Carbohydrate Research, 157 (1986) 139–156
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

THE SEQUENCE STATISTICS AND SOLUTION CONFORMATION OF A BARLEY (1→3, 1→4)-β-D-GLUCAN

GREGORY S. BULIGA, DAVID A. BRANT*,
Department of Chemistry, University of California, Irvine, California 92717 (U.S.A.)

AND GEOFFREY B. FINCHER
Department of Biochemistry, La Trobe University Bundoora, Victoria 3083 (Australia)

(Received September 3rd, 1985; accepted for publication in revised form, January 2nd, 1986)

ABSTRACT

The sequence statistics and aqueous solution conformation of the 40° water-soluble (1→3,1→4)-β-D-glucan isolated from barley *(Hordeum vulgare)* have been modeled realistically using the known sequence-distributions of (1→3) and (1→4) linkages, theoretical conformational analysis, and the statistical mechanical theory of polymer-chain conformation. This barley β-glucan fraction consists of (1→4)-β-glucooligosaccharides, predominantly of d.p. 4 or less, joined by single β-(1→3) linkages. Approximate treatments of the sequence statistics which do not take into account the small mole fraction (~2%) of (1→4)-β-glucooligosaccharides of d.p. ~10 significantly underestimate the chain extension in solution. A correct prediction of the observed chain extension is achieved when these longer, highly extended (1→4)-β-glucooligosaccharide blocks are included in a model which randomly incorporates all (1→4)-β-glucooligosaccharide segments in the proportions observed experimentally. Chain flexibility in the 40° water-soluble β-glucan fraction is shown to arise principally from the isolated β-(1→3) linkages; blocks of two or more contiguous β-(1→3) linkages provide a source of additional flexibility which may influence the properties of barley β-glucan fractions containing a significant proportion of such sequences.

INTRODUCTION

Endosperm cell-walls of barley and other cereal grains are comprised predominantly of (1→3,1→4)-β-D-glucan and arabinoxylan; in barley endosperm the (1→3,1→4)-β-D-glucan accounts for up to 70% of the cell wall[1]. In addition to their central role in the molecular organization of cell walls, the barley (1→3,1→4)-β-D-glucans are important in the malting and brewing processes where they may contribute to filtration difficulties or to the formation of hazes in beer[2].

Although the barley (1→3,1→4)-β-D-glucans probably consists of a large

*To whom correspondence should be addressed.

0008-6215/86/$ 03.50

family of polysaccharides of varying molecular size and fine structure, the 40° water-soluble (1→3,1→4)-β-D-glucan isolated from barley flour has been selected for detailed chemical and physical analyses[3,4]. This fraction represents up to 20% of the total (1→3,1→4)-β-D-glucan of barley endosperm cell-walls[1].

The 40° water-soluble (1→3,1→4)-β-D-glucan consists of ~70% (1→4)-linked β-glucosyl residues and ~30% (1→3)-linked β-glucosyl residues, which are organized predominantly into blocks of two or three (1→4)-linked residues separated by single (1→3)-linked residues[5]. The polysaccharide might therefore be considered as a chain of (1→3)-linked cellotriosyl and cellotetraosyl residues (Fig. 1). Mathematical analysis suggests, subject to a number of theoretical and technical constraints, that the cellotriosyl and cellotetraosyl units are arranged randomly in the chain[6]. However, longer blocks of up to 10 adjacent (1→4) linkages have also been observed in relatively low frequency[3,4,7], and in some barley (1→3,1→4)-β-D-glucans contiguous (1→3) linkages have been reported[8–11].

Hydrodynamic measurements reveal that in aqueous media the 40° water-soluble (1→3,1→4)-β-D-glucan adopts an extended conformation corresponding to an axial ratio of ~100, if a prolate ellipsoidal shape is assumed[3]. In the present work we use the known sequence-distributions of (1→3)- and (1→4) linkages in conjunction with realistic molecular modeling and the statistical mechanical theory of flexible polymer-chain conformation to predict the conformation of the polysaccharide in aqueous solution. The predictions are based on:

1. Treatment of the (1→3,1→4)-β-D-glucan as a second-order Markov chain; this model accounts for first- and second-neighbor interdependences in the sequence of (1→3) and (1→4) linkages but not for longer-range effects.

2. Treatment as a simple Bernoullian sequence of (1→3)-linked cellotriosyl and cellotetraosyl units; this model takes no account of the longer blocks of up to 10 adjacent (1→4) linkages.

3. Treatment as a Bernoullian sequence of (1→3)-linked (1→4)-β-glucooligosaccharides of up to 10 adjacent (1→4) linkages, in the proportions determined by Woodward *et al.*[4].

Comparison of predictions from the three models with experimentally determined conformational properties of the barley (1→3,1→4)-β-D-glucan have enabled us to identify the important structural determinants of the shape and solution behavior of the molecule and to examine the effects of contiguous (1→3)-

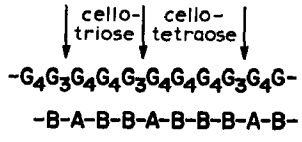

Fig. 1. Schematic representation of barley β-glucan structure. A represents a glucose residue linked at position 3 ($_3$G) or, alternatively, a β-(1→3) linkage. B represents a glucose residue linked at position 4 ($_4$G) or a β-(1→4) linkage. Arrows show that β-(1→3) linkages connect (1→4)-β-glucooligosaccharides.

linkages on these properties. Results of the successful treatment are presented as computer drawings of typical solution conformations of the polysaccharide.

MATERIALS

The 40° water soluble (1→3,1→4)-β-D-glucan purified from an Australian grown barley (*Hordeum vulgare* cv. Clipper) was compared with a commercially available preparation isolated in a similar manner from a barley harvested in the United Kingdom[3,4]. The commercial preparation was kindly donated by Mr. C. J. Dowzer (Bio-Con Australia Pty. Ltd.). Physical and chemical properties of the two (1→3,1→4)-β-D-glucan preparations are shown in Table I.

ANALYSIS OF THE SEQUENCE STATISTICS

Treatment as a second-order Markov process. — The known distribution of (1→3) and (1→4) linkages in the barley β-glucans in question[4] can be modeled as an n^{th} order Markov process[12]. The parameters of the model proliferate rapidly with increasing order, and, as will be shown here, a second-order Markov model is adequate to represent the present samples with reasonable accuracy. For the purposes of this discussion, let us denote glucose residues linked into the linear

TABLE I

CHEMICAL AND PHYSICAL PROPERTIES OF BARLEY (1→3,1→4)-β-D-GLUCAN[2-4].

	Source	
	Clipper	*Commercial*
Glucose (% of total monosaccharides)	98.3	97.9
Linkage position (% mol/mol)		
(1→3)-glucosyl	28	30
(1→4)-glucosyl	72	70
Adjacent (1→4)-linkages in blocks (% mol/mol)		
2	65	70
3	27	24
4	4	3
5	2	1
6–8	<1	<1
9–10	2	2
Intrinsic viscosity, $[\eta]$ (dl g^{-1})	6.90	4.26
Weight-average mol. wt., $\overline{M}_w$ (g mol^{-1})	290,000	160,000
Number-average mol. wt., $\overline{M}_n$ (g mol^{-1})	210,000	150,000
Osmotic second virial coefficient, A_2 (mL mol g^{-2})	~0	~0
Axial ratio, a/b		
(*a*) sedimentation	110	95
(*b*) viscosity	100	80
Characteristic ratio, C_∞	~18	~18

chain at C-3 ($_3$G) by the letter A and those linked at C-4 ($_4$G) by the letter B. All residues, except that at the reducing end of the chain, are also taken to be linked at C-1. This notation is shown in Fig. 1, where it may be observed that the residue at the non-reducing (left) end of any chain segment is not assigned as A or B, because the nature of its linkage to the preceding sugar may be unspecified. The same ambiguity does not exist for the reducing-end residue. With this scheme, when, as is usually the case, it becomes advantageous to consider the sequence of *linkages* rather than the sequence of residues, the (1→3) linkages are designated A and the (1→4) linkages, B. Because there is always one fewer linkage than residues in the chain, this scheme allows the type of every *linkage* to be specified. All linkage or residue sequences will, by convention, be read from left (non-reducing end) to right (reducing end).

The second-order Markov model assumes that the probability that a given linkage in the interior of a chain is of type k (k = A or B) depends on the type of its two predecessors i and j in the chain. We denote these probabilities by P_{ijk}, where P_{BBA}, for example, is the conditional probability that a particular linkage is of type A, given that its two immediate predecessors are both of type B. Eight such conditional probabilities P_{ijk} are defined for the second-order Markov model. In terms of these, the probability that an oligomer containing 11 residues is of sequence, BBABBBBABB is given by the product of 10 factors $P_B P_{BB} P_{BBA} P_{BAB} \cdot P_{ABB} P_{BBB} P_{BBB} P_{BBA} P_{BAB} P_{ABB}$. Two additional probabilities, P_i and P_{ij}, are required to describe the beginning of the sequence. Thus, P_B is the *a priori* probability that the first linkage at the reducing end of the chain is of type B. This is just the mole fraction of B linkages in the system; for long chains it becomes equal to the mole fraction of B residues. The parameter P_{BB} is the conditional probability that the second linkage of the chain is of type B, given that the first is also of type B. Complete treatments of Markov copolymer distribution statistics using matrix methods, are given by Hijmans[13] and Peller[14]. For our purposes it will suffice to employ the results of Ham[15], who showed that the probabilities P_i and P_{ij} are completely determined by the probabilities P_{ijk}. We need only observe here that the matrix product $\mathbf{P}_1\mathbf{P}_2\mathbf{P}^{x-2}$ generates all possible products of the type $P_A P_{AB} P_{ABB} P_{BBB} P_{BBA} \cdots$ which describe the probability of occurrence of each particular sequence of A and B linkages in a copolymer containing $x + 1$ residues. These probabilities are required in order to compute properly weighted averages of physical properties, *e.g.*, chain dimensions, over all sequences present in a macroscopic sample of chains of degree of polymerization $x + 1$. Here $\mathbf{P}_1$ and $\mathbf{P}_2$ are 4 × 4 diagonal matrices with elements P_A, P_A, P_B, P_B and P_{AA}, P_{AB}, P_{BA}, P_{BB}, respectively, and $\mathbf{P}$ is a 4 × 4 propagation matrix containing the eight P_{ijk} and eight zeros[13,14]. Because the probabilities are normalized, they sum to unity as expressed by the equation $[1,1,1,1]\,\mathbf{P}_1\mathbf{P}_2\mathbf{P}^{x-2}[1,1,1,1]^{tr} = 1$.

The barley β-glucan samples in question here are, to within a small experimental error, comprised of (1→4)-linked β-D-glucan segments, specifically the (1→4)-β-glucooligosaccharides *cellotriose* (–G_4G_4G–) and higher, linked by

isolated (1→3) linkages, as suggested in the depiction in Fig. 1. Exhaustive hydrolysis with barley (1→3,1→4)-β-D-glucan endohydrolase (EC 3.2.1.73), an enzyme which specifically hydrolyzes (1→4) linkages only where the glucosyl residue is substituted at O-3, and subsequent chromatographic fractionation of the hydrolyzate, have been employed[4] to establish for the present samples the distribution function (i.e., mole fraction) $p_{B,b}$ of (1→4)-β-glucooligosaccharides of each degree of polymerization b (Table I). Experimental values of $p_{B,b}$ for one of the two samples (Clipper) are plotted as the filled circles in Fig. 2. The corresponding experimental distribution function of (1→3) linkages $p_{A,a}$, plotted as filled squares in Fig. 2, has the simple form $p_{A,1} = 1.0$, $p_{A,a} = 0$ for $a > 1$. These are the distribution functions for the sample as a whole; individual polymer chains in the sample may deviate from them, but it is assumed here that for long chains $p_{A,a}$ and $p_{B,b}$ approach the reported sample-averages. In what follows, attention will be confined exclusively to the higher-molecular-weight sample, Clipper, since the sequence statistics and unpertubed dimensions for the two polymers are very similar.

The parameters of the second-order Markov model are related to the experimental observables $p_{A,a}$ and $p_{B,b}$ by the relations[15]:

$$p_{A,1} = P_{BAB} = 1 - P_{BAA} \qquad a = 1 \qquad (1)$$

$$p_{A,a} = P_{BAA}P_{AAA}^{a-2}P_{AAB} = P_{BAA}P_{AAA}^{a-2}(1 - P_{AAA}) \qquad a > 1 \qquad (2)$$

$$p_{B,1} = P_{ABA} = 1 - P_{ABB} \qquad b = 1 \qquad (3)$$

$$p_{B,b} = P_{ABB}P_{BBB}^{b-2}P_{BBA} = P_{ABB}P_{BBB}^{b-2}(1 - P_{BBB}) \qquad b > 1 \qquad (4)$$

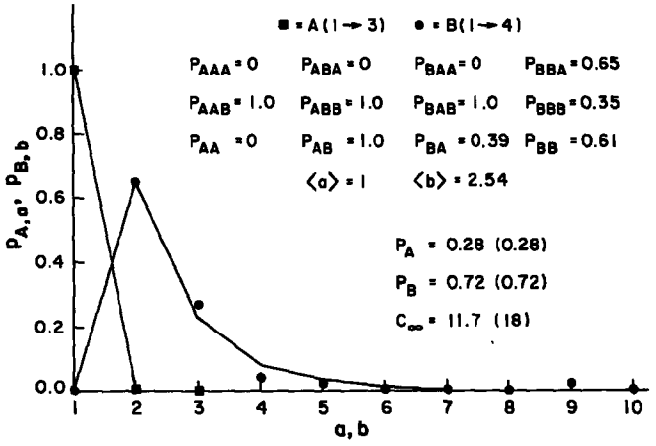

Fig. 2. Plot of the distribution function (mole fraction) $p_{A,a}$ of sequences of A linkages of length (a) *vs.* sequence length (a) and a similar plot of the distribution function $p_{B,b}$ of sequences of B linkages of length (b) *vs.* sequence length (b). Experimental results[4] are shown as filled squares [A, i.e., β-(1→3) linkages] and filled circles [B, i.e., β-(1→4) linkages]. Second-order Markov model calculations of $p_{A,a}$ and $p_{B,b}$, shown as solid curves, are based on the parameters P_{ijk} given. Calculated parameters P_{ij}, P_i, ⟨a⟩, ⟨b⟩, and C_∞ defined in the text are also shown. Experimental values of P_A, P_B, and C_∞ are shown in parentheses.

Using experimental values of $p_{A,1}$, $p_{A,2}$, $p_{B,1}$, and $p_{B,2}$ in conjunction with equations (*1–4*) and the four necessary relationships[15] $P_{ijA} + P_{ijB} = 1$, one may obtain values for all eight parameters P_{ijk}. Alternatively, as $p_{A,a+n}/p_{A,a} = P_{AAA}^{n} = (1 - P_{AAB})^{n}$ when $a > 1$ and n is a positive integer, the ratio of any two mole fractions $p_{A,a}$ for $a > 1$ can be used to obtain P_{AAA} and P_{AAB}, and any $p_{A,a}$ will then yield P_{BAA} and P_{BAB} once P_{AAA} and P_{AAB} are known. Parameters P_{BBB}, P_{BBA}, P_{ABB}, and P_{ABA} may obviously be evaluated from the mole fractions $p_{B,b}$ in similar fashion.

Equations (*2*) and (*4*) illustrate an important limitation of any Markov model of order y: The ratios $p_{A,a+n}/p_{A,a}$ and $P_{B,b+n}/p_{B,b}$ are each restricted to a single value for all $a,b > y$, and the distribution functions $p_{A,a}$ and $p_{B,b}$ must consequently decline monotonically for $a,b > y$. Because the distribution function $p_{B,b}$ has a maximum at $b = 2$ for the present barley glucans (Fig. 2), a first-order Markov model, which accounts only for first-neighbor interdependences, cannot be used to represent their sequence statistics. As $p_{B,b}$ has a small subsidiary maximum at $b = 9$, a Markov model of at least order 9 would be required to reproduce this feature of the Clipper distribution function.

Observables other than $p_{A,a}$ and $p_{B,b}$ can of course be used to help evaluate the Markov parameters P_{ijk}. In the present case, for example, $P_A = 1 - P_B = 0.28$ is experimentally determined for Clipper, and this information could be used. Here, however, we will consider the ability of the model to reproduce the experimental value of P_A as one test of the choice of parameters P_{ijk} evaluated from $p_{A,a}$ and $p_{B,b}$. It is also of interest to note that number-average sequence lengths $\langle a \rangle$ and $\langle b \rangle$, sometimes referred to as the "relative weight" of a and b units[15], are related to the P_{ijk} by:

$$\langle a \rangle = 1 + \frac{P_{BAA}}{P_{AAB}} \tag{5}$$

$$\langle b \rangle = 1 + \frac{P_{ABB}}{P_{BBA}} \tag{6}$$

We observe finally that the second-order Markov treatment reduces to a first-order treatment if $P_{AAA} = P_{BAA} = P_{AA}$, $P_{AAB} = P_{BAB} = P_{AB}$, $P_{ABA} = P_{BBA} = P_{BA}$, and $P_{ABB} = P_{BBB} = P_{BB}$. When, in addition, $P_{AA} = P_{BA} = P_A$ and $P_{AB} = P_{BB} = P_B$, the copolymer is Bernoullian or random.

Treatment as a Bernoullian sequence of (1→3)-linked cellotriosyl and cellotetraosyl units. — To a good approximation, the barley β-glucan samples in question may be represented as a Bernoullian copolymer of (1→3)-linked cellotriosyl and cellotetraosyl units[6]. If we let P_3 represent the mole fraction of cellotriosyl units and P_4, the mole fraction of cellotetraosyl units in the system, then the following obvious relationships allow one to connect the observables P_A, P_B, $p_{A,a}$ and $p_{B,b}$ with the parameters P_3 and P_4 of this Bernoullian model:

$$P_A = \frac{1}{1 + 2P_3 + 3P_4} \tag{7}$$

$$P_B = \frac{2P_3 + 3P_4}{1 + 2P_3 + 3P_4} \tag{8}$$

$$p_{A,1} = 1 \tag{9}$$
$$p_{A,a} = 0,\ a \geqslant 2 \tag{10}$$
$$p_{B,1} = 0 \tag{11}$$
$$p_{B,2} = P_3 \tag{12}$$
$$p_{B,3} = P_4 \tag{13}$$
$$p_{B,b} = 0,\ b \geqslant 4 \tag{14}$$

The mean sequence-lengths of A and B linkages are evidently given by:

$$\langle a \rangle = 1,\ \langle b \rangle = 2P_3 + 3P_4 \tag{15}$$

This model may be regarded as a special case of third-order Markov statistics, and equations (*7*)–(*15*) may all be derived from equations for the third-order model analogous to those for the second-order model already discussed[15]. (We have not obtained equations (*9*)–(*14*) from Ham's third order expressions for $p_{A,a}$ and $p_{B,b}$, which appear to be incorrect, but rather from straightforward extensions of equations (*1*)–(*4*) to the third-order case.)

Monte Carlo generation of sequence distributions. — We may also model the β-glucan chains more realistically as a Bernoullian sequence of (1→3)-linked (1→4)-β-glucooligosaccharides in the proportions specified by the full experimental distribution function $p_{B,b}$. The true distribution function has a maximum at b = 9 and 10 for Clipper and commercial, respectively, and decays to zero thereafter so that, by analogy with the Bernoullian model discussed in the previous section, a Markov model of order 9–10 would be required to carry out averaging over all possible A and B sequences using the matrix approach. Since this is impractical, we turn for this treatment to a Monte Carlo sampling technique in which the averaging is carried out over a representative computer-generated sample of copolymeric sequences for chains with a given degree of polymerization (d.p.). The Monte Carlo generation process is straightforward: The computer selects in sequence a series of random numbers between 0 and 1. Since $\Sigma_b p_{B,b} = 1$, each random number can be associated with one of the possible (1→4)-β-glucooligosaccharides. In this way, each random number in the sequence corresponds to a (1→4)-β-glucooligosaccharide of some length. These are incorporated into the polymer in the same sequence as the random numbers and in proportions consistent with the $p_{B,b}$. Evidently, the sequence of (1→4)-β-glucooligosaccharides in any given polymer chain of the Monte Carlo sample is Bernoullian and will depend on the random-number sequence used to generate that chain. Only for long chains will the distribution

function $p_{B,b}$ for a given chain match that of the sample as a whole, so that a sample large enough to be representative must be chosen. The requisite size of the sample can be determined by investigating the convergence of computed averages as a function of the number of members of the sample.

CALCULATION OF UNPERTURBED POLYMER-CHAIN DIMENSIONS

Conformational and sequence-distribution averaging by matrix multiplication. — The solution dimensions of the barley β-glucans, unperturbed by long-range volume exclusion, may be described by the characteristic ratio $C_x = \langle r_x^2 \rangle_0 / xL^2$, defined as the mean-square unperturbed end-to-end distance $\langle r_x^2 \rangle_0$ for the real chain consisting of x residues each of mean-square length L^2 normalized by the hypothetical mean-square unperturbed end-to-end distance xL^2 for the same chain with residues linked by free (universal) joints[16–18]. For flexible polymers, C_x increases with increasing x to a limiting value C_∞, which reflects the contribution of linkage geometry and glycosidic-bond torsional constraints on the mean extension of the chain.

The quantity $\langle r_x^2 \rangle_0$ is measurable. For the copolymers in question, the measured quantity reflects a conformational average over all of the accessible conformations of each chain in the sample, and, since these chains differ in general in copolymer sequence and d.p., over both the sequence and chain-length distributions of the sample. The nature of the average over chain length depends on the experimental technique employed to measure $\langle r_x^2 \rangle_0$; the quantity x in the denominator of C_x must be an average chosen in a consistent way in order to obtain a properly normalized C_∞ value[19,20]. The denominator quantity L^2 is also an average, which depends on copolymer composition[16]. It is defined as $L^2 = P_A L_A^2 + P_B L_B^2$, where L_A and L_B are the lengths of the glycosidic oxygen to glycosidic oxygen vectors (virtual bonds) spanning residues of type A and B, respectively. Using methods described earlier[19], experimental values of C_x (taken here to represent C_∞) may be estimated from the reported experimental molecular weight, molecular-weight distribution, intrinsic viscosity, and second virial coefficient (Table I) to equal 18 ±1.

The characteristic ratio C_∞ may also be calculated from realistic molecular theory[16]. In the treatments discussed here, which model the copolymer sequence distribution as a second-order Markov process, we have carried out the averages *simultaneously* over all chain conformations and sequence distributions, by using an elaboration of the matrix multiplication method of Miller, Brant, and Flory[21], who treated the first-order Markov case. Matrices $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}$, required to model the sequence distributions, have just been described. The required conformationally averaged coordinate transformation matrices $\langle \mathbf{T}_{ij} \rangle$, for i,j = A,B, may be evaluated for each kind of nearest-neighbor residue pair from the residue and linkage geometries and the associated conformational-energy surfaces estimated with "molecular mechanics" using approximate conformational-energy func-

tions[16–18]. (In fact we were able to use $\langle \mathbf{T}_{AA} \rangle = \langle \mathbf{T}_{BA} \rangle$ and $\langle \mathbf{T}_{BB} \rangle = \langle \mathbf{T}_{AB} \rangle$ to an excellent approximation. This simplification may be readily understood from structural drawings of the four dimers, where it may be observed that the nonreducing residue, because it is β-D-glucose in every case, contributes the same geometric and energetic characteristics to the linkage region; differences arise almost exclusively from the contributions of the reducing sugar. In effect, only two energy surfaces were needed, those corresponding to having either A or B in the reducing position at the glycosidic linkage.) The only additional parameters of the theory are the virtual-bond vectors $\mathbf{L}_A$ and $\mathbf{L}_B$ for residues of type A and B. All structural information for β-D-glucose was taken from Arnott and Scott[22]. The calculation is carried out in the usual approximation[16,18] that bond rotations ϕ and ψ at a given glycosidic linkage are independent of such rotations at neighboring and more remote linkages. Theoretical averaging over the chain-length distribution is obviated by proper choice of the denominator x in computing the experimental value[19] of C_∞.

Monte Carlo averaging over the conformational and sequence distributions. — In treatments involving Monte Carlo averaging over the sequence distribution, it is possible to average over the conformations of each sequence of given d.p. by using matrix multiplication methods[16]. This approach has been used by Hallman and Whittington[23] in their treatment of several copolysaccharides. It may be instructive, however, to average over both the conformational and sequence distributions using Monte Carlo methods. Chains of representative sequence and conformation may, for example, be extracted from the Monte Carlo sample for visualization by computer-graphics techniques. We have consequently chosen in the present case, where the average over the sequence distribution was done by Monte Carlo methods, to carry out the conformational average in the same way.

Monte Carlo averaging over the possible conformations of a copolymer chain of given sequence and length may be understood to proceed as follows: For each pair of linked residues in the chain, the computer assigns a conformation consistent with the conformational-energy surface appropriate to that linkage, so that lower-energy linkage conformations occur more frequently than those of higher energy in accordance with the Boltzmann distribution law. This process is repeated to generate a sample of chain conformations large enough to ensure that averages, e.g., of the squared end-to-end distance, over this sample yield results representative of the entire population of conformations available to the chain of given sequence and length. The details of the process for obtaining $\langle r_r^2 \rangle_0$ by this method have been described elsehwere[24].

RESULTS AND DISCUSSION

Second-order Markov model. — Using the experimental values of $p_{A,1}$, $p_{A,2}$, $p_{B,1}$, and $p_{B,2}$ for Clipper[4] in conjunction with equations *1–4*, the parameters P_{ijk} given in Fig. 2 are obtained. These yield the calculated second-order Markov distribution functions $p_{A,a}$ and $p_{B,b}$ plotted as solid lines in Fig. 2. Agreement with

experiment is necessarily complete for $p_{A,1}$, $p_{A,2}$, $p_{B,1}$ and $p_{B,2}$; it is reasonably good for $p_{B,b}$ with b >3, as one may observe by comparing the calculated and observed (filled circles) distribution-functions. Experimental values of P_A and P_B, shown in parentheses in Fig. 2, are reproduced exactly when calculated from the P_{ijk} using the relationships given by Ham[15], but the theoretical value of C_∞ is only about two-thirds of the experimental value (also given in parentheses). Calculated values of ⟨a⟩ and ⟨b⟩ are also reported in Fig. 2. No other choice of experimental observables from which the P_{ijk} may be evaluated produces any better agreement of the calculated and observed parameters ($p_{A,a}$, $p_{B,b}$ P_A, C_∞) within the constraints of the second-order Markov model.

Results of calculations presented in Fig. 3 demonstrate that increases in the average length of the highly extended (1→4)-β-glucooligosaccharide B sequences serve to increase the calculated value of C_∞. Here C_∞ and ⟨b⟩ are plotted versus P_B for a series of calculations in which the P_{ijk} were chosen to generate Markovian copolymers with increasing amounts of B ranging from the repeating copolymer $(-ABB-)_n$ with $P_B = 0.67$ to the pure homopolymer of B ($P_B = 1.00$). The model of Fig. 2 falls on these curves at $P_B = 0.72$, where ⟨a⟩ = 1.00, ⟨b⟩ = 2.54, and C_∞ = 11.7. The experimental value of C_∞ (18) is reproduced in Fig. 3 near $P_B = 0.79$, where ⟨a⟩ = 1.00 and ⟨b⟩ = 3.78. These larger values of C_∞ and ⟨b⟩ correspond to a flattened distribution-function $p_{B,b}$ which still has its maximum at b = 2 ($p_{B,2}$ = 0.36) but has only decayed at b = 5 to $p_{B,5} = 0.09$. Other calculations, e.g., treatment of the chain as a Bernoullian copolymer of A and B, likewise show that the calculated chain-extension correlates strongly with ⟨b⟩.

On the other hand, it is readily shown that introduction of sequences of A units longer than unity usually produces a reduction in calculated dimensions, be-

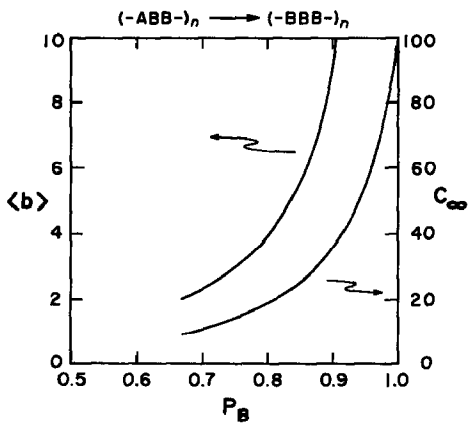

Fig. 3. Plots of mean length of B sequences ⟨b⟩ (left curve) and characteristic ratio C_∞ (right curve) *vs.* mole fraction P_B of B linkages in hypothetical barley β-glucan chains comprising (1→4)-β-glucooligosaccharides linked by isolated A, i.e., β-(1→3), linkages. Parameters P_{ijk} of the second-order Markov model were chosen to cause the character of the copolymer to change with increasing P_B, as shown along the top margin.

cause the greater inherent conformational freedom of the β-(1→3) linkage[25,26] confers flexibility on the chain, especially when several such linkages occur in sequence. Thus, calculations analogous to those in Fig. 3, which correspond to a transition from the repeating copolymer $(-AABB-)_n$ to homopolymeric B, show that the calculated C_∞ is always significantly smaller (20–200%) at a given value of P_B or ⟨b⟩ for the copolymers containing isolated AA units than for those containing isolated A units (data not shown).

These results suggest that a model which correctly accounts for the small maximum in $p_{B,b}$ at b = 9, may yield better agreement with the observed C_∞. This possibility is explored in a subsequent section. First, it is of interest to demonstrate briefly that certain justifiable modifications of the second-order Markov model cannot correct the discrepancy in calculated and measured values of C_∞ reported in Fig. 2.

Variations of the second-order Markov model. — To within experimental error both the Clipper and commercial β-barley glucans contain only isolated A units ($p_{A,1} = 1$, $p_{A,a>1} = 0$) but no isolated B units[4] ($p_{B,1} = 0$). We investigated in calculations reported in Fig. 4 the effect on the calculated chain dimensions of allowing $p_{B,1} = p_{A,2} = 0.05$ to account for a maximum possible experimental error of 5% in the chemical analysis of the polymer-sequence distribution. This change increases the calculated C_∞ from 11.7 (Fig. 2) to 12.7, a result still well below the value of 18 found experimentally. The increase of one unit in the theoretical C_∞ clearly derives from the flattened distribution $p_{B,b}$ and the correspondingly larger value of ⟨b⟩ found when the results in Fig. 4 are compared with those presented in Fig. 2. If we relax by 5% only the original constraint on $p_{B,1}$ and retain the apparent experimental finding $p_{A,2} = 0$, then theoretical results virtually indistinguishable from those of Fig. 4 are obtained. When the calculation is carried out with $p_{B,1} = 0$ and $p_{A,2} = 0.05$, the distribution function $p_{B,b}$ is again flattened in comparison with Fig. 2, but C_∞ increases negligibly, because of the compensating occurrence of some 10% of the A units in AA sequences, which, as already noted, promote chain flexibility. We conclude from these results that the second-order Markov model yields calculated unperturbed chain dimensions that are quite insensitive to any reasonable experimental uncertainties in the values of $p_{A,a}$ and $p_{B,b}$. (It should be noted that all of the results presented in this paragraph depend upon choices of the Markov parameters P_{ijk} consistent with the overall copolymer composition, i.e., the experimental result $P_A = 0.28$ was used in assigning values to the P_{ijk}.)

A second source of concern about the reliability of the calculations reported in Fig. 2 lies in the details of the theoretical model for calculating C_∞. The details in question pertain to the choice of skeletal geometry and the conformational potential-energy functions. The mean β-D-glucose geometry recommended by Arnott and Scott[22] was used throughout. For all of the calculations reported here the valence angle β at the glycosidic bridge[17] was taken to be 117°, unless otherwise indicated. Conformational potential-energy surfaces for β-(1→3) and β-(1→4) linkages were "unrefined" in the sense described by Burton and Brant[25]. Moreover,

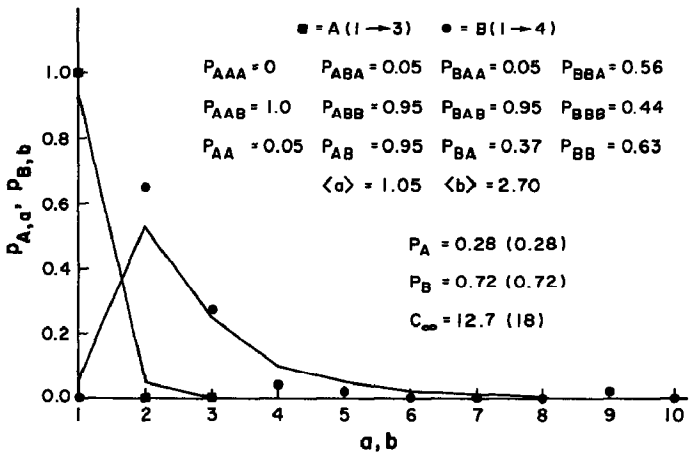

Fig. 4. Results of a calculation using the second-order Markov model presented as in Fig. 2 to illustrate the effect on the calculated results of a possible error of 5% in the experimental values of $p_{B,1}$ and $p_{A,2}$.

no terms were included in the potential functions to account explicitly for torsional strain arising from the electronic characteristics of the chemical bonding or for polymer–solvent interactions.

The energy surface for the β-(1→4) linkage is known to overestimate C_∞ for long cellulosic chains by a factor of about three and to underestimate the temperature coefficient of those dimensions by a similar factor[25]. This discrepancy has been explained[27,28] by the occasional occurrence of glucose ring conformers other than the most stable $^4C_1$(D) chair upon which the present β-(1→4) energy-surface is exclusively based. Occurrence of alternative glucose ring conformers has a marked effect on the directional persistence of long and highly extended cellulosic chains comprised primarily (~99%) of residues in the $^4C_1$(D) form. It will have no perceptible influence on the observable conformational properties of the barley β-glucans consisting of short (1→4)-β-glucooligosaccharides (d.p. predominantly less than 6) connected by β-(1→3) linkages possessing relatively great conformational freedom. We believe, therefore, on the basis of extensive prior work[25,27–29] that the energy surface employed here for β-(1→4) linkages describes adequately the conformational characteristics of these shorter (1→4)-β-glucooligosaccharides.

There has been no systematic effort to refine the energy surface for the β-(1→3) linkage against appropriate experimental data. In an effort to examine the sensitivity of the calculated barley β-glucan C_∞ to the details of the β-(1→3) energy-surface we have calculated C_∞ using β-(1→3)-surfaces based on a reasonable range of glycosidic bridge angles β. In general, a decrease in β engenders more steric constraint in an energy surface and thus produces less conformational freedom in the linkage. Variations of the glycosidic bridge angle within the range $115° \leqslant \beta \leqslant 119°$ in calculating the energy surface for the β-(1→3) linkage yielded no significant changes in the calculated value of C_∞ for the copolymer. A similar investigation of the sensitivity of the calculated C_∞ to the details of the β-(1→4) energy-surface

showed that the calculation was likewise insensitive to this feature of the calculation. We conclude, therefore, that C_∞ calculated from the model is rather strongly dependent on the sequence statistics as reflected in the distribution functions $p_{A,a}$ and $p_{B,b}$ but much less so on the details of either of the required conformational energy surfaces.

Monte Carlo treatment of a Bernoullian copolymer of β-(1→3)-linked (1→4)-β-glucooligosaccharides. — When the barley β-glucan is modeled as a Bernoullian sequence of β-(1→3)-linked (1→4)-β-glucooligosaccharides in the proportions specified by the actual experimental distribution function $p_{B,b}$, using Monte Carlo averaging over the sequence distribution and over the conformations available to each sequence, the calculated value of C_∞ is 18.2, in excellent agreement with the experimental estimate. The experimental value of P_A is, necessarily, reproduced exactly by the model. This calculation takes into account the 2% of (1→4)-β-glucooligosaccharide sequences possessing (approximately) 9 β-(1→4) linkages; these longer sequences are neglected in the second-order Markov model reported in Fig. 2. Although only 2% of the (1→4)-β-glucooligosaccharide sequences are of this type and, hence, ⟨b⟩ is not increased significantly in comparison with Fig. 2, some 7% of all of the β-(1→4) linkages occur within these long and highly extended cellodecaosyl sequences. This we believe to be the major reason for the markedly improved agreement between theory and experiment using this Bernoullian Monte Carlo approach. This contention is supported by calculations of C_∞ using the model[6] that approximates the barley β-glucans as a Bernoullian copolymer of β-(1→3)-linked cellotriosyl and cellotetraosyl units. Our Bernoullian Monte Carlo calculations based on this model, which also neglects the extended cellodecaosyl segments, yields $9.3 \leq C_\infty \leq 11.8$ when P_A is taken to be in the experimental range 0.28–0.30.

Fig. 5 illustrates the effect of Monte Carlo sample size on the average over all conformations of a barley β-glucan chain of given sequence. Here the calculated characteristic ratio C_x is plotted for two chains of d.p. $x = 1500$ (designated arbitrarily as sequences 3 and 8) as a function of the number of different chain conformations included in the Monte Carlo sample. The *running average* of C_x is reported. It may be seen to converge to $C_x \simeq 19$ for both sequences at a sample size of about 60 chains. Similar plots (not shown) of C_x *vs.* sample size for chains of d.p. $x = 50$ converge to their limiting values for sample sizes of ~20, because the total number of conformations is smaller for shorter chains, and a smaller sample is therefore required to provide a good approximation to the parent population. In this case, the limiting C_x values obtained for the two sequences at large sample-size are different, presumably because the composition $p_{B,b}$ of these short Bernoullian chains is different. All Monte Carlo results presented here are based on samples of 60 conformations for each chain sequence, regardless of d.p.

The influence of chain length and of copolymer composition on the calculated value of C_x are shown in Fig. 6. Here the filled circles represent the mean value of C_x calculated for values of the d.p. in the range $10 \leq x \leq 1500$. It is evident that C_x

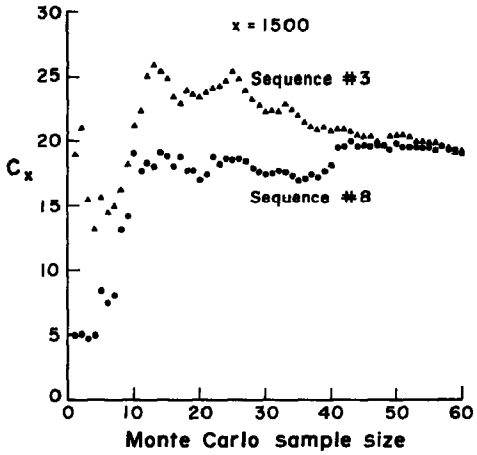

Fig. 5. A plot of the running average of C_x calculated from theory for two different copolymeric sequences of A and B linkages against the number of chain conformations in a Monte Carlo sample of the conformational population of each chain. Both chains are of d.p. $x = 1500$, and the sequences were selected to be consistent with the experimental distribution functions $p_{A,a}$ and $p_{B,b}$.

has effectively converged to C_∞ at $x = 1500$. The calculated mean values are the result of averaging over 10 sequences for each chain at d.p. $= x$; vertical bars represent the standard deviation (~1.5) at each d.p. for the sample of 10 copolymer sequences. Smooth convergence of C_x to its asymptote at large chain-length suggests that the Monte Carlo sample size used in averaging over copolymer composition is adequate.

The results of Fig. 6 disclose that the coincidence of C_x for the two sequences in Fig. 5 at large sample size is fortuitous. Very long Bernoullian chains, for which the distribution function $p_{B,b}$ must approach the experimental distribution, are

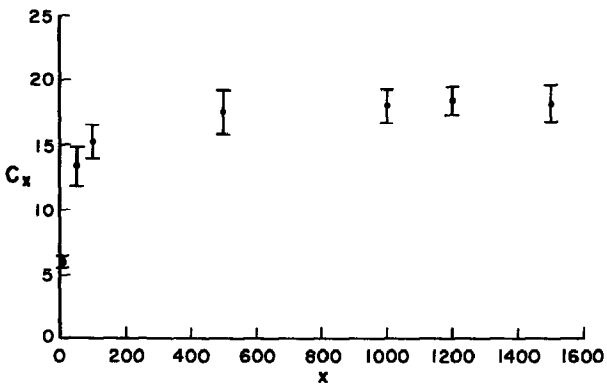

Fig. 6. The theoretical value of C_x plotted *vs.* d.p. x for a calculation involving Monte Carlo averaging over copolymer sequence-distribution and over the conformations available to each sequence. A Monte Carlo sample of 60 conformations was used to average over the conformations of each sequence. Averages over sequences were carried out with samples of 10 sequences; vertical bars represent the standard deviation of C_x for the latter samples.

expected to yield the same calculated value of C_∞ regardless of sequence, provided conformational averaging is adequate. Indeed, the conformationally averaged C_x will be the same for Bernoullian copolymers of any particular d.p., regardless of sequence, so long as the composition specified by $p_{B,b}$ is identical for all sequences. This is a direct consequence of the independent rotation approximation employed here[16,17] whereby each pair of interdependent torsion angles ϕ_i, ψ_i is assumed to be independent of all other such rotation angles in the chain. In any Monte Carlo sample of shorter chains the necessary departures of $p_{B,b}$ for the individual chains from the experimental distribution dictate a finite standard deviation for the sample. Hence, it is possible to contemplate a reduction in the standard deviation in the calculated value of C_∞, but, given the uncertainty in the present experimental estimate of C_∞, such efforts currently appear to be unwarranted.

CONCLUSIONS

These calculations demonstrate that the solution conformation of the 40° water-soluble barley β-glucan can be successfully modeled as a Bernoullian sequence of (1→4)-β-glucooligosaccharides linked together by isolated β-(1→3) linkages, if a quantitative theoretical fit to the measured unperturbed polysaccharide chain-dimensions is used as the criterion for success of the model. The proportions of the (1→4)-β-glucooligosaccharides used in the model are just those found by Woodward, Fincher, and Stone[4] (Table I), and the conformational properties of the elemental β-(1→3)- and β-(1→4)-linked D-glucose dimeric chain-segments have been modeled using methods described by Burton and Brant[25]. A second-order Markov model of the sequence statistics of β-(1→3) and β-(1→4) linkages reproduces the respective experimental distribution functions $p_{A,b}$ and $p_{B,b}$ of these linkages with reasonable accuracy. It leads to an underestimate of the polymer-chain extension in solution, as measured by the characteristic ratio C_∞ of the unperturbed dimensions, because it neglects the contributions from the small proportion of longer and highly extended (1→4)-glucooligosaccharides, namely, cellodecaosyl segments, in the chain. A model which regards the chain as a Bernoullian copolymer of β-(1→3)-linked cellotriosyl and cellotetraosyl units[6] has similar deficiencies.

A concise summary of the model-building exercise described here is provided in Figs. 7 and 8. The former figure shows computer drawn trajectories for two barley β-D-glucan chains, each frozen in a conformation typical of the countless conformations the chain will adopt as a function of time under the influence of collisions with the molecules of the solvent medium. In this rendition, only the glycosidic oxygen atoms are shown. They are connected by virtual bonds of fixed length which span the rigid sugar residues comprising the chain. Depth is conveyed by the size of the spheres representing the atoms, the smaller atoms being more remote from the viewer. The darker colored glycosidic oxygen atoms mark the positions of the β-(1→3) linkages in the chain depicted. The chains shown contain

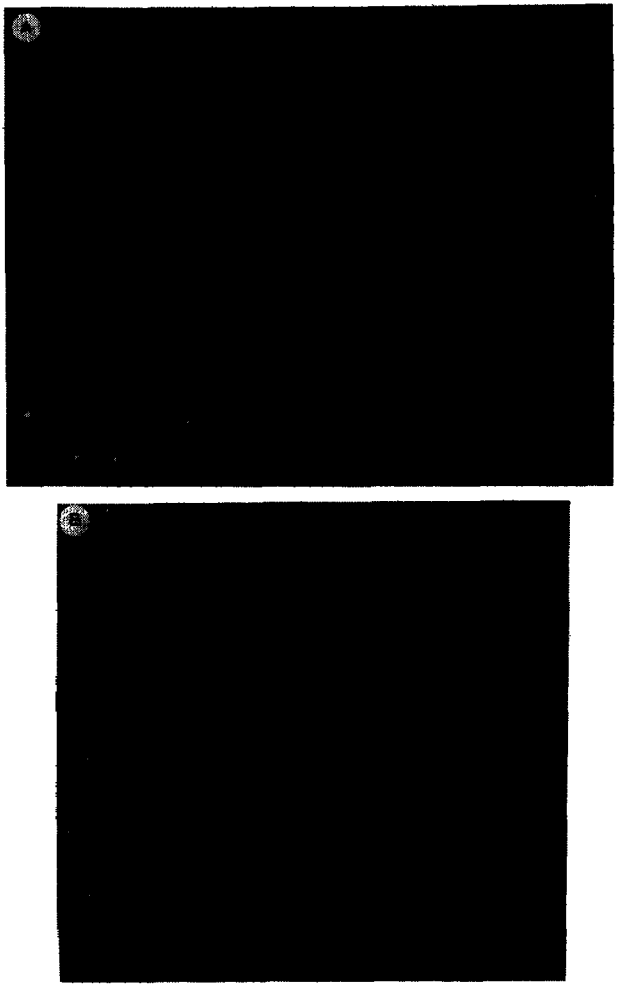

Fig. 7. (a) An instantaneous view of a barley β-glucan chain conformation chosen as typical from among the conformations in the Monte Carlo sample used to average over all chain conformations. Only the glycosidic oxygen atoms are shown; those at the β-(1→3) linkages appear as the darker spheres. Vectors (virtual bonds) spanning the sugar residues connect the glycosidic oxygens. (b) Same as (a) for a different barley β-glucan chain.

~50 residues. This seems to be long enough to disclose the essential features of the chain trajectory. Significantly shorter chains do not do this effectively, and it is difficult to produce intelligible images of chains that are much longer than this. The chains in Fig. 7 should be compared with similar drawings of the homopolymeric (1→4)-β- and (1→3)-β-D-glucans shown in Figs. 3 and 8, respectively, of the paper by Burton and Brant[25].

In Fig. 8 the same two barley β-glucan chains are presented as space-filling models. Carbon atoms are black, hydrogen atoms are white, and oxygen atoms are

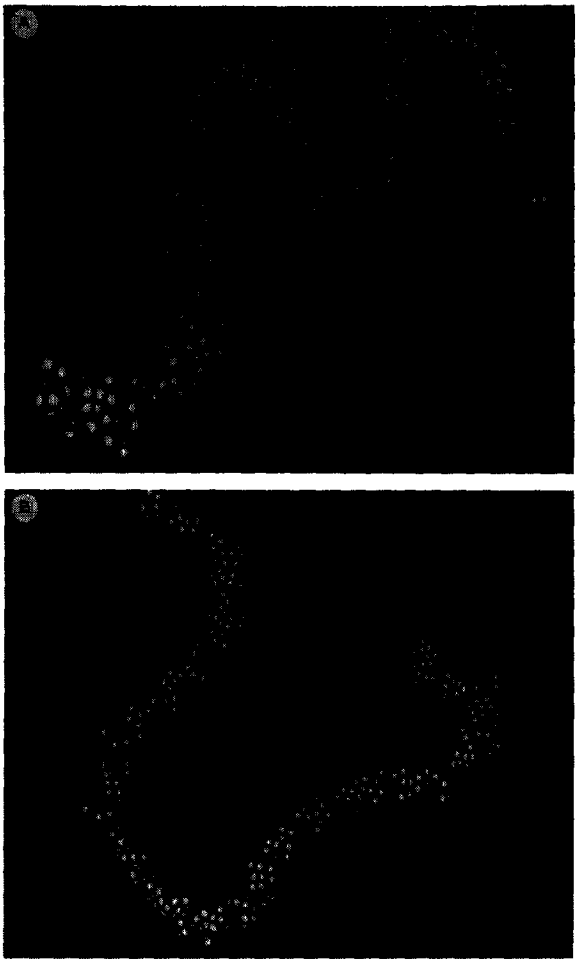

Fig. 8. (a) The same barley β-glucan chain conformation shown in Fig. 7(a), but now including all atoms in space-filling representations. Hydrogen atoms are white, carbon black, and oxygen gray. No attempt is made to identify the β-(1→3) linkages, which occur at the same positions as in Fig. 7(a). (b) Space-filling representation of the barley β-glucan chain shown in Fig. 7(b).

shown in gray; there has been no attempt to differentiate between β-(1→3) and β-(1→4) linkages, but the conformation, sequence, and orientation are identical with the chains shown, respectively, in Figs. 7a and 7b. It is apparent from the drawings that the positions along the backbone where the dissolved chain changes direction most abruptly are associated with the isolated β-(1→3) linkages, although one cannot conclude that bends are present in solution at all such linkages at all instants. The calculations indicate that consecutive β-(1→3) linkages can serve as loci of substantial chain flexibility such that barley-β-glucan fractions containing contiguous β-(1→3) linkages may be expected to display more compact solution

conformation, unless this effect is compensated by the occurrence of β-(1→4) linkages in longer blocks than are present in the 40° water-soluble fraction.

ACKNOWLEDGMENT

The authors are grateful to Mr. Todd Talashek for assistance in producing the computer drawings and to Professor W. J. Hehre for providing access to the computer graphics capabilities. This work has been supported by NIH Grant GM33062 (to DAB) and by grants from the Australian Research Grants Scheme (to GBF).

REFERENCES

1 G. B. FINCHER, *J. Inst. Brew.*, 81 (1975) 116–122.
2 J. R. WOODWARD AND G. B. FINCHER, *Brew. Dig.*, 58 (1983) 28–32.
3 J. R. WOODWARD, D. R. PHILLIPS, AND G. B. FINCHER, *Carbohydr. Polymers*, 3 (1983) 143–156.
4 J. R. WOODWARD, G. B. FINCHER, AND B. A. STONE, *Carbohydr. Polymers*, 3 (1983) 207–225.
5 F. W. PARRISH, A. S. PERLIN, AND E. T. REESE, *Can. J. Chem.*, 38 (1960) 2094–2104.
6 R. G. STAUDTE, J. R. WOODWARD, G. B. FINCHER, AND B. A. STONE, *Carbohydr. Polymers*, 3 (1983) 299–312.
7 W. W. LUCHSINGER, S.-C. CHEN, AND A. W. RICHARDS, *Arch. Biochem. Biophys.*, 112 (1965) 524–530; *ibid.*, 112 (1965) 531–536.
8 O. IGAROSHI AND Y. SAKURAI, *Agr. Biol. Chem.*, 29 (1965) 678–686.
9 M. FLEMMING AND D. J. MANNERS, *Biochem. J.*, 100 (1966) 4P–5P.
10 G. N. BATHGATE, G. H. PALMER, AND G. WILSON, *J. Inst. Brew.*, 80 (1974) 278–285.
11 M. FLEMMING AND K. KAWAKAMI, *Carbohydr. Res.*, 57 (1977) 15–23.
12 G. G. LOWRY, *Markov Chains and Monte Carlo Calculations in Polymer Science*, Marcel Dekker, Inc., New York, 1970.
13 J. HIJMANS, *Physica (Utrecht)*, 29 (1963) 1–17; *ibid.*, 29 (1963) 819–840.
14 L. PELLER, *J. Chem. Phys.*, 36 (1962) 2976–2986; *ibid.*, 43 (1965) 2355–2363.
15 G. E. HAM, *Copolymerization*, Wiley-Interscience, New York, 1964, Ch. 1.
16 D. A. BRANT AND K. D. GOEBEL, *Macromolecules*, 8 (1975) 522–530.
17 D. A. BRANT, *Q. Rev. Biophys.*, 9 (1976) 527–596.
18 D. A. BRANT, in J. PREISS (Ed.), *The Biochemistry of Plants*, Vol. 3, Academic Press, New York, 1980, pp. 425–472.
19 R. C. JORDAN AND D. A. BRANT, *Macromolecules*, 13 (1980) 491–499.
20 G. PARADOSSI AND D. A. BRANT, *Macromolecules*, 15 (1982) 874–879.
21 W. G. MILLER, D. A. BRANT, AND P. J. FLORY, *J. Mol. Biol.*, 23 (1967) 67–80.
22 S. ARNOTT AND W. E. SCOTT, *J. Chem. Soc., Perkin Trans. 2*, (1972) 324–335.
23 G. M. HALLMAN AND S. G. WHITTINGTON, *Macromolecules*, 6 (1973) 386–389.
24 R. C. JORDAN, D. A. BRANT, AND A. CESARO, *Biopolymers*, 17 (1978) 2617–2632.
25 B. A. BURTON AND D. A. BRANT, *Biopolymers*, 22 (1983) 1769–1792.
26 Y. KATO AND D. J. NEVINS, *Plant Physiol.*, 75 (1984) 745–752.
27 K. D. GOEBEL, C. E. HARVIE, AND D. A. BRANT, *Appl. Polym. Symp.*, 28 (1976) 671–691.
28 D. A. BRANT, *Carbohydr. Polymers*, 2 (1982) 232–237.
29 B. HSU, C. A. MCWHERTER, D. A. BRANT, AND W. BURCHARD, *Macromolecules*, 15 (1982) 1350–1357.