Structure of Cellulose. 1. Low-Energy Conformations of Single Chains

Istvan Simon† and Harold A. Scheraga*

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

R. St. John Manley

Pulp and Paper Institute of Canada and Department of Chemistry, McGill University, Montreal, Canada H3A 2A7. *Received June 23, 1987*

ABSTRACT: Conformational energy calculations were carried out on single chains of cellulose. By studying oligosaccharide chains of increasing length, it proved possible to determine low-energy conformations of the disaccharide portions, independent of the length of the chain. Three different sets of interaction patterns in the stable disaccharide portions of infinite chains were found; these patterns differ in the nature of the groups involved in hydrogen bonds. Considering the disaccharide as the repeating unit, the combination of these three patterns resulted in six different conformations of single-chain cellulose. All have very similar conformational energies, indicating that the actual abundance of these six structures, when packed in native and processed cellulose microfibrils, is determined largely by interchain interactions or intrachain interactions between elements of the chain far from one another in the covalent structure. All (meta)stable conformations found are extended helices. The calculations indicate that a (high-energy) sharp fold can occur in the chain only if it is long enough to include compensating attractive interactions; i.e., no low-energy folds, stabilized by short-range interactions, were found.

Introduction

Cellulose, the polymer of 1–4-linked β-D-glucose is the most abundant natural polymer. It forms crystalline microfibrils found in all higher plants as well as in many bacterial, fungal and algal systems. In its crystalline microfibril form, the cellulose chain has a ribbonlike structure with a periodicity of 10.3 Å (length of a disaccharide unit), the pyranose rings are in a chair conformation, and the glycosidic bonds connecting them are in the equatorial position relative to the pyranose rings.[1-3] The unit cell dimensions and space group of the crystal lattice were first reported more than half a century ago.[4,5] Crystallographic data obtained on microfibrils, however, provide very limited information about the internal structure of cellulose. To obtain a reasonably detailed picture, these data must be combined with X-ray diffraction data of small cellulose components such as glucose, cellobiose, etc.[6,7] Since these components are bonded together by glycosidic C–O–C bonds, it should be mentioned that the C–O–C bond angle in methyl β-D-glucopyranoside is 113.1°[8] while higher values (115–117°) have been reported for cellobiose.[9,10] For long chains, however, the lower value of 113° is preferred.[11]

The arrangement of the cellulose chain molecules in the crystalline fibril has been the subject of prolonged controversy. According to X-ray diffraction studies, the crystalline elementary fibrils are composed of cellulose chains that are fully extended in the axial direction.[2,5-7] In contrast, electron microscope observations of negatively stained preparations of elementary fibrils show an axial periodic structure, leading to the proposal that the constituent cellulose chains are folded in such a way that the straight chain segments are aligned parallel to the fibril axis.[12,13]

Several attempts have been made to derive the conformation of the cellulose chain and its packing in the crystalline fibril by calculating low-energy conformations of cellobiose and methyl β-cellobioside.[14-16] Pizzi and Eaton extended these calculations to tetrasaccharides.[17,18] Since the low-energy conformations obtained in these computations depend strongly on the end groups (cellobiose versus methyl β-cellobioside) and on the length of the molecules (disaccharides versus tetrasaccharides), extrapolation from these data to the structure of an infinite cellulose chain can introduce uncertainties.

The purpose of this and the following paper[19] is to determine the possible structures of native and processed cellulose by means of conformational energy calculations on single chains and lattice energy calculations on three-dimensional crystalline structures. We attempt to resolve the apparent contradictions between the electron microscope observations[12,13] and the results of X-ray diffraction studies.[2,5-7] We also obtain more detailed information about the possible conformations of the cellulose chains and their interactions than is currently available from X-ray fiber diffraction and other experimental and theoretical techniques.

In this first paper, we present low-energy conformations of single chains of cellulose which are independent of the length of the chain and can serve as building blocks for the crystalline fibril. The accompanying paper[19] reports results for the calculated low-energy packing of the chains in the crystalline lattice.

Table I
Four Atoms Used To Define Dihedral Angles in the Disaccharide Unit

dihedral angle	atoms[a]			
1 (φ)	O5	C1	O4′	C4′
2	CH_3	O4	C4	C5
3	C4	C5	C6	O6
4	C5	C6	O6	H6
5	C1	C2	O2	H2
6	C4	C3	O3	H3
7	O5′	C1′	(O1′)[b]	CH_3
8 (ψ)	C1	O4′	C4′	C5′
9	C4′	C5′	C6′	O6′
10	C5′	C6′	O6′	H6′
11	C1′	C2′	O2′	H2′
12	C4′	C3′	O3′	H3′

[a] See the bottom structure of Figure 1 for the definitions of these atoms. [b] In a larger chain, O1′ would be designated O4″.

Methods

Generation of Single Cellulose Chains. The nomenclature used in this paper is shown in Figure 1. In

† Institute of Enzymology, Hungarian Academy of Science, P.O. Box 7, Budapest 1502, Hungary.

0024-9297/88/2221-0983$01.50/0

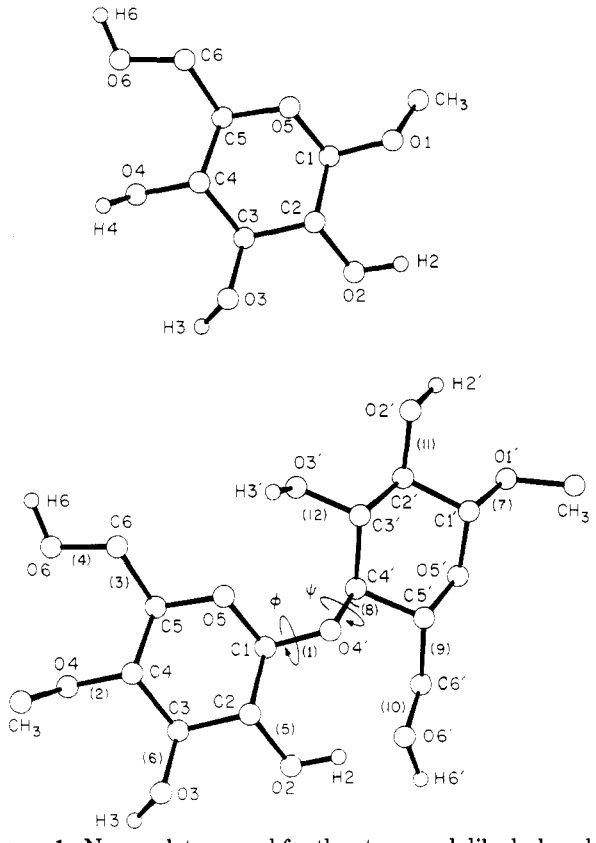

Figure 1. Nomenclature used for the atoms and dihedral angles in the mono- and disaccharide: top, methyl β-D-glucopyranoside (for glucose, the terminal CH_3 group is replaced by an H atom designated as H1); bottom, dimethyl β-D-diglucopyranoside (for cellobiose, the two terminal CH_3 groups are replaced by H atoms, designated as H4 and H1').

methyl β-D-glucopyranoside, the glycosidic oxygen is designated as O1. For di- and oligosaccharides, however, all glycosidic oxygens are designated as O4. Dihedral angles of 0° are assigned to the cis conformation. The four atoms used to define each variable dihedral angle are listed in Table I; rotation of the vector connecting the 3rd and 4th atoms relative to the vector connecting the 1st and 2nd atoms is positive if it is clockwise when viewed from the 2nd to the 3rd atom. These definitions of dihedral angles are similar to those used by Sathyanarayana and Rao,[20,21] the only difference being that carbon-bound hydrogens are not involved in the definitions because the united-atom approach is used in our calculations (see below).

Residues of the cellulose chain were generated from methyl β-D-glucopyranoside (obtained by replacing atom H1 of glucose by a methyl group—see Figure 1). The structure of this methyl derivative, which also includes a CH_3–O "glycosidic bond" (essential for our analysis), has been reported by Jeffrey and Takagi.[8] In this structure, we replaced atom H4 by a second methyl group. This second methoxy end group has a CH_3–O4 bond length of 1.380 Å and a C-O-C bond angle of 113.1°. The reason for substituting this extra CH_3 group is explained in section A of the Results. To generate the disaccharide, the CH_3–O4–C4 group of the second monomer unit becomes C1–O4'–C4' of the new glycosidic bond, as shown at the bottom of Figure 1. Larger oligosaccharides are generated in a similar manner, so that the chain always has a methyl group at each end. The geometries of the mono- and disaccharide units are given in Table II.

Conformational Energy Calculations. The choice of starting conformations for energy minimization is described in section B of the Results. The energy of any

Table II
Geometry of Mono- and Disaccharide Units

bond	length, Å	bond angle	value, deg
Methyl β-D-Glucopyranoside (Top of Figure 1)			
H4–O4	1.00	C1–O1–CH_3	113.1
O4–C4	1.43		
C1–O1	1.38		
O1–CH_3	1.43		
Dimethyl β-D-Diglucopyranoside (Bottom of Figure 1)			
CH_3–O4	1.38	CH_3–O4–C4	113.1
O4–C4	1.43		
C1–O4'	1.38	C1–O4'–C4'	113.1
O4'–C4'	1.43		
C1'–O1'	1.38	C1'–O1'–CH_3	113.1
O1'–CH_3	1.43		

oligosaccharide conformation is calculated[22] as a sum of energy terms arising from nonbonded, hydrogen bonding and electrostatic interactions between pairs of nonbonded atoms and intrinsic rotational barriers for each single bond. The partial charges of all the atoms, including the carbon-bound hydrogens of the 1,4-dimethyl β-D-diglucopyranoside, were determined by the CNDO/2 method.[23,24] The nonbonded energy was computed from a Lennard-Jones potential of the form

$$U_{ij} = \epsilon_{ij}[F(r_{ij}^0/r_{ij})^{12} - 2(r_{ij}^0/r_{ij})^6] \quad (1)$$

where ϵ_{ij} and r_{ij}^0 are the potential well depth and position of the minimum of the energy for the ijth pair of atoms and r_{ij} is the distance between atoms i and j. The same values of ϵ_{ij} and r_{ij}^0 were used as in our earlier calculations on oligosaccharides.[22] When atoms i and j were in a 1–4-position, i.e., separated by three covalent bonds, the value of $F = 0.5$; otherwise, $F = 1.0$. When an OH hydrogen interacted with an oxygen atom, a hydrogen-bonding function was used to calculate the interaction energy instead of the nonbonded potential function. The form of the hydrogen-bonding potential function is

$$U_{ij} = \epsilon_{ij}[(r_{ij}^0/r_{ij})^{12} - 2(r_{ij}^0/r_{ij})^{10}] \quad (2)$$

where ϵ_{ij} and r_{ij}^0 were the same as used earlier.[22] The electrostatic energy between atoms i and j was computed as the Coulombic energy between the partial charges of the atoms, with a dielectric constant of 2 (corresponding to an effective dielectric constant[24] of 4). The intrinsic torsional energy was calculated from the expression

$$U = (U_0/2)(1 + \cos 3\theta) \quad (3)$$

where U_0 is the barrier height when the dihedral angle θ was varied. U_0 was taken as 2.8 kcal/mol for >CH–CH_2– bonds and 0.6 kcal/mol for >CH–OH and >CH–O– bonds. All conformational energy calculations were based on the united atom approach;[22,25,26] i.e., the hydrogen atoms attached to carbon atoms were not treated explicitly. Rather, the parameters corresponding to a carbon atom in the calculation of nonbonded and electrostatic interactions were adjusted to simulate the effect of the hydrogen atoms bonded to it.

The FORTRAN program SMSNO[27] was used to find conformations corresponding to energy minima. During the energy minimization, the bond lengths and bond angles were held constant, and various conformations were generated by rotation around the single bonds that are not part of the pyranoside ring. The pyranoside ring was held fixed in the chair conformation.[1-3,8,22] The rigid chain conformation of the pyranoside ring was adapted from Jeffrey and Takagi[8] because it was based on a good crystal structure; in addition, the use of fixed bond lengths and bond angles is an essential feature of ECEPP (empirical

conformational energy program for peptides)[24] and its united-atom version UNICEPP.[26] The C–O–C bond angle of the glycosidic bond was generally set to 113.1° but, because of the uncertainty in the values reported for this angle,[3] we repeated the energy minimization calculations with (fixed) bond angles of 114°, 115°, 116°, 117°, and 118°, starting with conformations corresponding to each minimum of the ϕ,ψ map obtained with a bond angle of 113.1°.

Di-, tetra-, and hexasaccharides were studied. Since X-ray diffraction data show that the disaccharide is the repeating unit along the chain in any crystalline lattice, we were particularly interested in structures such as $i_1k_2i_3k_4i_5$ where i and k are separate (but may be similar) low-energy conformations of an overlapping disaccharide, and the subscripts represent the sequence number of the glycosidic link. In this example, monomer units 1 and 2 form a disaccharide with conformation type *i* (containing glycosidic bond number 1), and monomer units 2 and 3 form a disaccharide with conformation type k (containing glycosidic bond number 2), etc.; thus, $i_1k_2i_3k_4i_5$ represents a hexasaccharide. This is a useful nomenclature that can be applied to many complex polymer systems in which the repeating unit contain more than one subunit.

The conformations of two disaccharides were considered to be identical when all the corresponding dihedral angles differed by less than 15°. After energy minimization, a conformation k was considered to be independent of the chain length when the conformation k_2^4 (where the superscript indicates the length of the oligomer) in the structure of the tetrasaccharide $i_1^4k_2^4i_3^4$ was found to be the same as k_2^6 and k_4^6 in the minimum-energy conformation of the hexasaccharide $i_1^6k_2^6i_3^6k_4^6i_5^6$.

In order to eliminate the contributions from the ends of the chain, the conformational energy per disaccharide, characteristic of the conformation of a cellulose chain, was calculated as the difference between the conformational energy of the structure $i_1^6k_2^6i_3^6k_4^6i_5^6$ and that of the structure $i_1^4k_2^4i_3^4$, after the conditions $k_2^4 \equiv k_2^6 \equiv k_4^6$, $i_1^4 \equiv i_1^6$, and $i_3^4 \equiv i_5^6$ were found to be satisfied. In addition to these conformations, i.e., structures with disaccharides as the repeating units, we also carried out energy-minimization calculations on hexasaccharides in which the repeating unit was interrupted by one or more disaccharide structures corresponding to energy minimia of the (ϕ,ψ) map that were different from those of the rest of the hexasaccharide (see section C of Results), to assess the possible existence of low-energy sharp folds within the chain.

All the calculations reported in this paper were carried out on a Prime 550-II computer programmed in FORTRAN.

Results

A. Partial Charge Distribution in a Monosaccharide. Cellulose is a polymer of 1–4-linked β-D-glucose monomer units (see Figure 1) but, according to our CNDO calculation, the partial charges of O1 and O4 in glucose are different, and neither balances the sum of the partial charges of H1 and H4, even though the *net* charge on the glucose monomer is zero. A simple "polymerization" with such a constant partial charge distribution would result in a net charge proportional to the length of the chain as well as a false dipole moment because "elimination" of H_2O in the "polymerization" would remove a nonzero charge. This situation is the same when either of the hydrogens mentioned above is replaced by a methyl group. However, when *both* of these hydrogens are replaced by methyl groups, the CNDO calculation results in a partial charge distribution in which O1 and O4 are each found to have the same partial charges [−0.296 electronic charge units (ecu)], as does the sum of the two methyl groups (+0.148 ecu, each); i.e., the partial charge of one O1 (or O4) balances the partial charges of the *pair* of methyl groups, and the set of two methyl groups and an oxygen atom (which are "eliminated" upon "polymerization") has zero net charge. "Polymerization" of this monosaccharide with a constant charge distribution does not result in a net charge or a false dipole moment; therefore, dimethyl β-D-glucopyranoside is suitable for use as a monomer. The partial charges of all atoms of this monomer and united atoms are listed in Table III.

Table III
Partial Charges of Atoms and United Atoms of Dimethyl β-D-Glucopyranoside[a]

atom	partial charge, ecu[b]	atom	partial charge, ecu[b]
CH_3	0.1482	C3	0.1460
O4	−0.2964	O3	−0.3293
C4	0.1443	O6	−0.3086
C5	0.1447	H6	0.1706
C6	0.1415	H2	0.1743
O5	−0.3145	H3	0.1828
C1	0.3186	O1	−0.2964
C2	0.1405	CH_3	0.1482
O2	−0.3147		

[a] All carbon atoms are united atoms. Their charge contains the charge of the H atom(s) attached to them. C6 is bonded to two H atoms; the other carbon atoms are each bonded to one H atom. [b] Ecu = electronic charge unit.

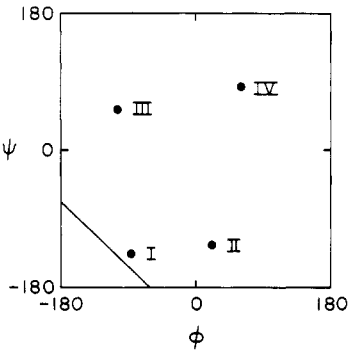

Figure 2. Location of the four energy minima and the line corresponding to the values of ϕ and ψ for a two-fold screw axis (a second portion of this line, near minimum IV, has been omitted). Relative energies in kcal/mol: I, 0.0; II, 18; III, 33; IV, 224.

It is worth noting that the significant influence of the terminal OH groups on the charge distribution of a disaccharide could be the origin of the higher value of the glycosidic bond angle in a disaccharide than in an oligosaccharide or in methyl β-D-glucopyranoside.[9-11]

B. Low-Energy Conformations of Disaccharides. To find all the minimum-energy conformations of the disaccharide, initial values of the dihedral angles ϕ and ψ (corresponding to rotation around the two single bonds of the glycosidic oxygen) were each taken every 40° between 0° and 320°. Initial values of the other dihedral angles, determining the conformations of the exocyclic functional groups and the end groups, corresponded to those found in the lowest energy monosaccharide obtained by first searching the dihedral-angle space of the monosaccharide. Energy minimizations were carried out from each of these 9 × 9 starting conformations and from all of the low-energy conformations found in our earlier paper,[22] allowing all 12 of the dihedral angles to vary.

Four regions of energy minima were found in the (ϕ,ψ) map (Figure 2). The lowest energy minimum is around

Table IV
High-Energy ($E > 5$ kcal/mol) Nonbonded Interactions in Dimethyl β-D-Diglucopyranoside

min	atom pair	energy, kcal/mol	min	atom pair	energy, kcal/mol
I	none	none	IV	O5–C6′	51.9
II	O5–C4′	13.5		O5–O3′	7.14
				C1–C5′	5.08
III	C1–C3′	19.2		C1–C3′	45.0
	C1–C5′	16.2		C2–C4′	34.4
				C2–C5′	43.2
				C2–C3′	26.5

($\phi = -86°$; $\psi = -136°$), called I, while the others are near II (21°, −124°), III (−103°, 55°), and IV (60°, 85°) with relative energies of 18, 33, and 224 kcal/mol, respectively. High energy ($E > 5$ kcal/mol) nonbonded interactions are listed in Table IV. In a second step, the dihedral angles 3, 4, 5, and 6 (see Table I) were set to all combinations of 0°, 120°, and 240° in the first residue with, in all cases, the corresponding dihedral angles (9, 10, 11, and 12) set to the same value in the second residue, while ϕ,ψ were set according to the positions of the four minima. Dihedral angles 2 and 7, which determine the conformations of the end groups, were set to the values of ψ and ϕ, respectively. Energy minimizations were carried out for these 4×3^4 conformations, varying all 12 of the dihedral angles during the minimization. In the lowest 4 kcal/mol energy range, seven conformations were found which differed in at least one of the dihedral angles. In all seven of these structures, the values of (ϕ,ψ) were close to (−86°, −136°) (± 10°), i.e., within region I.

Because of the uncertainty of the bond angle of the glycosidic bond, we repeated the energy-minimization calculations, starting with all conformations corresponding to energy minima in all the four regions of the (ϕ,ψ) map, by setting the angle of the glycosidic bond successively to 114°, 115°, 116°, 117°, and 118°. It was found that the energy differences among the regions of energy minima remained almost the same. All the lowest energy conformations still remain in region I, around (−86°, −136°) in the (ϕ,ψ) map, and the order of the energies of the conformations in this area remained unchanged when the valence angle of the glycosidic oxygen was altered. Therefore, all later calculations were carried out by retaining this bond angle at 113.1°.

C. Chain-Length-Independent Structure. Only some of the exocyclic functional groups of a disaccharide interact directly between the two monomer units. For example, when the structures are stabilized by O3′–H3′···O5 and O2–H2···O6′ hydrogen bonds, the values of dihedral angles 5, 9, 10, and 12 are thereby determined, but the values of the dihedral angles (3, 4, 6, and 11) of the other three functional groups can be changed without influencing these hydrogen bonds significantly. This fact made it possible to build 49 tetrasaccharide conformations from the seven low-energy disaccharide conformations, in which the interactions within the first two monomer units are the same as those within the last two monomer units, corresponding to one of the seven low-energy structures; at the same time, interactions within the central disaccharide correspond to the same or one of the other six low-energy disaccharide conformations. Thus, seven of these 49 tetrasaccharides are of type $i_1^4i_2^4i_3^4$ in which the interaction pattern within the three overlapped disaccharides are the same; and there are 21 pairs of additional conformations, each of types $i_1^4k_2^4i_3^4$ and $k_1^4i_2^4k_3^4$, respectively, in which the interaction pattern stabilizing the conformation of the central disaccharide is different from those stabilizing the outer disaccharides.

After energy minimization, in which all 24 dihedral angles were varied, the seven type i–i–i conformations fell into three minimum-energy classes, A, B, and C, resulting in three groups of tetrasaccharide, $A_1^4A_2^4A_3^4$, $B_1^4B_2^4B_3^4$, and $C_1^4C_2^4C_3^4$. It should be noted that the backbone dihedral angles (ϕ and ψ) differ very little (<10°) in classes A, B, and C and that the functional-group dihedral angles determining the positions of the atoms involving direct interactions within the disaccharide are the same in i_1^4, i_2^4, and i_3^4; it is only the dihedral angles determining the conformations of the exocyclic groups of the terminal monosaccharides that differ significantly. Energy minimization of the 21 pairs of structures mentioned above resulted in three pairs of structures, iki, kik, where i and k are stabilized by the same interaction patterns as A, B, or C. Therefore, the nine low-energy conformations found are $A_1^4A_2^4A_3^4$, $B_1^4B_2^4B_3^4$, $C_1^4C_2^4C_3^4$, $A_1^4B_2^4A_3^4$, $B_1^4A_2^4B_3^4$, $A_1^4C_2^4A_3^4$, $C_1^4A_2^4C_3^4$, $B_1^4C_2^4B_3^4$, and $C_1^4B_2^4C_3^4$. It should again be noted that only the dihedral angles determining the positions of the atoms interacting within the disaccharide structure are the same within a given structue type, e.g., A_2^4 in $A_1^4A_2^4A_3^4$, $B_1^4A_2^4B_3^4$, and $C_1^4A_2^4C_1^4$. The same statement is valid for all the other disaccharide structures as well.

In the next step, nine hexasaccharides were built from these nine tetrasaccharides by duplicating the central disaccharide. For example, $A_1^6B_2^6A_3^6B_4^6A_5^6$ was created by duplicating the central (overlapping) disaccharide in the center of the tetrasaccharide $A_1^4B_2^4A_3^4$. After energy minimization, which involved the variation of all 36 dihedral angles of the hexasaccharides, it was found that, in the three type $i_1^6i_2^6i_3^6i_4^6i_5^6$ hexasaccharides (i = A, B, or C) built from $i_1^4i_2^4i_3^4$ tetrasaccharides, $i_2^4 \equiv i_2^6 \equiv i_3^6 \equiv i_4^6$ within an accuracy of 2°, and $i_1^4 \equiv i_1^6$ and $i_3^4 \equiv i_5^6$ within an accuracy of 5° for all the dihedral angles of the corresponding disaccharide; i.e., in going from tetra- to hexasaccharides, i_1^4 and i_3^4 remained the same, and two i_2^4 disaccharides became inserted between these end disaccharides. Similarly, in the six type $i_1^6k_2^6i_3^6k_4^6i_5^6$ hexasaccharides (i, k = A, B, or C), built from the $i_1^4k_2^4i_3^4$ tetrasaccharides, $k_2^4 \equiv k_2^6 \equiv k_4^6$, $i_1^4 \equiv i_1^6$, and $i_3^4 \equiv i_5^6$ were found within an accuracy of 5°. For example, all the dihedral angles of B_2^4 in $A_1^4B_2^4C_2^4$ were found to be within 5° of those of B_2^6 and B_4^6 in the hexasaccharide $A_1^6B_2^6A_3^6B_4^6A_5^6$. It was also found, in the energy-minimized structures of $i_1^6k_2^6i_3^6k_4^6i_5^6$ and $k_1^6i_2^6k_3^6i_4^6k_5^6$, that $k_3^6 \equiv k_2^6 \equiv k_4^6$ for all the three pairs. This indicates that these remain minimum-energy structures regardless of the length of the chain.

The dihedral angles of the nine disaccharide structures found in the middle of the hexasaccharides are listed in Table V. This table also shows the conformational energy of these disaccharides relative to that of the most stable disaccharide, calculated as differences of the conformational energies of the corresponding hexa- and tetrasaccharide (see Methods section). In all of these low-energy structures ϕ and ψ are almost the same, near (−86°, −136°), point I in the (ϕ,ψ) map, and hence these are all extended structures. Increasing the glycosidic bond angle from 113.1° to 118° results in a decrease of the conformational energy of disaccharides A, B, and C by −1.45, −1.93, and −1.50 kcal/mol, respectively, but the order $E_B < E_C < E_A$ remains unaltered. Since the choice of various *fixed* bond angles added a constant energy, there was no need to take explicit account of the exoanomeric effect.

To check if long-range interactions, i.e., those between the two ends of the chain, can stabilize hairpinlike structures, one or more values of (ϕ,ψ) in the center of the chain

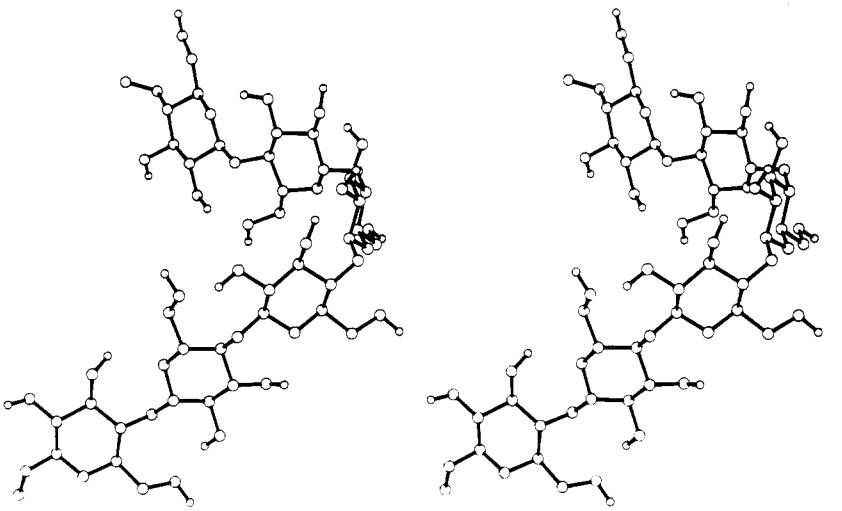

Figure 3. Stereoview of a hexasaccharide with a fold in the chain.

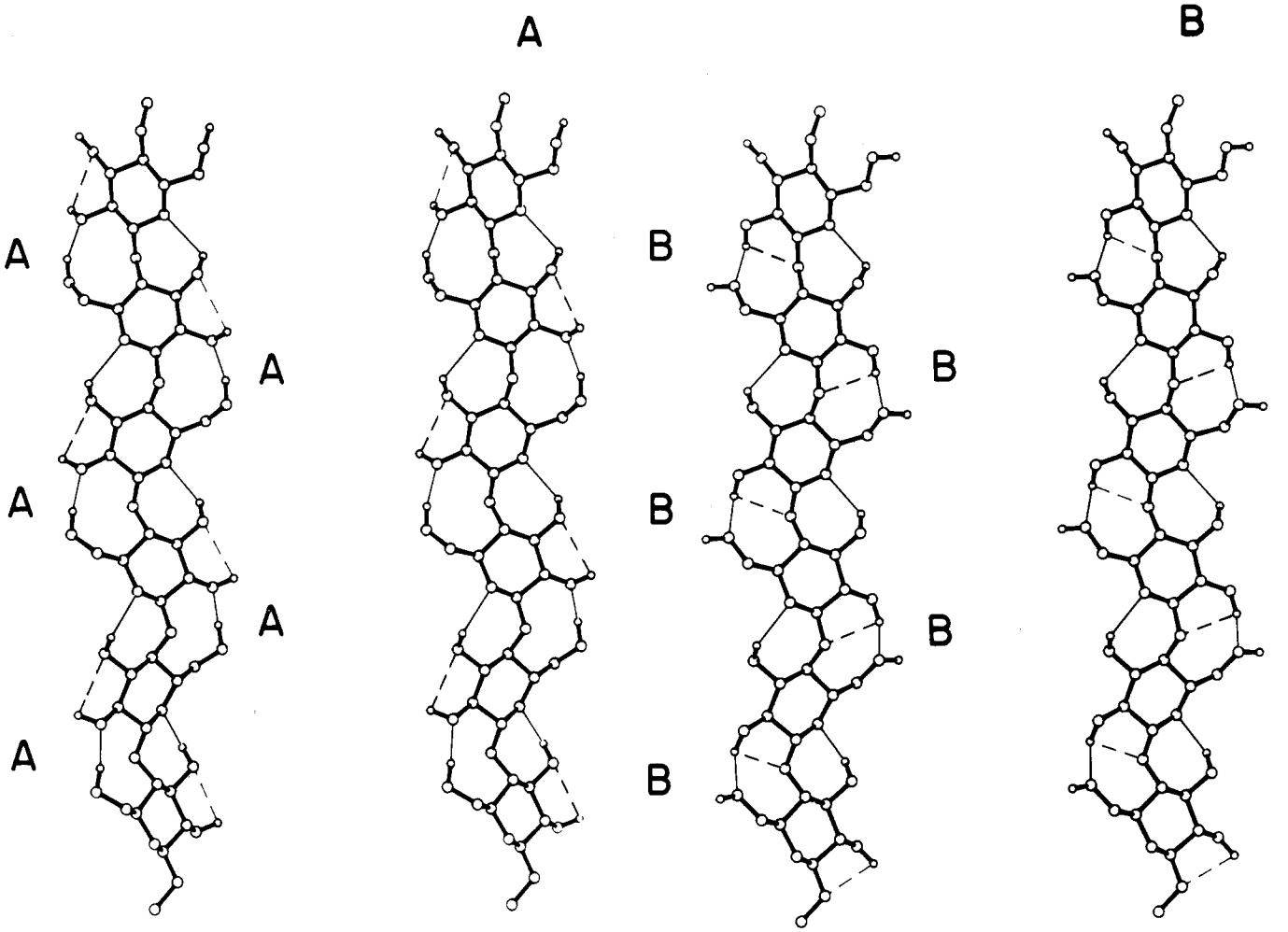

Figure 4. Stereoviews of the low-energy hexasaccharides (A) AA, (B) BB. Thin lines indicate strong hydrogen bonds ($E > 0.5$ kcal/mol); dashed lines indicate weak hydrogen bonds ($0.5 > E > 0.1$ kcal/mol).

were initially altered to values corresponding to the other three local minima (II, III, or IV) found in the (ϕ,ψ) map. Some of these initial hexasaccharide conformations placed the chain in a high-energy region of conformational space that was accessible to the global minimum without intervening potential energy barriers; consequently, after energy minimization, these initial conformations reverted to an extended structure. However, other of these initial hexasaccharide conformations placed the chain in high-energy local potential energy wells, with potential energy barriers separating them from the global minimum; consequently, after energy minimization, these chains maintained their folded forms (i.e., remained in high-energy local potential energy wells) with energies more than 40 kcal/mol higher than that of the extended chain. The structures with the fold always involved at least two higher energy minima [(ϕ,ψ) at II, III, or IV]. Figure 3 shows a folded structure whose energy is 65 kcal/mol higher than that of the extended chain. The backbone conformations of the glycosidic bonds on both sides of the third glucose unit correspond to minimum III.

D. Low-Energy Conformations of Long Cellulose

Table V
Dihedral Angles and Conformational Energies of Disaccharides Found in the Center of the Hexasaccharide Studied[a]

	AA*A*AA	BB*B*BB	CCCCC	AB*A*BA	BA*B*AB	AC*A*CA	CA*C*AC	BC*B*CB	CB*C*BC
1 (ϕ)	−89	−89	−86	−90	−90	−89	−86	−90	−86
2	222	216	225	217	222	225	223	225	217
3	261^{A}	275^{B}	62^{C}	275^{B}	261^{A}	62^{C}	260^{A}	62^{C}	277^{C}
4	106^{A}	214^{B}	80^{C}	193^{B}	105^{A}	85^{C}	105^{A}	82^{C}	188^{B}
5	157^{A}	64^{B}	153	158^{A}	65^{B}	156^{A}	157	66^{B}	154
6	86	96	82	86	91	84	84	85	84
7	270	271	274	275	271	273	271	274	275
8 (ψ)	−138	−144	−135	−138	−144	−138	−135	−144	−136
9	262^{A}	275^{B}	62^{C}	261^{A}	274^{B}	261^{A}	63^{C}	274^{B}	60^{C}
10	105^{A}	214^{B}	80^{C}	106^{A}	210^{B}	105^{A}	77^{C}	207^{B}	81^{C}
11	157^{A}	64^{B}	153	60^{B}	157^{A}	128	158^{A}	131	58^{B}
12	86	95	82	85	92	88	81	90	84
ΔE, kcal/mol	0.588	0.000	0.038	0.042	0.521	0.136	0.357	0.372	0.086
$\Delta \bar{E}$, kcal/mol	0.588	0.000	0.038	0.282		0.247		0.229	

[a] Superscripts A, B, and C indicate dihedral angles which define the positions of the atoms necessary to form A, B, or C interaction patterns. Italicized superscripts represent interactions appearing within the central disaccharide listed. Dihedral angles are quoted as positive angles between 0° and 360°, except for ϕ and ψ for which the range −180° to +180° is used to conform with the values in the (ϕ,ψ) map of Figure 2.

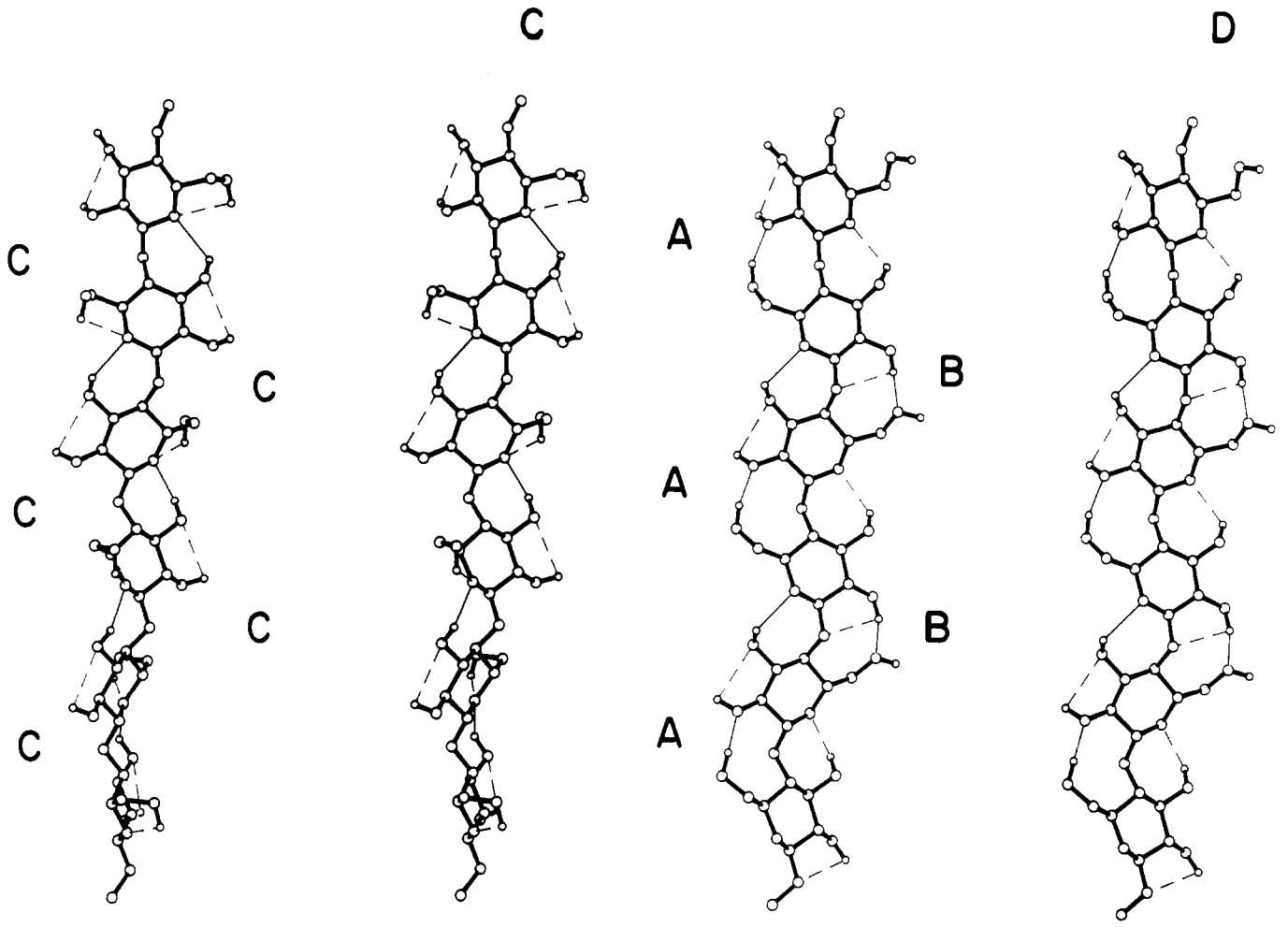

Figure 5. Stereoviews of the low-energy hexasaccharides (C) CC and (D) AB. Thin lines indicate strong hydrogen bonds ($E > 0.5$ kcal/mol); dashed lines indicate weak hydrogen bonds ($0.5 > E > 0.1$ kcal/mol).

Chains. Considering disaccharides as repeating units, six infinite chains can be built from the nine disaccharide structures found, since structures ...ikiki... and ...kikik... are the same for an infinite chain; these are $(AA)_n$, $(BB)_n$, $(CC)_n$, $(AB)_n$, $(AC)_n$, and $(BC)_n$. Stereoviews of the six hexasaccharides from which the six infinite chains can be built are shown in Figures 4–6.

Discussion

Low-energy conformations of single cellulose chains have been investigated. Three kinds of interaction patterns of glycosidic bonds, A, B, and C, stabilizing disaccharide conformations have been found. These define six infinite-chain structures. All of these structures are extended helices. Significant interactions occur only within glucose residues or between adjacent glucose residues in the chain. The strong O3′–H3′···O5 hydrogen bond ($E > 0.5$ kcal/mol) appears in all cases. Another hydrogen bond, different in interaction patterns A, B, and C, involves atom O6. In patterns A and B, O6′ and O2 of two consecutive residues are hydrogen bonded by H6′ and H2, respectively, while, in pattern C, H6 forms a hydrogen bond between O6 and

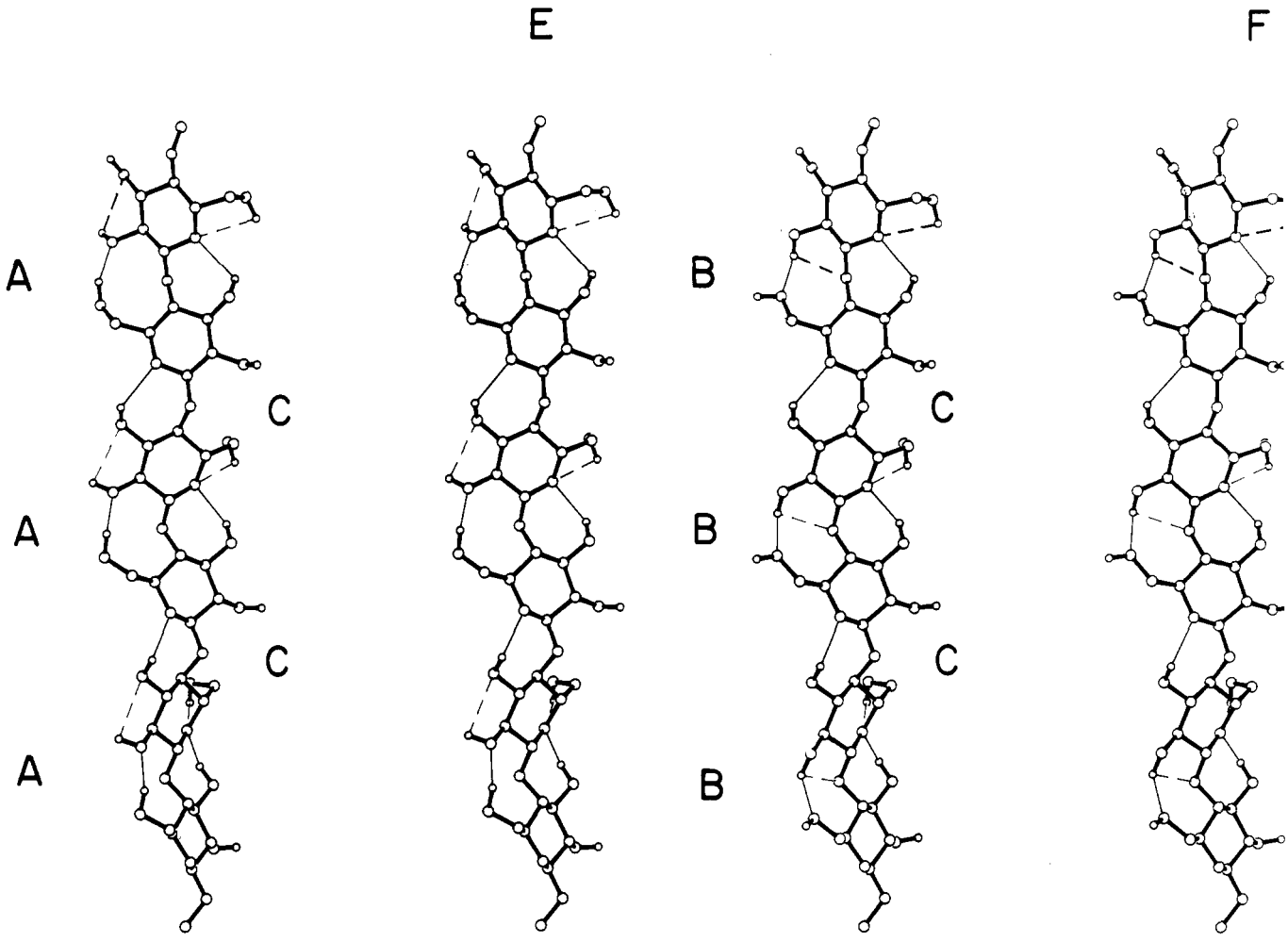

Figure 6. Stereoviews of the low-energy hexasaccharides (E) AC and (F) BC. Thin lines indicate strong hydrogen bonds ($E > 0.5$ kcal/mol); dashed lines indicate weak hydrogen bonds ($0.5 > E > 0.1$ kcal/mol).

O5 of the same glucose residue.

The last row of Table V shows that, in all six cases, the overall conformational energies are very similar. All the energies lie in a 0.6 kcal/mol interval suggesting that, in a real single cellulose chain, all the A, B, and C interaction patterns could be found with almost the same abundance. In an array of cellulose chains, the interchain interactions can shift this equilibrium and select one energetically favorable structure; this is the subject of the following paper.[19]

It is also interesting to note that, although $(ik)_\infty$ and $(ki)_\infty$ are the same infinite chain, there is a few tenths of a kilocalorie per mole difference between the conformational energy per disaccharide calculated in the two cases, because of the asymmetry in the disaccharide. These conformations should thus be characterized by their average energy ($\Delta\bar{E}$ in Table V).

The finding that the structure of a single cellulose chain is an extended one is in good agreement with previous calculations on smaller oligosaccharides.[14-18] One of the main results of our calculations is that the conformation of the disaccharide portion of the chain is independent of the chain length; therefore, it can be used as a building block of a crystalline lattice, as is done in the accompanying paper.[19]

Finally, we must mention that, although we found that the interaction of the two ends of a single hexasaccharide cannot support a hairpinlike turn structure, this does not rule out the possibility that an infinite chain can fold back on itself if the higher energy of the turning point is compensated by a *large* number of interactions in a crystalline lattice. This question will also be considered in the accompanying paper.[19]

Acknowledgment. This work was supported by research grants from the National Science Foundation (DMB84-01811), the National Institute of General Medical Sciences, National Institutes of Health (GM-14312), and the Natural Sciences and Engineering Research Council of Canada (Operating Grant A4529). Support was also received, in part, from the National Foundation for Cancer Research. We thank M. S. Pottle and S. Rumsey for help in setting up computer programs, Dr. M. R. Pincus and Dr. G. Némethy for stimulating discussions, and Prof. L. Glasser for help in preparing this manuscript.

Registry No. Cellulose, 9004-34-6; methyl β-D-glucopyranoside, 709-50-2; dimethyl β-D-diglucopyranoside, 112988-85-9.

References and Notes

(1) Hermans, P. H.; DeBooys, J.; Mann, C. J. *Kolloid-Z.* **1943**, *102*, 169–180.
(2) Gardner, K. H.; Blackwell, J. *Biopolymers* **1974**, *13*, 1975–2001.
(3) Marchessault, R. H.; Sundararajan, P. R. In *The Polysaccharides*; Aspinall, Gerald O., Ed.; Academic: New York, 1983; Vol. 2, pp 11–95.
(4) Meyer, K. H.; Mark, H. *Ber. Dtsch. Chem. Ges. B* **1928**, *61*, 593–614.
(5) Meyer, K. H.; Misch, L. *Helv. Chim. Acta* **1937**, *20*, 232–244.
(6) Gardner, K. H.; Blackwell, J. *Biochim. Biophys. Acta* **1974**, *343*, 232–237.
(7) Sarko, A.; Muggli, R. *Macromolecules* **1974**, *7*, 486–494.
(8) Jeffrey, G. A.; Takagi, S. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1977**, *33*, 738–742.
(9) Jacobson, R. A.; Wunderlich, J. A.; Lipscomb, W. N. *Acta Crystallogr. Sect. B: Struct. Crystallogr. Cryst. Chem.* **1961**, *14*, 598–607.

(10) Chu, S. S. C.; Jeffrey, G. A. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1968**, *24*, 830-838.
(11) Sundararajan, P. R. *Macromolecules* **1979**, *12*, 152-153.
(12) Manley, R. St. J. *Nature (London)* **1964**, *204*, 1155-1157.
(13) Manley, R. St. J. *J. Polym. Sci. Polym. Phys. Ed.* **1971**, *9*, 1025-1059.
(14) Rees, D. A.; Skerrett, R. J. *Carbohydr. Res.* **1968**, *7*, 334-348.
(15) Melberg, S.; Rasmussen, K. *Carbohydr. Res.* **1979**, *71*, 25-34.
(16) Pizzi, A.; Eaton, N. *J. Macromol. Sci. Chem.* **1984**, *A21*, 1443-1466.
(17) Pizzi, A.; Eaton, N. *J. Macromol. Sci., Chem.* **1985**, *A22*, 105-137.
(18) Pizzi, A.; Eaton, N. *J. Macromol. Sci. Chem.* **1985**, *A22*, 139-160.
(19) Simon, I.; Glasser, L.; Scheraga, H. A.; Manley, R. St. J. *Macromolecules*, following paper in this issue.
(20) Sathyanarayana, B. K.; Rao, V. S. R. *Biopolymers* **1971**, *10*, 1605-1615.
(21) Sathyanarayana, B. K.; Rao, V. S. R. *Biopolymers* **1972**, *11*, 1379-1394.
(22) Pincus, M. R.; Burgess, A. W.; Scheraga, H. A. *Biopolymers* **1976**, *15*, 2485-2521. Erratum: Ibid. **1977**, *16*, 468.
(23) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970; pp 57-153.
(24) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361-2381.
(25) Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 420-427.
(26) Dunfield, L. G.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1978**, *82*, 2609-2616.
(27) Gay, D. M. *Assoc. Comput. Mach. Trans. Math. Software* **1983**, *9*, 503-524.

Structure of Cellulose. 2. Low-Energy Crystalline Arrangements

Istvan Simon,[†] Leslie Glasser,[‡] and Harold A. Scheraga*

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

R. St. John Manley

Pulp and Paper Institute of Canada and Department of Chemistry, McGill University, Montreal, Canada H3A 2A7. *Received June 23, 1987*

ABSTRACT: Low-energy conformations and arrangements of cellulose chains in a crystalline structure have been investigated. Both metastable parallel and stable antiparallel arrangements exist. By analogy with the unit cell parameters and hydrogen bond networks suggested from X-ray diffraction data, the most stable parallel and antiparallel arrangements correspond to the cellulose I and II structures, respectively. Some low-energy structures obtained in the computations might correspond to cellulose III and IV, but the available experimental data are too incomplete to enable a definite identification to be made. Our data also suggest the presence of some structural inhomogeneity along and on the surface of the microfiber. Such inhomogeneities have been reported earlier, based on electron microscopy. A two-chain and an eight-chain unit cell model have been compared. The differences in conformation, packing arrangements, and energy are quite small and are at about the limit of experimental accuracy. A model is suggested for the transformation from metastable parallel-chain cellulose I to the more stable antiparallel chain cellulose II during mercerization.

Introduction

Early attempts to propose packing schemes for cellulose date back to the 1930s.[1] Although there are only a few stable conformations of a single chain,[2] these can be packed in several different arrangements. Thus, chains may be placed in parallel or in antiparallel alignment and, in both cases, one or more minimum-energy arrangements can be expected depending on the relative positions of the chains. By increasing the number of single-chain conformations involved, the number of theoretically possible packing arrangements increases dramatically. Many of these packing arrangements, however, may never appear because the corresponding local-minimum energies are too high relative to other minimum-energy arrangements. On the other hand, metastable packing arrangements may persist if large activation energies are necessary to alter the structure. Certainly, those transformations which involve a parallel ↔ antiparallel rearrangement must have large activation energies.

At least six different cellulose structures are known at present.[3] The best characterized are native cellulose (cellulose I), which is suggested to have a parallel arrangement[4-6] of chains, and mercerized cellulose (cellulose II), which can be produced from native cellulose by treatment with alkali and is suggested to have an antiparallel arrangement.[7-10] Less detailed information is available for cellulose III_I and III_{II}, which are produced from cellulose I and II, respectively, by liquid-ammonia treatment at −80 °C, and for cellulose IV_I and IV_{II}, which are products of glycerol treatment of cellulose III_I and III_{II} at 260 °C[3] (see Figure 1). It has been suggested that some among these six cellulose structures may represent the same conformation and packing but differ in the spatial arrangements of the microfibers.[3]

Most X-ray diffraction data indicate that the unit cell of the crystalline structure contains one disaccharide portion of each of two cellulose chains.[3] The more detailed diffraction patterns for cellulose I and II also show a few low-intensity reflections, the appearance of which depends on the source of the sample, which cannot be indexed on the basis of the simple two-chain unit cell model;[4,8] these data rather suggest that the unit cell dimensions normal to the fiber axis might each be twice as large as in the two-chain model so that the unit cell contains eight disaccharides.

The simplest, but fairly effective, hard-sphere calculations on cellulose packing[6,11,12] ignored the attractive terms involving the atomic interactions and took into account only repulsions. More detailed energy calculations, combined with X-ray, electron, and neutron diffraction data, led to the proposal that various hydrogen-bonded networks stabilize the structures of native and of processed cellulose.[5,7,13] Pizzi and Eaton[14-16] used conformational energy

[†] Institute of Enzymology, Hungarian Academy of Sciences, P.O. Box 7, Budapest 1502, Hungary.
[‡] Department of Chemistry, University of the Witwatersrand, Wits 2050, South Africa.

0024-9297/88/2221-0990$01.50/0