Conformational Statistics of Some Copolysaccharides

G. M. Hallman and S. G. Whittington*

Department of Chemistry, University of Toronto, Toronto 181, Ontario, Canada.

Received October 10, 1972

ABSTRACT: Unperturbed dimensions have been calculated for several polysaccharide copolymers. The effect of overall composition on the characteristic ratio has been investigated for the three polymers: glucomannan (a 1,4-linked copolymer of β-D-glucose and β-D-mannose), alginic acid (a 1,4-linked copolymer of α-L-guluronic acid and β-D-mannuronic acid) and pectic acid (considered as a copolymer of α-D-galacturonic acid and α-L-rhamnose, in which the galacturonic acid units are linked 1,4 and the rhamnose units are linked 1,2). In each case the effect of sequence distribution is also considered. For Bernoullian sequence statistics the characteristic ratio is a monotonic function of the composition for glucomannan and for pectic acid but it passes through a minimum for alginic acid. For a fixed monomer composition the characteristic ratio is sensitive to the sequence statistics of the copolymer.

In recent years a number of calculations of the unperturbed dimensions of polysaccharides have appeared in the literature. Polymers which have attracted particular attention are amylose,[1,2] cellulose,[3] poly(mannuronic acid) and poly(guluronic acid),[4] mannan,[5] and xylan.[5] Some work has also appeared on the effect of bonding geometry and local flexibility on the conformational statistics of homopolysaccharides.[6,7] Less attention has been paid to copolymers.[8,9] Many polysaccharides of biological importance contain two or more monomer residues and, although some of these can be idealised as alternating copolymers (*e.g.*, hyaluronic acid and the carageenans), others either contain one monomer in considerable excess so that the other monomer can be regarded as an impurity (*e.g.*, pectic acid) or the sequence statistics of the copolymer are more complex (*e.g.*, alginic acid). In either case it is interesting to enquire how the unperturbed dimensions of the copolymer depend on the relative amounts of the two monomers and, for a fixed monomer ratio, how they are influenced by the sequence statistics of the copolymer. For alginic acid there is good experimental evidence that samples from different sources have different monomer ratios and the sequence statistics are likely to depend on the source.[10,11] Different alginic acid samples also have very different unperturbed dimensions.[12,13] The purpose of this paper is to calculate the unperturbed dimensions of three copolysaccharides as the relative amounts of the monomers and as the monomer sequences are varied.

Methods

Calculation of Transformation Matrices. Assuming that the ring conformation of the monomer and the bridge angle is fixed, the glycoside structure is rigid, and it can be represented by a virtual bond[1] joining adjacent bridgehead oxygens. Let these virtual bonds be represented by a sequence of vectors $\{\mathbf{l}_i;\ i = 1, 2, \ldots, n\}$. The mean-square end-to-end length of such a sequence is given by

$$\begin{aligned}\langle r_n^2\rangle &= \langle(\textstyle\sum \mathbf{l}_i)\cdot(\sum \mathbf{l}_i)\rangle \\ &= \sum_{i=1}^{n} \mathbf{l}_i^2 + 2\sum_{i<j}^{n}\sum^{n}\langle \mathbf{l}_i\cdot \mathbf{l}_j\rangle\end{aligned} \qquad (1)$$

(The angular brackets denote averaging with respect to the Boltzmann distribution.)

Establishing a rectangular coordinate system on each monomer so that the x coordinate lies along the virtual bond, and defining $\mathbf{T}_i$ as the matrix which transforms from the coordinate system $(i + 1)$ to the system (i) (apart from translation), the above product is given by

$$\langle \mathbf{l}_i\cdot \mathbf{l}_j\rangle = (l_i\ 0\ 0)\langle \mathbf{T}_i\mathbf{T}_{i+1}\cdots\mathbf{T}_{j-1}\rangle\begin{pmatrix} l_j \\ 0 \\ 0\end{pmatrix} \qquad (2)$$

Assuming that under θ conditions the energy of the polymer is approximately equal to the sum of the dimer energies, then to a good approximation

$$\langle \mathbf{T}_i\mathbf{T}_{i+1}\cdots\mathbf{T}_{j-1}\rangle = \prod_{k=1}^{j-1}\langle \mathbf{T}_k\rangle \qquad (3)$$

The matrix $\langle \mathbf{T}_k\rangle$ transforms from the coordinates of $\mathbf{l}_{k+1}$ to $\mathbf{l}_k$, averaged assuming a Boltzmann distribution of the angles of rotation ϕ, ψ about the glycosidic and aglycone bonds respectively, and the angle χ, rotation about the C_5'–C_6' bond[7]

$$\langle \mathbf{T}_k\rangle = \frac{\int\int\int \mathbf{T}(\phi,\psi)\ \exp(-E_2(\phi,\psi,\chi)\beta)\mathrm{d}\phi\mathrm{d}\psi\mathrm{d}\chi}{\int\int\int \exp(-E_2(\phi,\psi,\chi)\beta)\mathrm{d}\phi\mathrm{d}\psi\mathrm{d}\chi} \qquad (4)$$

where β is the reciprocal of the product of Boltzmann's constant and the absolute temperature.

Calculation of the Characteristic Ratio. The characteristic ratio, C_∞, is given by the limit

$$C_\infty = \lim_{n\to\infty} C_n \qquad (5)$$

where

$$C_n = \langle r_n^2\rangle/nl^2 \qquad (6)$$

and l is the root-mean-square virtual bond length. From eq 1, 2, and 3 it follows that

$$C_n = 1 + (2/n)(1\ 0\ 0)\sum_{i<j}\sum\prod_{k=1}^{j-1}\langle \mathbf{T}_k\rangle\begin{pmatrix}1\\0\\0\end{pmatrix} \qquad (7)$$

For regular monomer sequences the summations can be carried out analytically, but for irregular sequences this expression is very inconvenient. However, Flory and Jerni-

---

[1] V. S. R. Rao, N. Yathindra, and P. R. Sundararajan, *Biopolymers*, **8**, 325 (1969).

[2] D. A. Brant and W. F. Dimpfl, *Macromolecules*, **3**, 655 (1970).

[3] N. Yathindra and V. S. R. Rao, *Biopolymers*, **9**, 783 (1970).

[4] S. G. Whittington, *Biopolymers*, **10**, 1481 (1971).

[5] N. Yathindra and V. S. R. Rao, *Biopolymers*, **10**, 1891 (1971).

[6] S. G. Whittington, *Macromolecules*, **4**, 569 (1971).

[7] S. G. Whittington and R. M. Glover, *Macromolecules*, **5**, 55 (1972).

[8] S. G. Whittington, *Biopolymers*, **10**, 1617 (1971).

[9] R. L. Cleland, *Biopolymers*, **10**, 1925 (1971).

[10] A. Haug, B. Larsen, and E. Baardseth, *Proc. Int. Seaweed Symp.*, **6**, 443 (1969).

[11] A. Haug, Report No. 30 of the Norwegian Institute of Seaweed Research.

[12] O. Smidsrød, *Carbohyd. Res.*, **13**, 359 (1970).

[13] O. Smidsrød and A. Haug, *Biopolymers*, **10**, 1213 (1971).

gan[14] have shown that the characteristic ratio may be obtained as

$$C_n = 1 + (2/nl^2)(1\ 0\ 0\ 0\ 0)\prod_{i=1}^{n-1}\mathbf{G}_i\begin{pmatrix}0\\1\\1\end{pmatrix} \tag{8}$$

where the **G** matrices are of the form

$$G_i = \begin{pmatrix}1 & \mathbf{l}_i'\langle\mathbf{T}_i\rangle & 0\\ \mathbf{0} & \langle\mathbf{T}_i\rangle & \mathbf{l}_i\\ 0 & \mathbf{0} & 1\end{pmatrix} \tag{9}$$

$\mathbf{l}_i'$ is the transpose of $\mathbf{l}_i$ and **0** is a null matrix of appropriate order. Using this formulation, C_n for a particular sequence of n monomers is obtained as an average over all conformations of that molecule by forming the appropriate product of **G** matrices.

Description of Program. In this work, polymers containing two different monomers were studied. Labeling these as monomers A and B, it was necessary to calculate four different average transformation matrices $\mathbf{T}_{AA}$, $\mathbf{T}_{AB}$, $\mathbf{T}_{BA}$, $\mathbf{T}_{BB}$, one for each of the four possible dimers. The **T** matrices were calculated by replacing the integrals over ϕ, ψ in eq 4 with summations over 10° intervals, and the integral over χ with a summation over 120° intervals, starting at the gt conformation. Thus the effect of rotation about the C_5'–C_6' bond has been explicitly included in the calculation of unperturbed dimensions. The dimer energies (E_2) were calculated using the Kitaygorodsky atomic pair potentials[15] with the parameters used in ref 1. The four corresponding **G** matrices were then calculated using eq 9.

The program was designed to generate random sequences of monomers according to at most second-order Markov statistics. Numerical values were assigned to the probabilities of adding a unit of A to each of the four possible terminal dimers. If a uniformly distributed random number, generated for each addition, was less than the probability corresponding to the current end state of the growing chain, the next unit was chosen to be an A; otherwise it was a B. Random numbers x_i were generated according to the formula

$$x_{i+1} \equiv 54891x_i(\text{mod } 2^{35}) \tag{10}$$

In principle the program forms the product of **G** matrices corresponding to the generated sequence of monomers, then calculates the value of C_n according to eq 8. Thus if the first monomer chosen was A and the second B, the first element in the product would be $\mathbf{G}_{AB}$. If the next monomer chosen was B, the product would be formed as $\mathbf{G}_{AB}\mathbf{G}_{BB}$, etc.

In practice it was seen to be a considerable computational saving if powers of the **G** matrices were stored, corresponding to blocks of poly(A), poly(B), and the alternating poly(A-B). Accordingly the products $(\mathbf{G}_{AA})^n$, $(\mathbf{G}_{BB})^n$, and $(\mathbf{G}_{AB}\mathbf{G}_{BA})^n$ were calculated and stored. Thus in order to form the product **G** matrices corresponding to part of a polymer containing, say, 106 units of A followed by 22 units of alternating poly(A-B), and 30 of B, only three further matrix multiplications would be needed, *viz.*: $(\mathbf{G}_{AA})^{106}\cdot(\mathbf{G}_{AB}\mathbf{G}_{BA})^{22}\cdot(\mathbf{G}_{BB})^{30}$. Using this scheme, matrix multiplication is only necessary when a block of alternating or homopolymer is terminated by the beginning of a different block. Since fewer arithmetic operations are required, rounding error is also reduced.

Powers of $\mathbf{G}_{AA}$ up to 128 in steps of one, and further powers 256, 512, and 1024 were stored. Thus a block of poly(A) up to 128 monomers long could be handled by one matrix multiplication and blocks up to 2048 by at most four multiplications. The matrices corresponding to alternating ($\mathbf{G}_{AB}\mathbf{G}_{BA}$) and poly(B) ($\mathbf{G}_{BB}$) were stored in powers up to 32 in steps of one, and to 1024 in further powers of 2. This scheme was found optimal for treating polymers in which one component (B) was present in small quantities, or at least in which the average block lengths of poly(B) and poly(A-B) were less than 32.

The program was designed to generate a given number of monomer sequences of a given length, and to print results averaged over many such sequences. Thus a typical calculation would involve averaging C_n for 500 polymer molecules containing 2000 monomers each. C_n was calculated for each molecule at given intervals, usually of 50 monomers, and C_∞ was found by linear extrapolation of C_n as a function of $(1/n)$ as $(1/n) \rightarrow 0$. This function was plotted for sample polymers and found to vary little in slope beyond about 1000 monomers.

Results

The sensitivity of the characteristic ratio to the overall composition and sequence distribution has been investigated for three copolymers: (i) glucomannan, in which both monomers are β, D, and *C1* and all glycosidic linkages are 1,4-; (ii) alginic acid, a copolymer of β-D-mannuronic acid (*C1*) and α-L-guluronic acid (*1C*) with all glycosidic linkages 1,4-; (iii) pectic acid, a copolymer of α-D-galacturonic acid (*C1*) and α-L-rhamnose (*1C*) in which the galacturonic acid units are linked 1,4- but the rhamnose units are linked 1,2-. (This is probably a serious oversimplification of the situation in pectic acid where other monomers may appear in the "backbone" and the rhamnose units may in fact be β.)

In order to enquire if any general conclusions can be drawn on these questions it is convenient to consider the three polymers together and to consider the different types of sequence statistics separately.

Bernoullian Statistics. In describing each type of statistics employed it is convenient to think of the polymer growing from the nonreducing towards the reducing end. In Bernoullian statistics the probability that the "next" monomer is A (say) is independent of the "last" monomer. Such sequence statistics are particularly simple and serve as a convenient baseline against which to compare more complex statistics, and this system has been used to investigate the gross features of dependence of characteristic ratio on overall composition of the polymer.

For each polymer we have calculated the characteristic ratio for several different compositions and the results are shown in Figure 1. For glucomannan and pectic acid the characteristic ratio is monotonic, decreasing as the proportion of mannose or rhamnose in the polymer is increased, and the relative insensitivity of characteristic ratio to composition for glucomannan presumably simply reflects the similarity of the characteristic ratios of the homopolymers. The result for pectic acid is interesting in that the characteristic ratio decreases very rapidly when the proportion of rhamnose is small, suggesting that, in some cases, the characteristic ratio can be extremely sensitive to small amounts of impurity. In this case, the characteristic ratio is decreased by a factor of two when the impurity is a little below the 10% level.

For alginic acid the situation is different in that the characteristic ratio–composition curve passes through a minimum at about 50%. For comparison we have calculated the characteristic ratio for the alternating copoly-

---

(14) P. J. Flory and R. L. Jernigan, *J. Chem. Phys.*, 42, 3509 (1965).

(15) A. I. Kitaygorodsky, *Tetrahedron*, 14, 230 (1961).

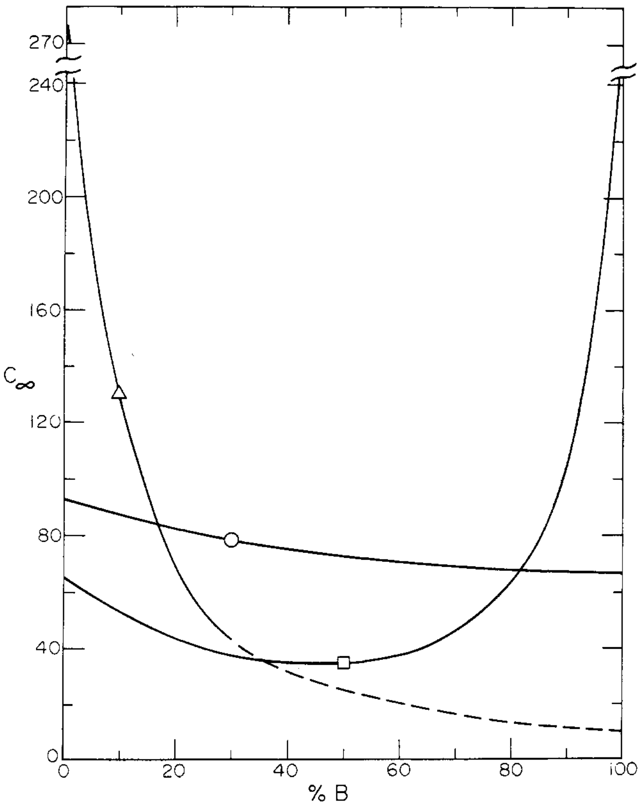

Figure 1. Dependence of characteristic ratio (C_∞) on composition for copolysaccharides with Bernoullian sequence statistics: (△) pectic acid, B = α-L-rhamnose; (○) glucomannan, B = mannose; (□) alginic acid, B = α-L-guluronic acid.

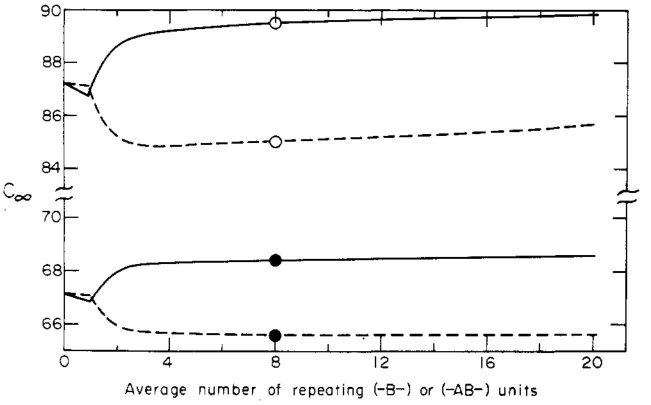

Figure 2. Dependence of characteristic ratio (C_∞) on average block length of impurity for glucomannan: (●) 10% glucose; (○) 10% mannose. Full line is first order Markov statistics and broken line is strictly alternating model. The points on the ordinate are the Bernoullian values.

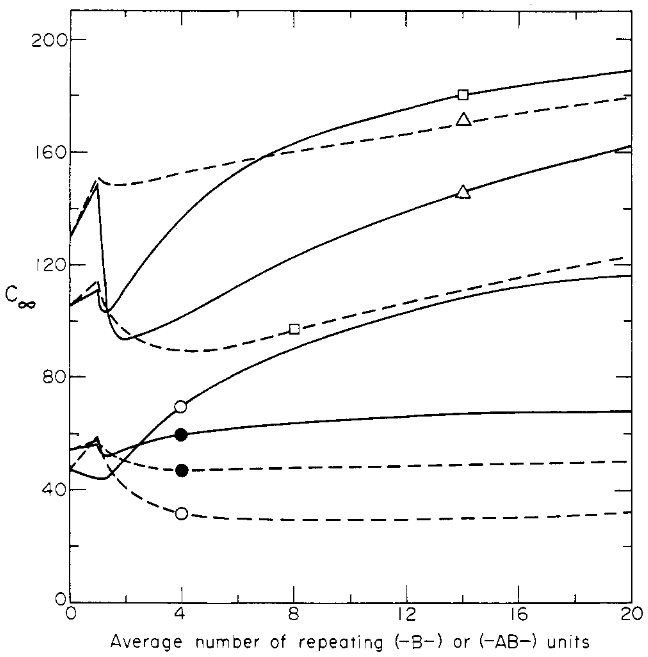

Figure 3. Dependence of characteristic ratio (C_∞) on average block length of impurity for: (△) 10% rhamnose in pectic acid; (□) 10% mannuronic acid in alginic acid; (○) 30% mannuronic acid in alginic acid; (●) 10% guluronic acid in alginic acid. Full line is first order Markov model and broken line is strictly alternating model. The points on the ordinate are the Bernoullian values.

mer, yielding a value of about 12, compared to the 50% Bernoullian value of about 35.

At least for homopolymers, it has been shown[6,7] that the characteristic ratio reflects both bonding geometry and local flexibility of the polymer (as measured, for instance, by the contribution, from rotation about the glycosidic and aglycone bonds, to the conformational partition function of the dimer). To some extent the above results can be rationalized in a similar way. For glucomannan the bonding geometry is constant and the linkage in the dimer is flexible if the nonreducing unit is mannose, but stiff if it is glucose. As the proportion of mannose increases the number of linkages with mannose as the nonreducing sugar increases, the overall flexibility increases and hence the characteristic ratio decreases. For pectic acid, linkages involving rhamnose are rather stiff but it has been pointed out by Rees and Wight[16] that insertion of a rhamnose unit in a galacturonic acid polymer produces a kink in the chain which causes a decrease in the characteristic ratio.

For alginic acid, the linkages increase in flexibility in the order gul–man, gul–gul, man–man, man–gul,[4,8,17] so that the proportions of the most flexible and of the least flexible linkages are largest near 50% composition. It may be that the most flexible linkage dominates the situation though one would not predict this from the contributions to the appropriate dimer partition functions. A second effect is that poly(guluronic acid) is 1a–4a linked, poly(mannuronic acid) is 1e–4e linked while, near 50%, there will be a considerable amount of 1e–4a and 1a–4e linkages. It is well known that amylose (with 1a–4e linkages) has a low characteristic ratio (~7) and calculations on β-galactan suggest that its characteristic ratio is much lower than, say, α-galactan.[7] It is therefore not clear whether increased flexibility or change in bonding geometry is responsible for the minimum at about 50% and, indeed, both effects may be important.

First-Order Markov Statistics. In this case the probability that the "next" monomer is A depends on the nature of the preceeding monomer. The transition probabilities were chosen to favor homopolymeric blocks, *i.e.*, p(A/A) is large and p(A/B) is small. The sequence statistics are governed by these two parameters so that, for fixed composition, one parameter is still available. In fact we chose p(A/B) and p(A/A) to vary the average length of the homopolymeric blocks, while keeping the composition fixed. We have investigated the dependence of characteristic ratio on average block length for glucomannan with 10% and 90% mannose, for pectic acid with 10% rhamnose and for alginic acid with 10, 30, and 90% mannuronic

---

(16) D. A. Rees and A. W. Wight, *J. Chem. Soc. B*, 469 (1971).

(17) O. Smidsrød, R. M. Glover, and S. G. Whittington, *Carbohyd. Res.*, in press.

acid. The results are shown by the continuous lines in Figures 2 and 3.

In each case, after some initial fluctuations, the characteristic ratio increases with block length and then levels off to a value above the Bernoullian value. The dependence on block length is certainly more striking in some cases than in others and there appears to be a close correlation between this dependence and the gradient, at the appropriate composition, of the characteristic ratio–composition curve. That is, the sensitivity to block length is related to the sensitivity to small changes in overall composition. This is well illustrated by the 30 and 90% mannuronic acid cases in alginic acid. Although the Bernoullian characteristic ratios are similar (47 and 54) the characteristic ratio–composition curve is changing more rapidly at 30% mannuronic acid than at 90% mannuronic acid and, with a block length of 20, the characteristic ratio has risen to 116 in the former case but to only 68 in the latter. The curious minimum observed in each case between an average block length of one and two is more difficult to rationalize. This effect is most pronounced in the pectic acid case and suggests that *dimers* of rhamnose, as an impurity in polygacturonan, are most effective at decreasing the characteristic ratio.

Strictly Alternating Impurities. For alginic acid, the polymer is known to contain blocks of alternating copolymer as well as homopolymeric blocks of each monomer.[18] Alternating blocks have also been suggested to occur in pectic acid.[19] For this reason we have investigated the situation in which the less abundant monomer appears *only* in alternating blocks. The particular model we have used is one with second order Markov statistics. We write $p(A/AB)$ for the probability that the next monomer is A given that the preceeding monomer is B and that the penultimate monomer is A. If B is the less abundant monomer, the model is characterised by $p(B/AA) = \delta$, $p(B/AB) = 0$, $p(B/BA) = 1 - \epsilon$ and $p(B/BB) = 0$. It is easy to show that the percentage of A in an infinite copolymer is $100\delta/(\epsilon + 2\delta)$ and that the average number of B monomers in an alternating block is δ^{-1}. We shall refer to $1/\epsilon$, the average number of repeating AB units in the alternating block, as the "block length," though one might regard $2/\epsilon$ as more correctly defining this quantity. We have investigated the dependence of characteristic ratio on block length for each of the six cases previously mentioned. The results are shown as the broken curves in Figures 2 and 3. In each case the characteristic ratio first decreases and then gradually increases as the block length increases and, for long blocks, the characteristic ratio for the alternating case is usually below that for the first-order Markov case. In general terms, the dependence of characteristic ratio on block length is rather less marked with alternating impurity than in the first-order Markov case.

Conclusions

The general result of this work is that the characteristic ratio of a copolysaccharide is sensitive both to overall composition and to sequence distribution but that the sensitivity is more marked in some cases than in others. Some simplifying generalisations have emerged but it is likely that, to obtain reliable predictions, individual cases must be treated on their own merits and it may be unwise to extrapolate the present results to copolymers having structures very different from the three cases discussed here.

For a copolysaccharide such as alginic acid it is fairly straightforward to determine experimentally the overall composition, *i.e.*, the mannuronic to guluronic acid ratio. On the other hand, the characterization of the sequence statistics is an extraordinarily difficult experimental problem. The amount of work involved in characterizing a single sample (*e.g.*, that isolated from *Laminaria digitata*) is enormous[20] and new approaches to this problem are urgently needed.

The problem of reproducing experimental characteristic ratios of well-characterized polysaccharides by calculations of the conformational statistics is still far from being solved. However, when we are in a position to predict characteristic ratios quantitatively for well characterized systems, the present results suggest that, in favorable cases, the use of calculated and experimental characteristic ratios may be a useful probe of the sequence statistics of copolysaccharides.

The quantitative aspects of the results reported here should be treated with some caution. There are a number of features (such as the constancy of the bridge angle and of the ring conformation, as well as the choice of potential function) which are important oversimplifications of the real situation, and these simplifications may have a considerable effect on the quantitative results obtained, although it seems unlikely that the qualitative conclusions will be severely affected.

Acknowledgments. The authors appreciate many helpful discussions with R. M. Glover, and the financial assistance of the National Research Council of Canada.

---

(18) B. Larsen, T. Painter, A. Haug, and O. Smidsrød, *Acta Chem. Scand.*, 23, 355 (1969).

(19) G. O. Aspinall, K. Hunt, and I. M. Morrison, *J. Chem. Soc. C*, 1080 (1967).

(20) A. Haug, S. Myklestad, B. Larsen, and O. Smidsrød, *Acta Chem. Scand.*, 21, 768 (1967).