Sarko and co-workers[93] proved to be close to the P_1 model of Gardner and Blackwell.[91] In assessing the method used by Sarko and co-workers[28,35,60,62,93] for the elucidation of the cellulose structure, it is necessary to note the following points. Firstly, the inclusion in the optimisation of "rigid" variables such as the valence angles and bond lengths has little justification.[35,62] The variation of these variables usually sharply impairs the convergence of the optimisation process, so that the target function becomes sensitive mainly to the "rigid" variables. The search for the optimum values of the "sensitive" parameters (torsional angles, the rotations of the chains, etc.) becomes greatly hindered owing to the oscillations in the "rigid" variables about their equilibrium values.

Yet another disadvantage of the method of Sarko and co-workers[93] is the absence, in an explicit form, of terms describing hydrogen bonds in the target function (5). For this reason, the endeavour by the system to achieve the maximum saturation by hydrogen bonds is not reflected in the optimisation of the target function.

Finally, it is essential to note also the serious error which Woodcock and Sarko[93] made in applying Hamilton's tests to the comparison of the best parallel (*P*) and antiparallel (*a*) models. As already stated, in determining the significance of the differences in the R'' factors between the *a* and *P* models, the authors assumed a one-dimensional hypothesis. This would evidently be correct if the *a* and *P* models differed only by the relative chain polarities. In reality the *a* and *P* models differ also by all the remaining parameters, whose number (23) actually determines the dimensions of the hypothesis. If

the dimensions of the hypothesis are assumed to be 23, then the a and P models are found to be statistically indistinguishable whatever the level of significance.

The insolubility of the problem of the relative chain polarities in CI was yet again demonstrated in French's study.[94] This was the best founded and interesting crystallographic investigation of CI based on thoroughly analysed and corrected experimental data (obtained from ramie cellulose) and a well balanced computational model.

As regards the experimental part of the study, the author was able to determine for the first time the intensities of the meridional 002 and 004 reflections, which proved to be extremely significant in finding the orientation of the hydroxymethyl group. From the standpoint of the selection of the changing variables, French's computational model occupies an intermediate position between the model of Gardner and Blackwell and the model of Sarko and co-workers. On the one hand, French's model is free from the defects of Gardner and Blackwell's model, since it assumes the deformation of the main macromolecular chain (as a result of the rotation of the monomer units about the virtual bond). On the other hand, in contrast to the model of Sarko and co-workers, in French's model there are no "rigid" parameters, since all the bond lengths and valence angles are assumed to be constant in the course of the solution of the structural problems (with the exception of the valence angle at the glucosidic oxygen atom).

Altogether French's computational model included nine changing variables, among which were the angles $\tau_1{}^k$, ϕ^k, and δ^k as well as the parameters p^k and S described previously. Initially French[94] assumed different values of $\tau_1{}^k$ and δ^k in the angular and central chains and also in two successive monomer residues, but this proved to be superfluous; neither the abandonment of the 2_1 symmetry nor the introduction of conformational differences between the angular and central chains entailed an appreciable improvement of the R'' factor.

The antiparallel model ($R'' = 0.158$), designated by the author[94] as the a_1 model, proved to be most suitable as regards the R'' factor. In this model, the macromolecules are linked into layers by intermolecular O(6)...O(3'') hydrogen bonds. Apart from the O(3)...O(5') intramolecular bond, when the position of the O(6) atom is close to tg, there is a possibility of the formation of two hydrogen bonds of the type O(2)...O(6') and O(6)...O(4). The existence of the latter bond is extremely doubtful according to French,[94] since it differs markedly from the linear bond and in addition the bridging O(4) oxygen atom is a weak proton acceptor.

The best "parallel" models found by French (the P_1 and P_2 models) differed very insignificantly from the a_1 model as regards the value of the R'' factor ($R''P_1 = 0.185$ and $R''P_2 = 0.175$). This made it impossible for the author[94] to reach the unambiguous conclusion that the packing of the chains in CI is antiparallel. Application of Hamilton's test[92] shows that the hypothesis of the parallel disposition of the chains in CI can be rejected only for an extremely low level of significance (95.0%).

The difficulties arising in the interpretation of the diffraction data for cellulose made it necessary to resort to non-crystallographic criteria, in particular to the energy criterion. Thus Nugmanov and co-workers[95,96] sought the structure of CI by optimising a target function of the type

$$\Phi = U + \omega R'' \qquad (6)$$

where U is the potential energy of the system and ω the weighting factor chosen to ensure that the maximum statistically insignificant changes in U and R'' make identical contributions to Φ ($\omega = 15$ kcal mol^{-1}). The potential U was calculated by the method of atom–atom potentials;[20,96,97,98] it included both intramolecular and intermolecular components. The optimum structure was sought in the space of 16 variable parameters: $\tau_i{}^k$, δ^k, ϕ^k, S, p^k, and B_T.

A considerable number of local minima in the target function were discovered in the course of the calculations.[95] The characteristics of the deepest of these are presented in Table 1. The fact that the values of the R'' factor for the best structural model lie within the comparatively narrow range 0.24–0.16 is striking when Table 1 is analysed. Application of Hamilton's statistical test[92] shows that, for the usual 99% level of significance, the vast majority of the models considered (with the exception of the a_3 and a_4 models) are statistically equivalent. This illustrates clearly yet again the ambiguity of the solution of the structural problem for CI on the basis of the available diffraction data even when stereochemical packing limitations are introduced.

Table 1. Certain characteristics of the most probably structural models of CI.[95,96]

Model*	R''	U, kcal mol^{-1}	Φ, kcal mol^{-1}	Model*	R''	U, kcal mol^{-1}	Φ, kcal mol^{-1}
$\bar{P}_1$	0.165	−19.9	−17.4	a_1	0.213	19.7	−16.5
$\bar{P}_2$	0.187	−19.7	−16.9	a_2	0 155	−18.6	−16.1
$\bar{P}_3$	0.182	−17.9	−15.2	a_3	0.240	−19.7	−16.0
$\bar{P}_4$	0.167	−17.5	−14.6	a_4	0.240	−19.7	−16.0

*The parallel models, designated by P_i and $\bar{P}_i$, correspond to the "average" polarity $p = (p^1 + p^2)/2 = \pm 1$, while the antiparallel a_i models correspond to the polarity $p = 0$.

It is significant that, in the course of the search for a global minimum in the target function Φ,[95] many structural models proposed previously for CI were discovered. Thus the models designated in Table 1 by P_1 and a_3 respectively are analogues of the Gardner–Blackwell–Sarko[64,93] and French[94] models just considered above. Here the analogy is manifested both by the similarity of the geometrical parameters of the models and by the complete identity of the hydrogen bond systems. Evidently the P_1 and a_2 models are not best as regards either the energetic standpoint or as regards the value of the R'' factor.

The global minimum in the target function was found within the framework of a parallel arrangement of the chains whose direction was opposite to that in the P_1 model (the $\bar{P}_1$ minimum in Table 1). The best "antiparallel" model a_1 was found to have similar energy to the $\bar{P}_1$ model and was statistically equivalent to the latter as regards the R'' factor. Projections of the P_1 and a_1 models onto the a^*c plane are illustrated in Figs. 7 and 8.

Despite the differences between the directions and polarities of the chains in the P_1, $\bar{P}_1$, and a_1 models, the chain conformations and the hydrogen bond systems are virtually similar in all three models. In each of the symmetrically independent layers, the chains are in a conformation close to $B1$ with the system of intramolecular hydrogen bonds O(3)H...O(5') and O(2)H...O(6') characteristic of $B1$. The macromolecules within the layer are linked via the intermolecular O(6)H...O(3'') hydrogen bonds, while the interlamellar hydrogen bonds are absent from the system.

An interesting feature of the $\bar{P}_1$ model is its ability to be converted into the next model ($\bar{P}_2$) as a result of the jump of the proton at the O(6) atom in the central chain, while the positions of all the remaining atoms remain virtually unchanged. The $\bar{P}_1 \rightarrow \bar{P}_2$ transition is accompanied by the rupture of the O(6)H...O(3″) intermolecular bond and the formation of the O(4″)...O(6)H...O(5″) interlamellar biacceptor bond. The similarity of the $\bar{P}_1$ and $\bar{P}_2$ models as regards structure and energy and also the small barrier to the transition make it possible to regard these two models as a single structural model with a labile system of hydrogen bonds.

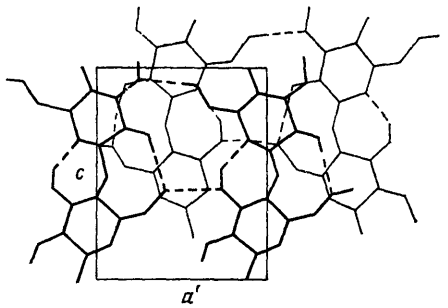


Figure 7. The $\bar{P}_1$ structural model corresponding to the global minimum in the target function Φ [Eqn.(7)] within the framework of the parallel version of the packing of the chains.[95,96]

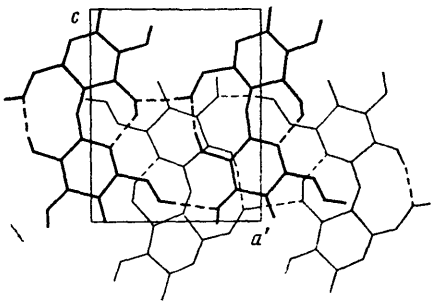


Figure 8. The a_1 structural model for CI corresponding to the global minimum in the target function Φ [Eqn.(7)] within the framework of the antiparallel version of the packing of the chains.[95,96]

A small energy barrier (<1 kcal mol^{-1}) is observed also for the $a_1 \rightarrow a_3$ and $a_1 \rightarrow a_4$ transitions. However, in these cases the probability of the transition is very insignificant owing to the large differences in the relative shift of the layers (~0.7 Å) between the models indicated. Clearly, such a shift requires the concerted collective displacement of the atoms within the limits of the target layer, which is an extremely improbable event. (Formally this implies that the pre-exponential factor in the expressions for the probabilities of the $a_1 \rightarrow a_4$ transitions is very small).

2. Cellulose II

The CII structural modification is formed when cellulose is regenerated from its derivatives or on reprecipitation from solutions [regenerated cellulose (RC)] and also when cellulose preparations are treated with concentrated solutions of alkali metal hydroxides [mercerised cellulose (MC)].[9-11] The regenerated and mercerised celluloses are characterised by virtually identical X-ray diffraction patterns and IR spectra and on this basis they are regarded as a single polymorphic modification.

Whereas well-resolved X-ray diffraction patterns typical of CII can be obtained on mercerising cotton, ramie, and wood cellulose, the attempts to obtain CII by mercerising the *Valonia ventricosa* fibres were not successful.[28,89,94] This confirms yet again the above hypothesis that cellulose from seaweed apparently possesses an individual crystal structure different from that of ramie, cotton, and wood cellulose.

It is important to note that the polymorphic transition from CI to CII is monotropic. The claim that CII is the more stable polymorphic modification of cellulose is based on precisely this experimental fact.

Analysis of the equatorial reflections in the X-ray diffraction patterns of CII shows that the CII unit cell is analogous to that in CI. One macromolecule intersects the *ab* face at the beginning of the crystallographic coordinate system, while the other makes an intersection of the centre of the face. It follows from the analysis of the intensities of the (110) and (020) reflections that the average plane of the glucopyranose rings in CII should be rotated by a somewhat greater angle relative to the *ac* plane than in CI.

The polymorphic transition from CI to CII is accompanied by significant changes in the IR spectra. Thus, whereas three bands with parallel dichroism (at 3275, 3350, and 3375 cm^{-1}) and two bands with perpendicular dichroism (at 3305 and 3405 cm^{-1}) are observed in the CI spectrum in the region corresponding to the vibrations of the OH groups, in the CII spectrum two bands with parallel dichroism (at 3407 and 3488 cm^{-1}) and three bands with perpendicular dichroism (at 3175, 3305, and 3350 cm^{-1}) are observed.[9,90] The above differences (both in the stretching vibration frequency of the OH groups and in the dichroism) indicate a significant rearrangement of the hydrogen bond system on passing from CI to CII. X-Ray diffraction patterns of CII were first obtained and indexed by Andress[99] within the framework of a monoclinic unit cell for the $P2_1$ space group. The unit cell parameters were $a = 8.14$ Å, $b = 9.19$ Å, $c = 10.3$ Å (fibre axis), and $\gamma = 118°$. Thus CII has lattice periods similar to those in CI but a significantly different monoclinic angle γ.

Crystallographic studies on CII were carried out simultaneously with such studies on CI using the same method and frequently the same workers were involved.[72,84,87,89,100,101] This obviates the need for .... structural CII models with parallel and antiparallel .... the CII models proper.†

As for CI, it is best to begin a detailed examination of structural models with the model put forward by Jones,[58,59] who examined structural CII models with parallel and antiparallel packing of the chains. A satisfactory agreement with IR spectroscopic and X-ray diffraction data was achieved for the model with the opposite direction of the chains and the shift $S = 0.29$. The positions of the O(6) atom in the angular and central chains were close to the *tg*-conformation

†Parts of the Russian text appear to be missing (Translator).

and *gt*-conformation respectively. The angular chain was stabilised by the two intramolecular hydrogen bonds O(3)H...O(5') and O(2)...O(6') and the single central bond O(3)H...O(5'). The intermolecular bonds were in the (110) plane. These are two symmetrically non-equivalent hydrogen bonds of the O(6)...O(3") type and the O(2)...O(2") bond.

In the analysis of the structure of CII, Kolpak and Blackwell[63,102] used the approach proposed by Gardner and Blackwell[64] for the elucidation of the structure of CI. The same four trial models P_1, P_2, a_1, and a_2 were used as the initial approximation as in the analysis of the packing of CI. Both regenerated and mercerised celluloses were examined.

The optimised a_2 model, which is also free from short contacts, was found to be most suitable as regards the R'' factor. Application of Hamilton's test showed that, in terms of the R'' factor, the models obtained from a_1, P_1, and P_2 can be rejected for a level of significance higher than 99%. Comparison of the a_2 models optimised with respect to X-ray diffraction data for RC on the one hand, and MC, on the other showed that they are fairly similar but not identical. The main difference is in the value of the parameter S, which is 0.216 for RC and 0.227 for MC. Significant differences are observed also in the optimised values of the average isotropic temperature factor B_T, which are 20 Å^2 for RC and 32 Å^2 for MC.

The system of hydrogen bonds in the a_2 model resembles the system of such bonds in the Jones model.[59] The O(6)...O(3") bond is located in the plane of the central chains, while the O(6)...O(2") bond is in the plane of the angular chains. The O(2)...O(2") bond is located in the (110) plane. According to the authors,[63,102] the presence of the last bond is in fact the main and fundamental difference between the structures of CI and CII and is the main cause of the increased stability of the CII polymorphic modification.

The structure of CII was also examined by Stipanovic and Sarko.[35] The optimum structural model of CII was selected from the condition for the minimum in the target function (5) with all the geometrical variables, including the bond lengths and the valence and torsional angles, varied. The average isotropic temperature multiplier was likewise varied. The model with the antiparallel disposition of the chains, which is close in terms of its parameters to the structural models of RC and MC proposed by Kolpak and co-workers,[63,102] was found to be most suitable. However, according to the author's calculations,[35] apart from the O(2)...O(2") bond in the (110) plane there is also a possibility of the formation of the interlamellar O(6)...O(3") bond.

A structural model of CII with a parallel arrangement of the chains has been proposed recently.[103] The unit cell in this model is made up of the diagonals of the Wellard cell and includes one symmetrically independent chain. This study[103] was based on the existing definite similarities between the X-ray diffraction patterns and IR spectra of CII and cellotetraose, whose unit cell contains one crystallographically non-equivalent molecule.[104] However, the above model[103] had an inadmissibly high value of the R factor ($R = 0.50$) and in addition conflicted with data for the IR dichroism of the stretching vibrations of the OH groups. This attempt demonstrates perhaps the error of the "mechanical" transfer of data obtained solely on the basis of a single model compound (cellotetraose) to the structure of a complex polymer such as cellulose.

The energy criteria have been resorted to in order to determine the structure of CII.[98,105,106] The optimum structural model was sought in a manner fully analogous to the method described above in the discussion of the studies by Nugmanov and co-workers.[61,95,96] The characteristics of the most probable structural models of CII are presented in Table 2. The following factors are striking in the analysis of Table 2. Firstly, it is evident that in the case of CI the available diffraction data are too few for an unambiguous solution of the structural problem (for a 99% level of significance, all the models presented are statistically equivalent). Secondly, comparison of the data in Tables 1 and 2 shows that the optimum models of CII are more favourable on energy grounds than the optimum models of CI by approximately 1–2 kcal mol^{-1}, which agrees with the view that CII is a more stable polymorphic modification of cellulose. Finally, it is seen from Table 2 that, among the models most suitable as regards energy, there are no models with a parallel arrangement of the chains. The difference of approximately 2 kcal mol^{-1} obtained between the best parallel model (P_1) and the best antiparallel models is sufficiently convincing to reject the hypothesis of a parallel arrangement of the chains in CII.[103] Among the antiparallel models in Table 2, mention should be made of the a_6 model, which is in essence a structural analogue of the familiar model of Stipanovic and Sarko[35] and also of Kolpak and co-workers.[63,102] Evidently the a_6 model is appreciably inferior to the best antiparallel models as regards energy and the value of the target function.

Table 2. Certain characteristics of the most probable structural models of CII.[98,105,106]

Model*	R''	U, kcal mol^{-1}	Φ, cal mol^{-1}	Model*	R''	U, kcal mol^{-1}	Φ, cal mol^{-1}
a_1	0.179	−21.4	−18.7	a_5	0.186	−20.4	−17.6
a_2	0.178	−21.2	−18.5	a_6	0.124	−19.4	−17.5
a_3	0.183	−21.2	−18.5	P_1	0.167	−19.4	−16.9
a_4	0.164	−20.1	−17.6	P_1	0.164	−17.5	−15.1

*The same notation is used as in Table 1.

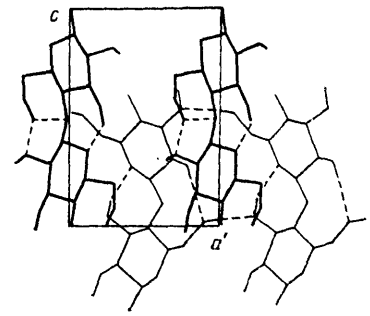
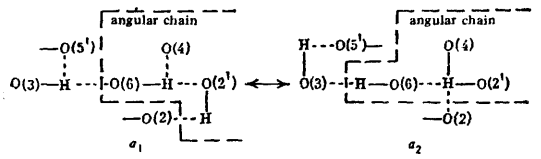


Figure 9. The optimum structural model a_1 for CII.[98,105,106]

The a_1 model, whose projection on to the $a'c$ plane is illustrated in Fig. 9, is best in terms of the value of the target function. Evidently the a_1 model represents alternating layers parallel to the ac plane. The (020) layer in the a_1 model is constructed analogously to the layers in the structural a_1 and P_1 models for CI described above in the discussion of the studies of Nugmanov and co-workers.[95,96] In the (020) layer there are the intramolecular hydrogen bonds O(3)...O(5') and O(2)H...O(6') and the intermolecular bond

O(6)H...O(3″). The conformation of the macromolecules in the (010) layer of angular chains is close to the *A*1 conformation with its typical O(4)...O(6)H...O(2) biacceptor bond. There are no intermolecular hydrogen bonds in the (010) plane. The neighbouring layers in the a_1 model are linked by intermolecular bonds: the strong O(2)H...O(2″) bond and the weak O(3)H...O(6″) bond. The proton at O(3) is then found to be involved in the O(5′)...O(3)H...O(6″) biacceptor bond.

The a_2 structural model can be obtained from a_1 by changing the orientation of the O(6)H and O(2)H hydroxyls of the angular chain and the a_3 model can be obtained by altering the orientation of the O(6)H hydroxyl in the central chain. The rearrangement of the hydrogen bond system in the transition from the a_1 model to the a_2 model is illustrated schematically below (for clarity, only the hydroxyls involved in the rearrangement are shown); the OH groups of the angular chain are differentiated by a dashed line):



The hydrogen bond system remains uninvolved in the $a_1 \rightarrow a_3$ transition and only the position of the proton at the O(6) atom changes relative to the line connecting the donor and acceptor oxygen atoms.

It is seen from Table 2 that the a_1, a_2, and a_3 structures have similar energies and in addition the barriers to the transitions between them are low (0.9 and 0.2 kcal mol^{-1} for the $a_1 \rightarrow a_2$ and $a_1 \rightarrow a_3$ transitions respectively). As for the $\bar{P}_1$ and $\bar{P}_2$ models of CI, this factor suggests that all three structures can exist simultaneously in CII and can be interconverted as a result of the thermal migration of protons.

IV. CONCLUSION

Analysis of the data presented shows that the problem of the determination of the molecular-crystal structure of cellulose is still far from its final solution. In fact only the data for the symmetry and size of the unit cell as well as the conformation of the glucopyranose residue in the main macromolecular chain can now be regarded as fully non-controversial information about the structure of cellulose. The data enumerated do not, however, allow the complete reproduction of the architecture of the crystal owing to the extensive conformational possibilities associated with the side groups and the wide variety of different versions of the disposition of the cellulose chains in the crystal arising from this. The available diffraction data are too limited to make an unambiguous choice between several alternative versions of the structure.

The semiempirical energy calculations, which apparently possess a greater "resolving capacity" than crystallographic criteria such as the *R* factor and the principle of close packing, can apparently help significantly in solving the structural problem. However, even in this case calculation yields several possible versions of the structure which have extremely similar energies.

In our view, a more well defined situation obtains for the stable CII polymorphic modification of cellulose where the most suitable model is separated from the remaining possible versions of the structure by an "energy gap" of approximately 1 kcal mol^{-1}. Bearing in mind that the error in the calculation of the lattice energy by the method of atom–atom potentials is usually systematic and is to a large extent compensated when different versions of packing of the same molecules are compared, one may expect that the ~1 kcal mol^{-1} energy difference is entirely significant. An interesting feature of the most probable model of CII is the lability of its hydrogen bond system, i.e. the capacity for ready rearrangements as a result of the thermal movement of individual protons.

In the case of CI the best models are distributed more or less uniformly with respect to energy and the best parallel and the best antiparallel models ($\bar{P}_1$ and a_1 in Table 1) are almost equivalent as regards energy and are statistically indistinguishable with respect to the *R*″ factor. An important argument in support of the antiparallel version of the structure is the finding that the fibres swell only slightly on mercerisation leading to the transition of CI into CII. An appreciable structural rearrangement (in particular in the change of the arrangement of the chains) cannot then occur, which entails a different similarity of the structures of CI and CII. Comparison of the a_1 models for CI and CII does indeed reveal a similarity of significant parameters such as the orientation of the hydroxymethyl groups and the relative shift of the chains.

REFERENCES

1. H.S.Hon, J.Polym.Sci., Polym.Chem.Ed., 1976, **14**, 2497.
2. R.D.Brown and I.L.Jurasek (Editors), "Hydrolysis of Cellulose: Mechanism of Enzymatic and Acid Catalysis", Plenum Press, Washington, 1979, p.399.
3. De.Ellefsen and B.A.Toennesen, in "Cellulose and its Derivatives" (Edited by N.Bicles and L.Segal) (Translated into Russian),# Izd.Mir, Moscow, 1974, p.154.
4. Kh.U.Usmanov and G.V.Nikonovich, "Elektronnaya Mikroskopiya Tsellyulozy" (Electron Microscopy of Cellulose), Izd.FAN, Tashkent, 1962.
5. V.A.Marikhin and L.M.Myasnikova, "Nadmolekulyarnaya Struktura Polimerov", (The Supermolecular Structure of Polymers), Izd.Khimiya, Leningrad, 1977.
6. V.I.Sharkov, Uspekhi Khim., 1967, **36**, 312 [Russ. Chem.Rev., 1967(2)].
7. I.A.Tarchevskii and G.N.Marchenko, "Biosintez i Struktura Tsellyulozy" (The Biosynthesis and Structure of Cellulose), Izd.Nauka, Moscow, 1985.
8. Kh.U.Usmanov and K.Kh.Razikov, "Atlas Morfologicheskikh Struktur Khlopka", (Atlas of the Morphological Structures of Cotton), Izd.FAN, Tashkent, 1978.
9. N.Bicles and L.Segal (Editors), "Cellulose and its Derivatives" (Translated into Russian), Izd.Mir, Moscow, 1974.
10. N.I.Klenkova, "Struktura i Reaktsionnaya Sposobnost' Tsellyulozy" (The Structure and Reactivity of Cellulose), Izd.Nauka, Leningrad, 1976.
11. R.M.Brown (Editor), "Cellulose and Other Natural Polymer Systems", Plenum Press, New York–London, 1982, p.519.

[#This may be a reference to a series of reviews on this theme appearing in High Polym., 1971, Part 4, (Ed.of Translation)].

12. O.P.Golova and V.I.Ivanov, "O Molekulyarnom Vese Tsellyulozy" (The Molecular Weight of Cellulose), Izd.Akad.Nauk SSSR, Moscow–Leningrad, 1949.
13. G.V.Schultz and M.Marx-Figini, Makromol.Chem., 1954, **14**, 1.
14. J.Honeyman (Editor), "Advances in the Chemistry of Cellulose and Starch" (Translated into Russian), Inostr.Lit., Moscow, 1962.
15. R.G.Zhbankov "Infrakrasnye Spektry Tsellyulozy i Ee Proizvodnykh" (Infrared Spectra of Cellulose and its Derivatives), Izd.Nauka i Tekhnika, Minsk, 1964, p.338.
16. E.L.Eliel, "Stereochemistry of Carbon Compounds" (Edited by V.M.Potatov) (Translated from English into Russian), Izd.Mir, Moscow, 1965.
17. N.K.Kochetkov, A.F.Bochkov, B.A.Dmitriev, A.I.Usov, O.S.Chizhov, and V.M.Shibaev, "Khimiya Uglevodov" (Carbohydrate Chemistry), Izd.Khimiya, Moscow, 1967.
18. V.P.Panov and R.G.Zhbankov, "Konformatsii Sakharov" (The Conformations of Sugars), Izd.Nauka i Tekhnika, Minsk, 1975.
19. J.F.Stoddart, "Stereochemistry of Carbohydrates" (Edited by Yu.A.Zhdanov) (Translated from English into Russian), Izd.Mir, Moscow, 1975.
20. V.G.Dashevskii, "Konformatsiya Organicheskikh Molekul" (The Conformations of Organic Molecules), Izd. Khimiya, Moscow, 1974.
21. V.G.Dashevskii, "Konformatsionnyi Analiz Organicheskikh Molekul" (Conformational Analysis of Organic Molecules), Izd.Khimiya, Moscow, 1982.
22. N.A.Martynov and K.A.Vylegzhanina, "Rentgenografiya Polimerov" (X-Ray Diffraction of Polymers), Izd.Khimiya, Leningrad, 1972.
23. F.H.Jael, "Polymeric Single Crystals" (Translated into Russian), Izd.Khimiya, Leningrad, 1968.
24. B.Wunderlich, "The Physics of Macromolecules" (Edited by Yu.K.Godovskii and V.S.Popkov) (Translated from English into Russian), Izd.Mir, Moscow, 1976.
25.§ D.W.Jones, in "Cellulose and Its Derivatives" (Edited by N.Bicles and L.Segal), (Translated into Russian), Izd.Mir, Moscow, 1974, p.119.
26. H.G.Wellard, Nature, 1954, **174**, 471.
27. H.G.Wellard, J.Polym.Sci., 1954, **13**, 471.
28. A.Sarko and R.Muggli, Macromolecules, 1974, **7**, 486.
29. G.Honjo and M.Watanabe, Nature, 1958, **181**, 326.
30. D.G.Fisher and J.Maan, J.Polym.Sci., 1960, **42**, 189.
31. G.Legrant, Acta Cryst., 1952, **5**, 800.
32. R.Norman, Textile Res.J., 1963, **33**, 711.
33. A.I.Kitaigorodskii and D.Ya.Tsvankin, Vysokomol. Soed., 1959, **1**, 269.
34. A.I.Kitaigorodskii and D.Ya.Tsvankin, Vysokomol. Soed., 1959, **1**, 279.
35. A.J.Stipanovic and A.Sarko, Macromolecules, 1976, **9**, 851.
36. R.H.Atalla and J.C.Gast, Amer.Chem.Soc., Adv. Chem.Ser., 1979, **55**, 3249.
37. R.H.Atalla, L.E.William, and D.L.Van der Hart, Amer. Chem.Soc., Adv.Chem.Ser., 1979, **55**, 3251.
38.¶ J.Blackwell and R.H.Marchessault, in "Cellulose and Its Derivatives" (Edited by N.Bicles and L.Segal) (Translated into Russian), Izd.Mir, Moscow, 1974, p.9.
39. I.Nieduszynski and E.D.T.Atkins, Biochim.Biophys. Acta, 1970, **222**, 109.
40. J.Blackwell, F.J.Kolpak, and H.Gardner, Tappi, 1978, **61**(1), 71.
41. A.Sarko, Tappi, 1978, **61**(2), 59.
42. W.G.Terrier, Acta Cryst., 1963, **16**, 1023.
43. S.C.Chy and G.A.Jeffrey, Acta Cryst., 1967, **23**, 1038.
44. V.G.Dashevskii, in "Organicheskiya Khimiya.Itogi Nauki i Tekhniki" (Organic Chemistry. Advances in Science and Engineering), Izd.VINITI, Moscow, 1975, Vol.1, p.99.
45. V.P.Panov, V.N.Gritsan, and R.G.Zhbankov, Dokl. Akad.Nauk Beloruss.SSR, 1977, **21**, 236.
46. V.N.Gritsan, R.G.Zbankov, and V.G.Kacur, Acta Polym., 1982, **33**(1), 20.
47. R.G.Zhbankov and V.P.Kozlov, "Fizika Tsellyulozy i Ee Proizvodnykh" (The Physics of Cellulose and Its Derivatives), Izd.Nauka i Tekhnika, Minsk, 1983.
48. R.E.Reeves, J.Amer.Chem.Soc., 1949, **71**, 212.
49. R.E.Reeves, J.Amer.Chem.Soc., 1949, **71**, 215.
50. R.E.Reeves, J.Amer.Chem.Soc., 1951, **73**, 957.
51. S.C.Chy and G.A.Jeffrey, Acta Cryst., 1968, **B24**, 830.
52. R.A.Jecobson, J.A.Wunderjich, and W.N.Lipscomb, Acta Cryst., 1961, **14**, 598.
53. C.J.Brown, J.Chem.Soc., A, 1966, 927.
54. S.Arnott and W.E.Scott, J.Chem.Soc., Perkin Trans. II, 1972, **3**, 324.
55. P.R.Sunderarajan and V.S.R.Rao, Biopolymers, 1969, **8**, 305.
56. B.K.Sathyanarayana and V.S.R.Rao, Biopolymers, 1971, **10**, 1605.
57. A.V.Nikitin, R.M.Myasnikova, V.I.Andrianov, V.N.Molchanov, S.I.Firgang, and A.I.Pertsin, Kristallografiya, 1984, **29**, 489.
58. D.W.Jones, J.Polym.Sci., 1958, **32**, 124.
59. D.W.Jones, J.Polym.Sci., 1958, **43**, 173.
60. P.Zygenmayer and A.Sarko, Amer.Chem.Soc., Symp. Ser., 1980, **141**, 226.
61. A.I.Pertsin, O.K.Nugmanov, V.F.Sopin, G.N.Marchenko, and A.I.Kitaigorodskii, Vysokomol.Soed., 1981, **A23**, 2146.
62. A.Sarko, J.Soutwick, and L.Hayashi, Macromolecules, 1976, **9**, 857.
63. F.J.Kolpak, M.Weih, and J.Blackwell, Polymer, 1978, **19**(2), 123.
64. K.H.Gardner and J.Blackwell, Biopolymers, 1974, **13**, 1975.
65. M.Sundaralingam, Biopolymers, 1958, **6**, 189.
66. P.Zugenmayer and A.Sarko, Acta Cryst., 1968, **B28**, 3158.
67. K.Kildeby, S.Melberg, and K.Rasmussen, Acta Chem.Scand., 1977, **A31**, 1.
68. S.Melberg and K.Rasmussen, Carbohydr.Res., 1979, **71**, 25.
69. B.J.Poppleton and A.M.Mathieson, Nature, 1968, **218**, 1046.

[§ This may be a reference to a series of reviews on this theme appearing in High Polym., 1971, Part 4, (Ed. of Translation)].

[¶ This may be a reference to a series of reviews on this theme appearing in High Polym., 1971, Part 4, (Ed. of Translation)].

70. B.J.Poppleton and B.M.Gatehouse, J.Polym.Sci., 1972, **375**, 1.
71. K.H.Meyer and H.Mark, Ber.Deutsch.Chem.Ges., 1928, **61**, 593.
72. K.H.Meyer and L.Misch, Helv.Chim.Acta, 1957, **20**, 232.
73. P.H.Hermans, J.Boys, and C.Maan, Kolloid, 1943, **2**, 169.
74. D.A.Rees and J.Skerret, Carbohydr.Res., 1968, **7**, 534.
75. D.A.Rees and W.E.Scott, Chem.Comm., 1969, 1.
76. V.P.Panov, V.N.Gritsan, and R.G.Zhbankov, Zhur. Prikl.Spektrosk., 1979, 293.
77. M.D.Newton, G.A.Jeffrey, and S.Takagi, J.Amer. Chem.Soc., 1979, **101**, 1997.
78. M.D.Newton, Acta Cryst., 1983, **39**, 104.
79. G.A.Jeffrey and H.Maluszynska, Int.J.Quantum Chem., Quantum Biol.Symp., 1981, **8**, 231.
80. G.A.Jeffrey and J.Mitra, Acta Cryst., 1983, **39**, 469.
81. S.Nishikawa and S.Ono, Proc.Nath.Phys.Soc., 1913, **7**, 131.
82. O.L.Sponsler and W.H.Dore, Cellulose Chemie, 1930, **11**, 186.
83. F.T.Pierce, Trans.Faraday Soc., 1946, **42**, 545.
84. F.T.Pierce, Trans.Faraday Soc., 1946, **42**, 560.
85. A.Frey-Wyssling, Biochem.Biophys.Acta, 1955, **18**, 166.
86. H.J.Marrinan and J.Maan, J.Polym.Sci., 1956, **21**, 301.
87. J.Maan and H.J.Marrinan, J.Polym.Sci., 1958, **32**, 125.
88. C.Y.Liang and R.H.Marchessault, J.Polym.Sci., 1959, **37**, 385.
89. R.H.Marchessault and C.Y.Liang, J.Polym.Sci., 1960, **43**, 71.
90. J.Dechant, R.Danz, W.Kimmer, and R.Schmolki, "Infrared Spectroscopic Studies of Polymers", (Translated into Russian), Izd.Khimiya, Moscow, 1976.
91. K.H.Gardner and J.Blackwell, Biopolymers, 1974, **13**, 1975.
92. W.G.Hamilton, Acta Cryst., 1965, **18**, 502.
93. C.Woodcock and A.Sarko, Macromolecules, 1980, **13**, 1183.
94. A.D.French, Carbohydr.Res., 1978, **61**, 67.
95. O.K.Nugmanov, A.I.Pertsin, and G.N.Marchenko, Dokl.Akad.Nauk SSSR, 1984, **277**, 108.
96. A.J.Pertsin, O.K.Nugmanov, and G.N.Marchenko, Polymer, 1986, **27**, 597.
97. A.J.Kitaigorodsky, "Molecular Crystals and Molecules", Plenum Press, New York–London, 1973.
98. O.K.Nugmanov and A.I.Pertsin, Symposium, Konformatsii i Funktsii Biologicheskikh Molekul.Teoreticheskie Aspekty" (The Conformations and Functions of Biological Molecules. Theoretical Aspects), Izd. Zinatne, Riga, 1984, p.143.
99. K.R.Andress, Z.Phys.Chem., 1929, **4**, 190.
100. M.K.Sen and S.C.Roy, Nature, 1954, **173**, 298.
101. C.Legrand, A.Bonnemayere, A.Guinier, and P.Antreuberger, Nature, 1955, **175**, 1045.
102. F.Kolpak and H.J.Blackwell, Macromolecules, 1976, **9**, 273.
103. R.E.Hunter and N.E.Dweltz, J.Appl.Polym.Sci., 1979, **23**, 249.
104. J.Maan, L.Roldan-Ganzales, and H.J.Wellard, J.Polym.Sci., 1960, **42**, 165.
105. O.K.Nugmanov, A.I.Pertsin, and G.N.Marchenko, Dokl.Akad.Nauk SSSR, 1983, **272**, 648.
106. A.I.Pertsin, O.K.Nugmanov, G.N.Marchenko, and A.I.Kitaigorodsky, Polymer, 1984, **25**, 107.

Nesmeyanov Institute of
Organoelemental Compounds,
USSR Academy of Sciences,
Moscow