A General Treatment of the Configurational Statistics of Polysaccharides[1]

David A. Brant*[2] and Kenneth D. Goebel

Department of Chemistry, University of California, Irvine, California 92664.
Received September 24, 1974

ABSTRACT: A treatment of the configurational statistics of polysaccharides is given in the isomeric state approximation. All classes of linear polysaccharides of specified chemical sequence are treated simultaneously. Chain tortuosity arising from torsional motions about the chemical bonds of the glycosidic linkages is recognized explicitly as is the possibility for conformational isomerism of the sugar residues. Valence angles and lengths are taken to be fixed at the equilibrium values, and pyranose residues in their chair conformations are treated as inflexible constituents of the skeletal structure. Pyranose and furanose forms capable of pseudorotation may be incorporated as rigid skeletal entities as well, provided suitable attention is given to the selection and interpretation of the conformational isomeric states included. Separation of the configuration energy into independent contributions is shown to be impossible in general. Methods are described for assessing the influence of neighbor interactions on the populations of the several conformers of the sugar residues. The relative conformational free energy of the flexible and chair form conformers of pyranose sugars is discussed, and appropriate measures of polysaccharide chain flexibility and stiffness are suggested.

Attempts to relate the random coil chain dimensions of polysaccharides to the geometric details of the polymer skeletal structure were carried out as early as 1948 by Benoit,[3] who treated specifically the cellulose chain in the approximation of fixed valence lengths and angles and inflexible sugar residues while assuming both free and independent hindered rotations about the bonds of the glycosidic linkages connecting adjacent residues. Eliezer and Hayman subsequently extended the treatment of Benoit to amylose and pectic acid.[4] Similar considerations of a variety of other polysaccharides at the same levels of approximation have been put forth more recently by Burchard,[5] by Cleland,[6] and by Yathindra and Rao.[7] The inadequacy of the free rotation approximation was recognized in the earliest work, and steric hindrances to rotation about the bonds of the glycosidic bridge were acknowledged through the introduction of simple torsional hindrance potentials.[3-5] However, the uncomplicated functional forms employed in these treatments to render the development analytically tractable bore little relation to the true rotational hindrance potentials, the varied and irregular forms of which are highly dependent upon the details of the geometric and electronic structure of the polymer chain. Moreover, these treatments failed to account for the interdependence of the torsional hindrance potentials for rotations about the two bonds of a given glycosidic linkage, which is evident from a cursory inspection of space-filling molecular models, much less for interdependences which may be operative at longer range in the chain sequence.

Recently methods have become available which permit the calculation of configuration-dependent properties of polymer chains based on chain models which can incorporate more realistic approximations to the conformational potential energy surface for representative skeletal segments.[8] The statistical mechanical methods and theoretical foundations have been reviewed in detail by Flory.[9] Within the context of these newer methods several treatments of the configurational statistics of polysaccharides have appeared.[10-13] In each case cited the interdependence of potentials for rotations about the two C–O bonds of the glycosidic bridge has been taken into account in conjunction with a skeletal model which attributes all of the tortuosity of the chain to rotations about these two bonds. Constancy of valence lengths and angles is also assumed, and sugar residues are restricted to a single rigid ring conformation. It has furthermore been assumed, sometimes without critical consideration of the validity, that rotations at one glycosidic bridge, although mutually interdependent, are independent of rotations at all other such sites in the chain. This approximation is probably justified only in circumstances under which intramolecular hydrogen bonding of the polysaccharide is suppressed.[13] Treatments based on more elaborate chain models have also been presented.[14] These acknowledge changes in sugar ring conformation as a possible source of chain tortuosity. Dunfield and Whittington[14b] have assumed Bernoullian sequence statistics for the distribution of alternate ring conformers in the chain, whereas Brant and Goebel[14a] employ a statistical mechanical method which accounts for the influence of neighbor interactions in determining the sequence statistics. The latter method is developed in detail in what follows.

The usual approximation of rigidity in all structural parameters save the glycosidic torsions requires critical examination. It has commonly been argued that the force constants associated with distortions of valence lengths and angles from their equilibrium values are much larger than those for torsional strain involving internal rotations about the skeletal bonds with the consequence that distortions of the former types are of relatively small thermal amplitude and, hence, have a negligible effect upon the molecular conformation.[15] Pseudorotation in furanose residues[16] clearly may render inappropriate the assumption of rigid ring geometry for such sugars. The strong dependence of polynucleotide chain dimensions on ribose ring geometry has already been demonstrated.[17] Likewise, the importance of flexibility in the five-membered pyrollidine ring for the configuration of poly(L-proline) and poly(γ-hydroxy-L-proline) has been investigated.[18] Pyranose sugars on the other hand normally assume the so-called "chair" geometry as the lowest energy conformation, and this form is incapable of pseudorotation.[19,20] One must, however, anticipate the existence of vibrational modes of the pyranoid chair which distort the ring through coupled displacements of endocyclic valence and torsion angles from their equilibrium values. Indeed, analysis of the infrared and Raman spectra of the skeletal pyranoid tetrahydropyran[21] reveals the existence of several such ring distortion modes occurring at frequencies below 500 cm^{-1} for which relatively large amplitudes must be expected owing to an appreciable contribution to these modes from the "soft" torsional displacements.[15]

Six-membered rings are also capable of adopting confor-

mations of lower symmetry than the chair form known commonly as the "boat" and the "twist" or "skew" forms.[19,20] These conformers, which occur at or near minima on the conformational energy surface for the molecule, may be interconverted continuously by pseudorotation and collectively constitute what may be termed the "flexible" forms.[16,21] The conformational energy profile along the pseudorotation itinerary for pyranose sugars is unknown, although for tetrahydropyran the pseudorotation coordinate for the flexible forms apparently carries the molecule through four equivalent minimum-energy conformations separated by two nonequivalent pairs of higher energy conformations which possess energies about 0.2 and 1.6 kcal/mol, respectively, greater than the conformations of minimum energy.[21] These same investigations[21] suggest that the least energetic flexible forms lie some 6.5 kcal/mol above the chair conformer. A similar energy surface has been described for pseudorotation of the skeletal furanose, tetrahydrofuran; however, in this case there exists no analog of the rigid chair conformations.[21] Although the energy surfaces for various pyranose sugars must surely differ dramatically from that of tetrahydropyran, the results quoted for the latter molecule do illustrate that the several flexible conformers of pyranose sugars may in some cases be interconverted by surmounting only relatively small energy barriers. The possible impropriety of treating pyranose sugars in their flexible ring conformations as rigid constituents of the polysaccharide skeleton is thus quite obvious. The flexible forms are, however, invariably less stable than one or the other chair form counterpart, and there is no evidence for the occurrence of substantial proportions of flexible form sugar residues in any polysaccharide.[16] This circumstance notwithstanding, to ignore altogether the possible existence in polysaccharides of occasional flexible form residues may be to overlook a most important if, indeed, infrequent, locus of chain flexibility.[14a] In general, introduction into a polysaccharide chain of a sugar ring conformer alternate to the preponderant ring form produces a large change in the direction of chain propagation and hence constitutes a potential source of chain tortuosity.[14]

It is evident that several sources of polysaccharide chain flexibility in addition to torsions about the bonds of the glycosidic linkage may make a substantial contribution to the tortuosity of the chain. Efforts to describe rigid, e.g., helical or folded, conformations of minimum energy may be seriously compromised by failure to recognize the important sources of flexibility already cited as well as other sources such as the valence angle at the oxygen atom of the glycosidic bridge.[13a] For calculations of the dimensions of randomly coiling polysaccharide chains on the other hand it should suffice to adopt fixed bond lengths and angles which may be assigned appropriate equilibrium values based on structural data from analogous small molecules.[15] The high-frequency modes associated with changes in these coordinates should be nearly independent of the conformation and environment of the chain,[9b] and thermal fluctuations about the mean values will normally be in the harmonic range so that positive and negative fluctuations may be presumed to have nearly compensating influences upon the chain configuration.[9a] It is in this sense that the fixed glycosidic valence angle adopted in all treatments to date[10–14] of the chain configuration of polysaccharides should be understood. Pyranose rings in their chair conformations may likewise be treated as rigid structural units for the purpose of calculating configuration-dependent properties of random chains. The observed ring deformation frequencies for tetrahydropyran[21] all exceed kT at room temperature, so that the ring distortions, although

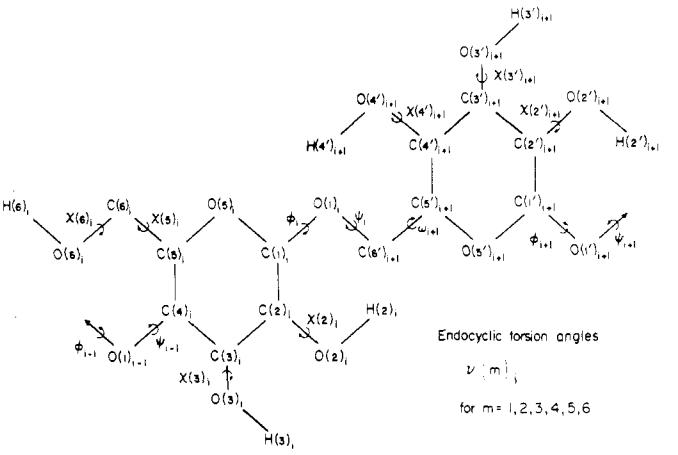

Figure 1. A representative segment of a polyhexopyranose chain illustrating 1,4 and 1,6 glycosidic linkages.

conceivably of considerable amplitude, will be approximately harmonic thereby ensuring that the deformations of opposing sense will cancel approximately in their effect upon the chain configuration. Sugar conformers capable of pseudorotation may also be regarded as rigid entities in treating the configurational statistics, provided suitable attention is given to the selection and interpretation of the conformational isomeric states included. This aspect of the treatment will be elaborated in the Discussion.

In this paper we develop a treatment of the configurational statistics of polysaccharides which takes into account the possibility for conformational isomerism of the sugar residues. Bond angles and bond lengths are assumed to be fixed at the equilibrium values, and pyranose residues in their chair conformations are treated as inflexible units in keeping with the rationale supplied above; pyranose and furanose forms capable of pseudorotation may be incorporated under the same prescription subject to a judicious choice of conformational isomeric states. The usual approximation[10–13] of independence of groups of glycosidic torsions is retained for chains possessing a fixed sequence of sugar ring conformers. However, separation of the configuration energy will be shown to be impossible when alternative conformations of a given sugar residue are allowed.[13,14a] The treatment is otherwise general, within the confines of the isomeric state approximation,[9] and all classes of linear polysaccharides of specified chemical sequence are treated simultaneously. Random and Markoffian copolymers may be treated using the present results by employing well-known Monte Carlo methods.[9a]

Geometry and Notation[22]

Consider a linear polysaccharide chain containing a specified chemical sequence[23] of sugar residues. A representative dimeric segment of this chain is shown in Figure 1. Atomic symbols, e.g., $C(m)_i$, are indexed in parentheses to denote position in the sugar ring; a serial index which identifies sequential position of sugar residues appears as a subscript on the atomic symbols and increases in the direction of the reducing end of the chain. Hydrogens $H(m)_i$ bonded to carbons $C(m)_i$ are omitted from Figure 1. Primes are used when appropriate, e.g., when serial subscripts are omitted, to denote atoms of that residue of a dimeric segment which is closer to the reducing end. For generality of illustration the segment depicted comprises hexopyranose residues i and $i + 1$ connected in 1,6 glycosidic linkage. Connection to residue $i - 1$ is taken in Figure 1 to proceed

through a 1,4 linkage with residue i.

Let the conformation of hexopyranose ring i be described by a *set* of coordinates ξ_i, for example, a set of valence lengths, valence angles, and torsion angles which describe the relative positions of the six endocyclic atoms plus any exocyclic atoms bonded directly to the sugar ring. It is convenient to assign specific symbols to the torsion angles of the linkage region as follows:[24]

$$\begin{aligned} \varphi_i &= \theta[\mathrm{H}(1)_i, \mathrm{C}(1)_i, \mathrm{O}(1)_i, \mathrm{C}(m)_{i+1}] \\ \psi_i &= \theta[\mathrm{C}(1)_i, \mathrm{O}(1)_i, \mathrm{C}(m)_{i+1}, \mathrm{H}(m)_{i+1}] \\ \omega_{i+1} &= \theta[\mathrm{O}(1)_i, \mathrm{C}(6)_{i+1}, \mathrm{C}(5)_{i+1}, \mathrm{H}(5)_{i+1}] \end{aligned} \quad (1)$$

The index m in eq 1 adopts the value appropriate to 1,m glycosidic linkages, e.g., $m = 6$ for the 1,6 linkage depicted in Figure 1. The torsion angle ω_{i+1} is defined only for 1,6 linkages. To render the definition of ψ_i unambiguous for 1,6 linkages, the hydrogen, designated $\mathrm{H}^1(6)_{i+1}$, used to define ψ_i is that one of the two hydrogens bonded to $\mathrm{C}(6)_{i+1}$ which lies cis to $\mathrm{O}(5)_{i+1}$ when $\omega_{i+1} = 0°$. For purposes of generality the development to follow treats explicitly the case of 1,6 linked hexopyranose units. Its ready adaptation to the occurrence of other linkages and other types of sugar residues will be noted.

In the succeeding treatment the set of variables $\{\varphi,\psi,\omega,\xi\}$ are regarded as the configuration variables of the chain. That is, the multitude of allowed chain conformations arises as a consequence of changes in these coordinates, all other coordinates such as the bond lengths and bond angles in the linkage region being regarded as fixed at their equilibrium values. The exocyclic torsion angles $\chi(m)_i$[24] will be ignored in the present development on the assumption that the pendant atoms can be treated as spherically symmetrical *groups*, e.g., $\mathrm{OH}(m)_i$, whose positions relative to the sugar ring are defined by valence angles and lengths of the set ξ_i which locate the exocyclic atoms bonded directly to the ring. The positions of the exocyclic hydrogen atoms $\mathrm{H}(m)_i$ are likewise specified as fixed for each ring conformation defined by ξ_i. Treatment of substituents on the polymer backbone in the structureless group approximation effectively reduces the dimensionality of the problem without serious sacrifice of physical reality.[9,15] The approximation is by no means necessary[11,12] and would be difficult to justify in circumstances where intramolecular hydrogen bonding was important. We consider the configurational statistics intractable for systems characterized by such interactions however,[13] and the treatment presented applies only when intramolecular hydrogen bonding is suppressed, for example, by ether or ester substitution of the hydroxyls or by confining attention to strongly-solvating, hydrogen-bonding solvents such as water.

In the approximation specified the distance $\mathrm{O}(1)_{i-1} \cdots \mathrm{O}(1)_i$ in Figure 1 is fixed for a particular conformation of sugar ring i, i.e., for fixed ξ_i. The corresponding "virtual" bond vector may be labeled $\mathbf{l}_i$. Let the chemical bond vector $\mathrm{O}(1)_i$–$\mathrm{C}(6)_{i+1}$ be called $\mathbf{b}_i$. The residue vector is then defined as $\mathbf{L}_i = \mathbf{l}_i + \mathbf{b}_i$. In similar fashion the distance $\mathrm{C}(6)_{i+1} \cdots \mathrm{O}(1)_{i+1}$ is recognized as the length of virtual bond $\mathbf{l}_{i+1}$. The residue vector $\mathbf{L}_{i+1}$ will be given by $\mathbf{L}_{i+1} = \mathbf{l}_{i+1} + \mathbf{b}_{i+1}$ or $\mathbf{L}_{i+1} = \mathbf{l}_{i+1}$ depending on whether the connection to residue $i + 2$ is through a 1,6 or 1,m (for $m = 2$, 3, 4) linkage, respectively.

Calculation of the Unperturbed Chain Dimensions

The chain vector $\mathbf{r}$ for the polysaccharide chain illustrated in Figure 1 is given in terms of the above definitions by

$$\mathbf{r} = \sum_{0<i\leq x} \mathbf{L}_i$$

where x is the degree of polymerization of the chain. The vector $\mathbf{r}$ connects $\mathrm{C}(6)_1$ and $\mathrm{O}(1)_x$; contributions to the chain vector from the preceding bonds $\mathrm{H}(6)_1$–$\mathrm{O}(6)_1$ and $\mathrm{O}(6)_1$–$\mathrm{C}(6)_1$ and the succeeding bond $\mathrm{O}(1)_x$–$\mathrm{H}(1)_x$ are unimportant for all cases of interest and are ignored for convenience.[25] The mean-square end-to-end distance follows as usual[9a] from the scalar product $\mathbf{r}\cdot\mathbf{r}$.

$$\langle r^2\rangle_0 = \sum_{0<i\leq x} \langle \mathbf{L}_i \cdot \mathbf{L}_i\rangle + 2 \sum_{0<i<j\leq x} \langle \mathbf{L}_i \cdot \mathbf{L}_j\rangle \quad (2)$$

Here angle brackets imply an average over all conformations of the chain, and the subscript zero connotes neglect of long range volume exclusion.[9a] In contrast to earlier treatments[10-13] the diagonal terms require averaging not only because of the variable orientation of $\mathbf{b}_k$ with respect to $\mathbf{l}_k$ when residues k and $k + 1$ are conjoined in 1,6 linkage but also because the coordinates ξ_k are among the configuration variables to be considered. The latter factor persists in the treatment of chains containing no 1,6 linkages for which no vectors $\mathbf{b}_k$ contribute to the chain vector and the torsion angle ω_k does not appear as a configuration variable. Application of the present treatment to these simpler chains will be described but will not require separate exposition here.

Let there be established for each residue k of the chain a coordinate system with reference to which the residue vector $\mathbf{L}_k = \mathbf{L}_k(\varphi_k,\xi_k)$ may be expressed. The orthogonal matrix $\hat{\mathbf{T}}_k = \hat{\mathbf{T}}_k(\varphi_k,\psi_k,\omega_{k+1},\xi_k,\xi_{k+1})$ is defined to transform to coordinate system k any vector $\mathbf{v}^{(k+1)}$ referred originally to basis system $k + 1$: $\mathbf{v}^{(k)} = \hat{\mathbf{T}}_k\mathbf{v}^{(k+1)}$. (The dependences of $\mathbf{L}_k$ on φ_k and of $\mathbf{T}_k$ on ω_{k+1} occur only when residue k is the residue of lower serial index for a pair of residues connected in 1,6 linkage and may otherwise be suppressed. These dependences are carried here explicitly for generality.) The scalar products $\mathbf{L}_i\cdot\mathbf{L}_i$ and $\mathbf{L}_i\cdot\mathbf{L}_j$ required in eq 2 may then be expressed by

$$\mathbf{L}_i \cdot \mathbf{L}_i = \mathbf{L}_i^{\mathrm{T}}\mathbf{L}_i \quad (3a)$$

$$\mathbf{L}_i \cdot \mathbf{L}_j = \mathbf{L}_i^{\mathrm{T}}\left(\prod_{i\leq k<j} \hat{\mathbf{T}}_k\right)\mathbf{L}_j \quad (3b)$$

where the vectors on the right-hand side are understood to be written in matrix notation.[9a]

Averages called for in eq 2 over all conformations of the chain require specification of the configuration energy $V\{\varphi,\psi,\omega,\xi\}$.[9a,26] This quantity, which we have chosen to depend explicitly on the configuration variables $\{\varphi,\psi,\omega,\xi\}$ must be understood as a *potential of mean force* subject to prior averaging over the configuration space of the molecules of the medium and the several "hard" configuration variables of the macromolecule which have been fixed for convenience at their equilibrium values.[9b] It has therefore the character of a free energy.[15] As such it cannot be anticipated that $V\{\varphi,\psi,\omega,\xi\}$ can be specified ab initio but rather that a judiciously chosen functional form can be satisfactorily parameterized by appeal to appropriate experimental information.[15] Let the conformational energy V_k of a pair of residues k and $k + 1$ be expressed for $0 < k < x$ by[26]

$$\begin{aligned} V_k &= V_k(\varphi_k, \psi_k, \omega_{k+1}, \xi_k, \xi_{k+1}) \\ &= V_k^{\mathrm{s}}(\xi_k) + V_k^{\mathrm{m}}(\varphi_k, \psi_k, \omega_{k+1}, \xi_k, \xi_{k+1}) \end{aligned} \quad (4)$$

where V_k^{s} is the self energy of residue k and V_k^{m} is the mutual energy of residues k and $k + 1$; for $k = x$ the mutual energy vanishes and $V_x = V_x^{\mathrm{s}}(\xi_x)$. The assumed dependences of V_k^{s} and V_k^{m} upon the configuration variables are shown explicitly in eq 4. The self energy of a residue is taken to depend solely upon the ring conformation ξ_k of

that residue; the mutual energy of a pair of residues depends upon their respective ring conformations ξ_k, ξ_{k+1} and upon the mutual orientations of the residues as governed by the torsion angles φ_k, ψ_k, ω_{k+1} of the glycosidic linkage. In terms of the conformational energies V_k the configuration energy of the chain may be written

$$V\{\varphi, \psi, \omega, \xi\} = \sum_{0<k\leq x} V_k \tag{5}$$

The averaged scalar products required in eq 2 may be written formally using the expressions in eq 3 as

$$\langle \mathbf{L}_i \cdot \mathbf{L}_i \rangle = \frac{\sum_{\{\varphi,\psi,\omega,\xi\}} (\mathbf{L}_i^{\mathrm{T}}\mathbf{L}_i) \exp(-V\{\varphi,\psi,\omega,\xi\}/RT)\Delta\{\varphi,\psi,\omega,\xi\}}{\sum_{\{\varphi,\psi,\omega,\xi\}} \exp(-V\{\varphi,\psi,\omega,\xi\}/RT)\Delta\{\varphi,\psi,\omega,\xi\}} \tag{6a}$$

$$\langle \mathbf{L}_i \cdot \mathbf{L}_j \rangle = \frac{\sum_{\{\varphi,\psi,\omega,\xi\}} \left[\mathbf{L}_i^{\mathrm{T}} \left(\prod_{i\leq k<j} \hat{\mathbf{T}}_k\right) \mathbf{L}_j\right] \exp(-V\{\varphi,\psi,\omega,\xi\}/RT)\Delta\{\varphi,\psi,\omega,\xi\}}{\sum_{\{\varphi,\psi,\omega,\xi\}} \exp(-V\{\varphi,\psi,\omega,\xi\}/RT)\Delta\{\varphi,\psi,\omega,\xi\}} \tag{6b}$$

Sums over all values of the configuration variables appear in place of integrals in consonance with the isomeric state approximation upon which the succeeding treatment depends.[9] The quantity $\Delta\{\varphi,\psi,\omega,\xi\}$ represents the volume in configuration space to be associated with each isomeric state included in the summation. For a given $\{\xi\}$, i.e., for a given stereochemical sequence of the chain, the summations over $\{\varphi,\psi,\omega\}$ may be written as a product of summations, if, as assumed in writing eq 4, V_k depends only on those variables φ_k, ψ_k, and ω_{k+1} from the set $\{\varphi,\psi,\omega\}$ which correspond to torsions associated with a single glycosidic linkage, namely, that between residues k and $k + 1$. After appropriate grouping of factors with identical index we find

$$\langle \mathbf{L}_i \cdot \mathbf{L}_i \rangle = \frac{\sum_{\{\xi\}} \left(\prod_{0<k<i} u_{\xi\eta;k}\right)(l_{\xi\eta;i}^{\,2})\left(\prod_{i<k\leq x} u_{\xi\eta;k}\right)}{\sum_{\{\xi\}} \left(\prod_{0<k\leq x} u_{\xi\eta;k}\right)} \tag{7a}$$

$$\langle \mathbf{L}_i \cdot \mathbf{L}_j \rangle = \frac{\sum_{\{\xi\}} \left(\prod_{0<k<i} u_{\xi\eta;k}\right)(\mathbf{l}_{\xi\eta;i}^{\mathrm{T}})\left(\prod_{i<k<j} \mathbf{t}_{\xi\eta;k}\right)(\mathbf{l}_{\xi\eta;j})\left(\prod_{j<k\leq x} u_{\xi\eta;k}\right)}{\sum_{\{\xi\}} \left(\prod_{0<k\leq x} u_{\xi\eta;k}\right)} \tag{7b}$$

where for $0 < k < x$

$$u_{\xi\eta;k} = u_{\xi;k} z_{\xi\eta;k} \tag{8}$$

with $u_{\zeta;k} = C_{\zeta;k} \exp(-V_k^{\mathrm{s}}/RT)$ and

$$z_{\xi\eta;k} = C_{\xi\eta;k} \sum_{\varphi_k,\psi_k,\omega_{k+1}} \exp(-V_k^{\mathrm{m}}/RT)$$

The quantities $C_{\zeta;k}$ and $C_{\zeta\eta;k}$ account for the factor $\Delta\{\varphi,\psi,\omega,\xi\}$ of eq 7; a detailed description of these factors will be postponed until the Discussion. For $0 < i < x$

$$l_{\xi\eta;i}^{\,2} = u_{\xi;i} C_{\xi\eta;i} \sum_{\varphi_i,\psi_i,\omega_{i+1}} \mathbf{L}_i^{\mathrm{T}}\mathbf{L}_i \exp(-V_i^{\mathrm{m}}/RT) \tag{9}$$

and for $0 < i < k < j < x$

$$\mathbf{l}_{\xi\eta;i}^{\mathrm{T}} = u_{\xi;i} C_{\xi\eta;i} \sum_{\varphi_i,\psi_i,\omega_{i+1}} \mathbf{L}_i^{\mathrm{T}}\hat{\mathbf{T}}_i \exp(-V_i^{\mathrm{m}}/RT) \tag{10}$$

$$\mathbf{t}_{\xi\eta;k} = u_{\xi;k} C_{\xi\eta;k} \sum_{\varphi_k,\psi_k,\omega_{k+1}} \hat{\mathbf{T}}_k \exp(-V_k^{\mathrm{m}}/RT) \tag{11}$$

$$\mathbf{l}_{\xi\eta;j} = u_{\xi;j} C_{\xi\eta;j} \sum_{\varphi_j,\psi_j,\omega_{j+1}} \mathbf{L}_j \exp(-V_j^{\mathrm{m}}/RT) \tag{12}$$

Subscripts $\zeta\eta$ of eq 7–12 denote the ring conformations respectively of residues k and $k + 1$ as specified by the variables ξ_k and ξ_{k+1}. For $i, j, k = x$

$$u_{\xi\eta;x} = u_{\xi;x} = C_{\xi;x} \exp(-V_x^{\mathrm{s}}/RT) \tag{13}$$

$$l_{\xi\eta;x}^{\,2} = l_{\xi;x}^{\,2} = u_{\xi;x}\mathbf{L}_x^{\mathrm{T}}\mathbf{L}_x \tag{14}$$

$$\mathbf{l}_{\xi\eta;x} = \mathbf{l}_{\xi;x} = u_{\xi;x}\mathbf{L}_x \tag{15}$$

since $V_x = V_x^{\mathrm{s}}$ depends solely on the ring conformation of residue x.

The denominator of eq 7 is the classical configuration partition function for the chain of specified chemical sequence.[9,27] It is an explicit sum over all stereochemical sequences $\{\xi\}$, summation over the other configuration variables $\{\varphi,\psi,\omega\}$ having been accomplished formally for each $\{\xi\}$ through separation into factors specified in eq 8 (and 13) by the statistical weights $u_{\zeta\eta;k}$. Evaluation of the latter quantities requires summation of the Boltzmann factor of the conformational energies V_k^{m} over φ_k, ψ_k, ω_{k+1} for each combination $\zeta\eta$ of ring conformations accessible to the pair of residues k and $k + 1$ to yield the dimer partition functions $z_{\zeta\eta;k}$. Estimates of the mutual energy V_k^{m} are conventionally obtained from approximate conformational energy calculations.[15] Self energies V_k^{s} are in principle subject to estimation by the same methods, but reliable calculations of the relative energies of the several ring conformations of sugars have not yet been made owing to the lack of symmetry in the molecules and the difficulty of coping adequately with the significant contributions to the self energy from strain energy contributions. Further considerations of evaluation of V_k^{s} will be reserved for the Discussion section below. Summations in eq 8 over the variables φ_k, ψ_k, ω_{k+1} can be carried out on a judiciously chosen grid of points spaced as closely as required to sample adequately the empirical conformational energy surface.[9b,28] The criterion for adequacy of sampling is convergence of the calculated values of configuration-dependent properties with reduction in grid spacing; the accuracy of the surface can be judged only by its ability to produce agreement of calculated values for a variety of configuration-dependent properties with experimentally observed values. The numerators of eq 7 are likewise explicit sums over all stereochemical sequences $\{\xi\}$ of the chain. Summations over $\{\varphi,\psi,\omega\}$ in the numerators have been factored into summations over φ_k, ψ_k, ω_{k+1} as specified in eq 8–15 for each accessible combination $\zeta\eta$ of ring conformations for the residue pair $k,k + 1$.

Calculations of polysaccharide chain configuration published to date (with two exceptions[14]) treat chains capable of existence in only a single fixed stereochemical sequence in which case summations over $\{\xi\}$ are suppressed in eq 7. Under these conditions eq 7 is equivalent to the earlier treatments[10–13] which uniformly assume separability of the configuration energy $V\{\varphi,\psi,\omega,\xi\}$ for fixed $\{\xi\}$ into independent contributions from each residue pair $k,k + 1$. This approximation, which asserts independence of sets of torsions at each glycosidic linkage but acknowledges interdependences within each set, is predicated on the spatial separation of adjacent glycosidic linkages effected by the large intervening sugar residue.[15] As noted, however, it probably fails if intramolecular hydrogen bonding among sugar residues makes a significant contribution to the conformational energy,[13] and evidently must fail in the more general case treated here in which the configuration energy is clearly not separable owing to its dependence on a particular ξ_k in more than a single term of eq 5. The partial separation

achieved in eq 7 results in the present case from retention of the assumption that sets of torsions at a given glycosidic linkage, while mutually interdependent, are independent of all other such sets of torsions for each fixed sequence of ring conformations. When adjoining residues occur in 1,2 linkage, the assumption of independent sets of glycosidic torsions is probably vitiated by the necessary proximity of first neighboring glycosidic bridges. The present treatment must be presumed unsatisfactory for polysaccharides possessing 1,2 linkages. Particular sugar ring conformations may also precipitate the close approach of first neighbor bridges for 1,3-, 1,4-, and 1,6-linked residues, but consideration of models discloses that in most, if not all, such cases the conformational energies of the chain segments involved would be so large as to preclude important contributions to the chain properties from the offending conformations. With the exceptions noted, the assumption of independence of sets of torsions for a chain of fixed stereochemical sequence is a plausible working hypothesis which has proven sufficient in numerous treatments of the configurational statistics of biopolymers.[15] More demanding experimental tests of the theory than currently available may eventually require that it be revised.

The isolation of independent groups of glycosidic torsion angles for fixed $\{\xi\}$ permits averaging with any desired degree of precision over the conformation space of the independent units, i.e., summations in eq 8–15 can be conducted on as fine a grid of points in conformation space as is warranted by the accuracy of the potential energy surfaces.[15] Because V_k depends simultaneously on ξ_k and ξ_{k+1} ,the summations over $\{\xi\}$ specified in eq 7 can be carried out only if the sugar residues are restricted to a modest number γ_k of conformational isomers. Within this restriction, which will be examined for validity below, the quantities $\langle \mathbf{L}_i \cdot \mathbf{L}_i \rangle$ and $\langle \mathbf{L}_i \cdot \mathbf{L}_j \rangle$ in eq 7 can be computed using matrix methods and the sums in eq 2 evaluated following Flory.[9a]

To this end we define for $0 < k < x$ statistical weight matrices $\mathbf{U}_k = [u_{\zeta\eta;k}]$ having dimensions $\gamma_k \times \gamma_{k+1}$ governed by the number of ring conformations accessible to each residue of the pair $k,k+1$ and indexed for the several respective ring conformation states ζ and η of each. In addition we construct the $\gamma_x \times \gamma_x$ diagonal matrix $\mathbf{U}_x = [u_{\zeta;x}]$ and the vectors $\mathbf{J}_1$ and $\mathbf{J}_x$ comprising respectively γ_1 and γ_x elements, each unity. The denominator, or configuration partition function, of eq 7 is then given by

$$Z = \mathbf{J}_1^{\mathrm{T}} \mathbf{U}_1^{(x)} \mathbf{J}_x \tag{16}$$

The notation $\mathbf{U}_1^{(x)}$ signifies the serial product of factors $\mathbf{U}_k$, beginning with $\mathbf{U}_1$ and containing a total of x factors, the last of which in this case is the diagonal matrix $\mathbf{U}_x$.[9] Similarly, using the $\gamma_i \times \gamma_{i+1}$ matrices $\mathfrak{L}_i^2 = [l_{\zeta\eta;i}^2]$ for $0 < i < x$ and the $\gamma_x \times \gamma_x$ diagonal matrix $\mathfrak{L}_x^2 = [l_{\zeta;x}^2]$ for $i = x$, the numerator of eq 7a can also be evaluated as a matrix product to yield

$$\langle \mathbf{L}_i \cdot \mathbf{L}_i \rangle = Z^{-1} \mathbf{J}_1^{\mathrm{T}} \mathbf{U}_1^{(i-1)} \mathfrak{L}_i^2 \mathbf{U}_{i+1}^{(x-i)} \mathbf{J}_x \tag{17a}$$

The sum in the numerator of eq 7b can likewise be expressed in terms of a matrix product employing for $0 < i < k < j < x$ the $\gamma_i \times 3\gamma_{i+1}$ matrix $\mathfrak{L}_i^{\mathrm{T}} = [\mathbf{l}_{\zeta\eta;i}^{\mathrm{T}}]$, the $3\gamma_k \times 3\gamma_{k+1}$ matrix $\mathfrak{I}_k = [\mathbf{t}_{\zeta\eta;k}]$, and the $3\gamma_j \times \gamma_{j+1}$ matrix $\mathfrak{L}_j = [\mathbf{l}_{\zeta\eta;j}]$ constructed from the matrices defined in eq 10–12. To construct $\mathfrak{I}_k$, for example, the elementary matrices $\mathbf{t}_{\zeta\eta;k}$ enter in pseudorows indexed for the γ_k ring conformations ζ accessible to residue k and in pseudocolumns indexed for the γ_{k+1} conformations η of residue $k+1$. Since the elements $\mathbf{t}_{\zeta\eta;k}$ are square of order 3, the resulting matrix $\mathfrak{I}_k$ is $3\gamma_k \times 3\gamma_{k+1}$. A pseudodiagonal terminal matrix $\mathfrak{L}_x = [\mathbf{l}_{\zeta;x}]$ of dimensionality $3\gamma_x \times \gamma_x$ is also required when $j = x$. Diagonal elements consist of the 3×1 vectors $\mathbf{l}_{\zeta;x}$ indexed for each accessible ring conformation ζ of the residue at the reducing terminus. Off-diagonal elements are 3×1 null vectors. With this device eq 7b may be rewritten as

$$\langle \mathbf{L}_i \cdot \mathbf{L}_j \rangle = Z^{-1} \mathbf{J}_1^{\mathrm{T}} \mathbf{U}_1^{(i-1)} \mathfrak{L}_i^{\mathrm{T}} \mathfrak{I}_{i+1}^{(j-i-1)} \mathfrak{L}_j \mathbf{U}_{j+1}^{(x-j)} \mathbf{J}_x \tag{17b}$$

Substitution from eq 17 into eq 2 and evaluation of the summations in the latter equation using the matrix method of Flory[9a] yields

$$\langle r^2 \rangle_0 = 2Z^{-1} \mathbf{P}_1^{\mathrm{T}} \mathbf{G}_1^{(x)} \mathbf{P}_x \tag{18}$$

Here $\mathbf{P}_1$ is the $5\gamma_1 \times 1$ vector comprising γ_1 ones followed by $4\gamma_1$ zeros, $\mathbf{P}_x$ is the $5\gamma_x \times 1$ vector commencing with $4\gamma_x$ zeros and terminated by γ_x ones, and $\mathbf{G}_i$ is the $5\gamma_i \times 5\gamma_{i+1}$ matrix

$$\mathbf{G}_i = \begin{bmatrix} \mathbf{U} & \mathfrak{L}^{\mathrm{T}} & \tfrac{1}{2}\mathfrak{L}^2 \\ \mathbf{0} & \mathfrak{I} & \mathfrak{L} \\ \mathbf{0} & \mathbf{0} & \mathbf{U} \end{bmatrix}_i \tag{19}$$

where the subscript i applies to each element of $\mathbf{G}_i$, and the null matrices $\mathbf{0}$ are of dimensions required to complete the $5\gamma_i \times 5\gamma_{i+1}$ matrix. Note that $\mathbf{G}_x$ differs in detail from $\mathbf{G}_i$ for $0 < i < x$ because of the special definitions of the terminal factors $\mathbf{U}_x$, $\mathfrak{L}_x^2$, and $\mathfrak{L}_x$. In similar fashion the mean-square radius of gyration for the polysaccharide chain of specified chemical sequence is given by[9a]

$$\langle s^2 \rangle_0 = 2(x+1)^{-2} Z^{-1} \mathbf{Q}_1^{\mathrm{T}} \mathbf{H}_1^{(x)} \mathbf{Q}_x \tag{20}$$

where $\mathbf{Q}_1$ is the $7\gamma_1 \times 1$ vector with γ_1 ones succeeded by $6\gamma_1$ zeros, $\mathbf{Q}_x$ is the $7\gamma_x \times 1$ vector with $6\gamma_x$ zeros followed by γ_x ones, and $\mathbf{H}_i$ is the $7\gamma_i \times 7\gamma_{i+1}$ matrix

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{U} & \mathbf{U} & \mathfrak{L}^{\mathrm{T}} & \tfrac{1}{2}\mathfrak{L}^{\mathrm{T}} & \tfrac{1}{2}\mathfrak{L}^2 \\ \mathbf{0} & \mathbf{U} & \mathfrak{L}^{\mathrm{T}} & \tfrac{1}{2}\mathfrak{L}^2 & \tfrac{1}{2}\mathfrak{L}^2 \\ \mathbf{0} & \mathbf{0} & \mathfrak{I} & \mathfrak{L} & \mathfrak{L} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{U} & \mathbf{U} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{U} \end{bmatrix}_i \tag{21}$$

Other configuration dependent properties are subject to calculation by extension of the method described here.[9] Normalization of $\mathbf{U}_i$, $\mathbf{G}_i$, and $\mathbf{H}_i$ with the largest eigenvalue of $\mathbf{U}_i$ will be required for chains exceeding modest length.[9a]

Evaluation of the Characteristic Ratios and Mean Stereochemical Compositions

Let characteristic dimensionless ratios of the mean-square unperturbed end-to-end distance and radius of gyration be defined by dividing $\langle r^2 \rangle_0$ and $\langle s^2 \rangle_0$ by the mean-square end-to-end distance of a chain of *uncorrelated* virtual and chemical bond vectors $\mathbf{l}_i$ and $\mathbf{b}_i$ corresponding to a polymer with the same chemical and stereochemical composition as the real chain of interest.[23]

$$C_x^{\mathrm{r}} = \frac{\langle r^2 \rangle_0}{\sum_p \sum_{\zeta(p)} \langle n_{\zeta(p)} \rangle L_{\zeta(p)}^2} \tag{22a}$$

$$C_x^{\mathrm{s}} = \frac{\langle s^2 \rangle_0}{\sum_p \sum_{\zeta(p)} \langle n_{\zeta(p)} \rangle L_{\zeta(p)}^2} \tag{22b}$$

Here the sum over p is over all chemically different kinds of sugar residues in the chain, the sum over $\zeta(p)$ is over all accessible ring conformations of residues of type p, and $\langle n_{\zeta(p)} \rangle$ is the mean number of residues of type p in conformation $\zeta(p)$. The quantity $L_{\zeta(p)}^2$ is the mean-square magnitude of the residue vector $L_{\zeta(p)}$ for a residue of type p in conformation $\zeta(p)$.[29,30] For a residue linked 1,6 to its successor, e.g., residue i of Figure 1, $L_{\zeta(p)}^2 = l_{\zeta(p)}^2 + b_{\zeta(p)}^2$, where $l_{\zeta(p)}^2$ and $b_{\zeta(p)}^2$ are respectively the squared magni-

tudes of the virtual and chemical bond vectors $\mathbf{l}_{\zeta(p)}$ and $\mathbf{b}_{\zeta(p)}$ which compromise the residue vector. For a residue linked 1,m (m = 2, 3, 4) to its successor $L_{\zeta(p)}{}^2 = l_{\zeta(p)}{}^2$.

One may evaluate $\langle n_{\zeta(p)} \rangle$ by differentiation of the configuration partition function[9a]

$$\langle n_{\zeta(p)} \rangle = \frac{\partial \ln Z}{\partial \ln u_{\zeta(p)}} \tag{23}$$

The mean frequency of coincidences of neighboring residues of types p and q in conformations ζ(p) and η(q), respectively, is likewise given by

$$\langle n_{\zeta(p)\eta(q)} \rangle = \frac{\partial \ln Z}{\partial \ln u_{\zeta(p)\eta(q)}} \tag{24}$$

The derivatives in eq 23 and 24 may be evaluated conveniently by a matrix method after Jernigan and Flory.[9a] Thus

$$\langle n_{\zeta(p)} \rangle = Z^{-1}\mathbf{R}_1{}^{\mathrm{T}}\hat{\mathbf{U}}^{(x)}_{\zeta(p);1}\mathbf{R}_x \tag{25}$$

$$\langle n_{\zeta(p)\eta(q)} \rangle = Z^{-1}\mathbf{R}_1{}^{\mathrm{T}}\hat{\mathbf{U}}^{(x)}_{\zeta(p)\eta(q);1}\mathbf{R}_x \tag{26}$$

in which $\mathbf{R}_1$ is the vector of γ_1 ones followed by γ_1 zeros and $\mathbf{R}_x$ is the vector with γ_x zeros followed by γ_x ones. The $2\gamma_i \times 2\gamma_{i+1}$ matrices $\hat{\mathbf{U}}_{\zeta(p);i}$ and $\hat{\mathbf{U}}_{\zeta(p)\eta(q);i}$ are defined by

$$\hat{\mathbf{U}}_{\zeta(p);i} = \begin{bmatrix} \mathbf{U} & \mathbf{U}'_{\zeta(p)} \\ \mathbf{0} & \mathbf{U} \end{bmatrix}_i \tag{27}$$

$$\hat{\mathbf{U}}_{\zeta(p)\eta(q);i} = \begin{bmatrix} \mathbf{U} & \mathbf{U}'_{\zeta(p)\eta(q)} \\ \mathbf{0} & \mathbf{U} \end{bmatrix}_i \tag{28}$$

where $\mathbf{0}$ is the $\gamma_i \times \gamma_{i+1}$ null matrix and

$$\mathbf{U}'_{\zeta(p);i} = \frac{\partial \mathbf{U}_i}{\partial \ln u_{\zeta(p);i}} \tag{29}$$

$$\mathbf{U}'_{\zeta(p)\eta(q);i} = \frac{\partial \mathbf{U}_i}{\partial \ln u_{\zeta(p)\eta(q);i}} \tag{30}$$

Subscripts i of eq 27 and 28 apply to each element of $\hat{\mathbf{U}}_{\zeta(p);i}$ and $\hat{\mathbf{U}}_{\zeta(p)\eta(q);i}$. Application of eq 25–30 evidently requires attention to the chemical sequence of sugar residues in the chain. Matrices $\mathbf{U}'_{\zeta(p);i}$ are null of order $\gamma_i \times \gamma_{i+1}$ unless a residue of type p occurs at sequential position i; the same is true of $\mathbf{U}'_{\zeta(p)\eta(q);i}$ unless residues i and $i + 1$ are simultaneously of types p and q, respectively. Methods for calculating the temperature coefficients of Z, $\langle n_{\zeta(p)} \rangle$, and $\langle n_{\zeta(p)\eta(q)} \rangle$ are given by Flory,[9a] where procedures for calculating the a priori and conditional probabilities of a residue at a particular sequential position being in a particular conformation are also described. The temperature coefficients of $C_x{}^{\mathrm{r}}$ and $C_x{}^{\mathrm{s}}$ may be computed most readily using eq 18 and 20 employing different values of the temperature in equations 8–14. It is important to recognize in so doing that the assumption of temperature independence for the configuration energy $V\{\varphi,\psi,\omega,\xi\}$ is rendered unavoidable by the inscrutable character of the temperature dependence of this free energy parameter.

Discussion

The Mutual Free Energy Effect. It is of interest to inquire, for a chain of given chemical sequence, into the frequency of occurrence of residues of type p in conformational isomeric state $\zeta(p)$ in order to examine the effect upon some configuration-dependent property of the chain from the presence of such residues. Likewise the frequency of occurrence of the dimeric stereochemical sequence $\zeta(p)\eta(q)$ for a given chemical sequence pq may also be of interest. These quantities may be computed using eq 25 and 26.[31] It is apparent that for long chains

$$\langle n_{\zeta(p)} \rangle = \sum_q \sum_{\eta(q)} \langle n_{\zeta(p)\eta(q)} \rangle \tag{31}$$

where the outer sum is over all chemically different kinds of residues in the chain and the inner sum is over all conformations accessible to each kind of residue; for shorter chains eq 31 will not reproduce the results of eq 25 owing to neglect in eq 31 of effects due to terminal residues.

The statistical weight $u_{\zeta(p)\eta(q);i}$ governing the probability of occurrence at positions i and $i + 1$ of the stereochemical sequence $\zeta(p)\eta(q)$ depends according to eq 8 on the self energy of residue i *and* on the mutual energy of residues i and $i + 1$. The mean frequency $\langle n_{\zeta(p)\eta(q)} \rangle$ is evidently dependent on the magnitude of $u_{\zeta(p)\eta(q)}$ in relation to the magnitudes of the statistical weights of other stereochemical sequences. In view of the foregoing and eq 31 it is clear that the mean population $\langle n_{\zeta(p)} \rangle$ of residues of type p in conformation $\zeta(p)$ cannot be judged solely on the basis of the relative self energies of the several conformers of residues of type p. Insofar as the mutual energy of residue i is a different function of its orientation (as governed by φ_i, ψ_i, ω_{i+1}) with respect to succeeding residues $i + 1$ of varying chemical types, $\langle n_{\zeta(p)} \rangle$ may depend as well on the chemical identity of the residues which compose the chain, on the ring conformations accessible to these several residue types, and, indeed, on the sequence distribution of the sugar residues.

Scrutiny of eq 25, 27, and 29 reveals that the proper measure of the importance of neighbor interactions in determining the mean stereochemical composition of the chain is the mutual free energy $\Delta A_{\zeta(p)\eta(q)}$ defined for the stereochemical sequence $\zeta(p)\eta(q)$ by $\Delta A_{\zeta(p)\eta(q)} = -RT \ln z_{\zeta(p)\eta(q)}$. The quantity $z_{\zeta(p)\eta(q)}$ is recognized as the configuration partition function for the dimeric unit $\zeta(p)\eta(q)$ of fixed chemical and stereochemical sequence; it appears as the second factor in eq 8. The factor $C_{\zeta(p)\eta(q)}$ which precedes the summation in the definition of $z_{\zeta(p)\eta(q)}$ arises in changing from the classical integration over the configuration variables φ_i, ψ_i, ω_{i+1} to a summation over rotational isomeric states each specified by a set of discrete values for these variables. It accounts for the volume in configuration space associated with each rotational isomeric state and is given by[32]

$$C_{\zeta(p)\eta(q)} = \left(\frac{2\pi}{N_{\varphi_i}}\right)\left(\frac{2\pi}{N_{\psi_i}}\right)\left(\frac{2\pi}{N_{\omega_{i+1}}}\right) \tag{32}$$

where the symbols N_{φ_i}, N_{ψ_i}, and $N_{\omega_{i+1}}$ represent the numbers of values taken in the summation at equal intervals over each of the respective torsion angles for the particular residue pair and ring conformations in question. For linkages other than 1,6 the factor $(2\pi/N_{\omega_{i+1}})$ is absent. The factors $C_{\zeta(p)\eta(q)}$ evidently cancel from numerator and denominator of eq 7, provided N_{φ_i}, N_{ψ_i}, and $N_{\omega_{i+1}}$ are independent of ring conformation for residues i and $i + 1$.[27] Contributions to the mutual free energy from the average mutual energy $\Delta E_{\zeta(p)\eta(q)}$ and the mutual entropy $\Delta S_{\zeta(p)\eta(q)}$ may be calculated using standard methods.

$$\Delta A_{\zeta(p)\eta(q)} = \Delta E_{\zeta(p)\eta(q)} - T\Delta S_{\zeta(p)\eta(q)} \tag{33}$$

$$\Delta E_{\zeta(p)\eta(q)} = RT^2\left(\frac{\partial \ln z_{\zeta(p)\eta(q)}}{\partial T}\right) = \frac{\sum_{\varphi_i,\psi_i,\omega_{i+1}} V_i{}^{\mathrm{m}} \exp(-V_i{}^{\mathrm{m}}/RT)}{\sum_{\varphi_i,\psi_i,\omega_{i+1}} \exp(-V_i{}^{\mathrm{m}}/RT)} \tag{34}$$

$$\Delta S_{\zeta(p)\eta(q)} = R \ln z_{\zeta(p)\eta(q)} + RT\left(\frac{\partial \ln z_{\zeta(p)\eta(q)}}{\partial T}\right) \tag{35}$$

Mutual energies, entropies, and free energies calculated according to eq 33–35 assume values relative to the value zero in a hypothetical reference state for the dimer characterized by unhindered rotation of torsion angles φ_i, ψ_i, ω_{i+1} throughout a domain in conformation space having dimensions of one cubic radian (or one square radian for residue pairs linked other than 1,6) and possessing a mutual conformational energy everywhere zero within this domain. The average mutual energy and the mutual entropy and free energy so calculated may therefore take on positive, negative, or zero values. They evidently contain no contributions from degrees of freedom for which the associated hard coordinates have previously been fixed at their equilibrium values. Chemical and stereochemical sequences $\zeta(p)\eta(q)$ possessing relatively small mean mutual energies and/or relatively large extents of mutual conformational freedom, as reflected by relatively large mutual entropies, will be characterized by smaller mutual free energies and, hence, will sustain enhanced probabilities of occurrence relative to sequences not stabilized by these factors. Comparison of mutual free energies among dimers of varying chemical and stereochemical sequence requires for consistency that the factors $C_{\zeta(p)\eta(q)}$ be taken into account. It is likewise necessary that mutual conformational energies V_i^m be calculated on a consistent basis. This can normally be accomplished by letting the nonbonded and electrostatic contributions approach zero as the lengths of the bonds flanking the glycosidic oxygen approach infinity while setting the inherent torsional energies for the linkage bonds to zero in their energy minima.[15]

The collective influence of neighbor interactions in determining the mean population $\langle n_{\zeta(p)} \rangle$ may be expressed in terms of a mean mutual free energy $\Delta A_{\zeta(p)}$ defined by

$$\frac{\langle n_{\zeta(p)} \rangle}{n(p)} = \frac{u_{\zeta(p)} \exp(-\Delta A_{\zeta(p)}/RT)}{\sum_{\zeta(p)} u_{\zeta(p)} \exp(-\Delta A_{\zeta(p)}/RT)} \quad (36)$$

where the sum in the denominator is over all accessible conformations $\zeta(p)$ for residues of type p and $n(p)$ is the number of residues of type p in the chain. Employing values of $\langle n_{\zeta(p)} \rangle$ calculated from eq 25, 27, and 29, one may use eq 36 to evaluate $\Delta A_{\zeta(p)}$ for each ring conformation relative to the value zero assigned to some arbitrarily chosen conformer. In the case of cellulosic chains it has been found that the frequencies of occurrence of certain conformations of β-D-glucose possessing self energies high in relation to that of the least energetic conformer are materially enhanced by favorable mutual free energy effects.[14a] Hence, for cellulosics Bernoullian sequence statistics do not obtain for the distribution of alternate ring conformers in the chain.[14b]

Incorporation of Pseudorotating Pyranose Conformers. When bond angles and bond lengths of the set of variables ξ_i describing the ring conformation of a pyranose residue i are rigidly fixed, it may be shown by very general considerations[20] that the chair form conformers possess no residual degrees of conformational freedom. That is, remaining variables of the set ξ_i, namely, the endocyclic torsion angles $\nu(m)_i$,[24] are also "hard" and may not vary. Hence, the chair conformers can justifiably be regarded as rigid constituents of the polysaccharide skeleton spanned by virtual bonds of constant length. The degree of approximation is no greater than that normally made in treating the configurational statistics of structurally simpler polymer chains when bond angles and bond lengths are assigned to their fixed mean values. It is furthermore appropriate to assign one or both chair forms of pyranose residue i among the γ_i conformational isomeric states accessible to that residue. Such states are structurally well defined, i.e., possess fixed ξ_i, and evidently exist at discrete minima on the conformational energy surface for the molecule.[16,21] Consistent with the conventional presumptions of isomeric state theory the chair forms may be considered to make significant contributions to the equilibrium configurational properties of the chain, the relative importance of which depends, of course, on their respective self energies and mutual energies of interaction with neighboring residues.

In contrast to the chair conformers just described unsubstituted six-membered ring structures characterized by fixed bond angles and lengths and possessing C_2 symmetry retain a single "soft" degree of conformational freedom among the variables of the set ξ_i. The value of any one torsion angle, say $\nu(1)_i$, may be selected arbitrarily, but the remaining five are then fixed by conditions imposed by the constraints of ring closure and the specified symmetry.[19,20] Progress along the pseudorotation intinerary for the flexible forms of this idealized ring may be measured conveniently in terms of the single angular variable δ_i.[33] Pseudorotation of pyranose flexible forms, the ring skeletons of which lack true C_2 symmetry, may likewise be described approximately by the single variable δ_i, acknowledging that relatively small distortions of bond angles (and/or lengths) must accompany the changes in δ_i.

To the extent that the section through the conformational energy surface along the pseudorotation coordinate lacks relief, identification of some or all of the γ_i conformational isomeric states of residue i with flexible conformers at minima on this surface becomes invalid, because all such conformations possess similar energies under these conditions and no discrete and limited set of conformers can be viewed as making the predominant contributions to the equilibrium configurational properties of the chain. Lacking justification for identifying discrete flexible conformations of low energy, the isomeric states may be chosen instead to sample the space of the variable δ_i in a representative fashion. This is in complete analogy to the procedure described above for systematic sampling of the energy surface when integrating over the space of the linkage coordinates φ_i, ψ_i, ω_{i+1} for fixed ξ_i and ξ_{i+1}.[9b,28] In the latter case, because of the assumed independence of sets of linkage coordinates, there was no practical limitation to the frequency of sampling and, hence, to the precision to which the integral could be approximated. Limitations do exist on the number of conformational isomeric states γ_i which can be employed, owing to the inseparability of the configuration energy, but the accuracy with which the energy surface for pseudorotation is currently known does not warrant sampling with a frequency greater than the practical constraints imposed by the dimensions of matrices $\mathbf{G}_i$ and $\mathbf{H}_i$ on the number of states γ_i. The above comments with regard to treatment of the flexible pyranose forms apply equally to the furanose sugars.

Lack of knowledge concerning the details of the variation of conformational energy with progress along the pseudorotation coordinate may dictate that the self energies associated with the several flexible ring conformations of a given sugar be regarded as adjustable parameters of the calculation.[35] From this point of view it may be possible to gain information about the relative self energies of the several conformers of a given sugar from the values required to bring theoretical calculations of configuration-dependent properties into conformity with a body of experimental findings. Interpretation of relative self energies determined under these conditions will depend upon the characteristics assumed for the energy surface along the pseudorotation coordinate. Two limiting cases will be described here: essentially free pseudorotation and strongly hindered pseudorotation.

When there is reason to believe that the flexible conformers of a residue of chemical type p are capable of interconversion along a path of low energetic relief, it is appropriate to define the conformational free energy of the flexible *class* by $\Delta A_{\zeta'(p)} = -RT \ln z_{\zeta'(p)}$, where $z_{\zeta'(p)}$ is the conformational partition function for the flexible class of residues of type p given by

$$z_{\zeta'(p)} = \sum_{\xi'(p)} C_{\xi'(p)} \exp(-V_{\xi'(p)}{}^{s}/RT) \qquad (37)$$

Here the summand is just $u_{\zeta(p)}$, introduced with reference to eq 8,[30] and the sum is over the $\gamma'(p)$ conformational isomeric states $\zeta'(p)$ taken to represent the flexible class of residues of type p. The factor $C_{\zeta'(p)}$ accounts in part for the volume in configuration space associated with each isomeric state of the chain.[27] When variations in energy along the pseudorotation coordinate are of the order of kT, $C_{\zeta'(p)}$ may be approximated in eq 8–15 and 37 by $2\pi/\gamma'(p)$, provided the $\gamma'(p)$ isomeric states are selected with uniform spacing along the pseudorotation coordinate. In this limiting case $V_{\zeta'(p)}{}^{s}$ is understood as the potential of mean force for the flexible conformer in question subject to prior averaging over all mutual arrangements of the polymer molecule and the other molecules of the medium and over the vibrational motions of all of the hard variables of the polymer molecule.[9b] The factor $2\pi/\gamma'(p)$ reflects the entropic stabilization afforded the flexible class relative to the chair form which arises from the additional soft degree of freedom among the conformational variables of the flexible class.

The quantity $\Delta A_{\zeta'(p)} - V_{\zeta'(p)}{}^{s}$ provides a measure of the free energy difference between conformations of the readily interconverting flexible class and a chair conformer, designated $\zeta^*(p)$, which is effectively independent of the arbitrary number $\gamma'(p)$ of flexible conformers selected to sample the energy surface for pseudorotation; normally $V_{\zeta'(p)}{}^{s}$ will be assigned the value zero for one or the other chair form.[36] No factor analogous to $2\pi/\gamma'(p)$ enters explicitly in the free energy $V_{\zeta'(p)}{}^{s}$ of the chair form because all of the variables $\nu(m)$ (or alternatively μ[33]) are among the hard variables for chair conformers, and integration over these coordinates is taken into account implicitly in assigning the parameters of the conformational energy functions.[9b,15] Hence, for chair conformers $C_{\zeta(p)} = 1$ in eq 8–15, and whenever chair and readily interconverting flexible conformations are both accessible to residues of the chain, the quantities $C_{\zeta(p)}$ cannot be factored from the numerator and denominator of eq 7.[27] Although the quantity $2\pi/\gamma'(p)$ could evidently be absorbed into the exponential factor of $u_{\zeta(p)}$ for the readily interconverting flexible isomers, its explicit appearance has an important conceptual virtue in that it renders the adjustable parameter $\Delta A_{\zeta'(p)}$ essentially independent of the number $\gamma'(p)$ of flexible class isomeric states. Moreover, the entropic advantage of the flexible class is explicitly acknowledged.

When relief along the pseudorotation coordinate is large and energy barriers separating the most stable conformers are several times kT, δ becomes a hard variable as it is for the chair conformers, and its displacements are vibrations rather than rotations. Under these circumstances it is no longer appropriate to group together as a class the forms possessing approximate C_2 symmetry. They are preferably regarded as distinct conformational isomers and treated equivalently to the chair forms. In this case all $C_{\zeta(p)}$ should be equated to unity in eq 8–15, all self energies being understood as potentials of mean force averaged over the configuration space of the molecules of the medium and all of the hard configuration variables of the macromolecule. Situations intermediate between the limiting examples presented can be treated in a fashion analogous to that of strongly hindered rotation. It will be recognized in such cases, however, that the self energies for the flexible forms represent potentials of mean force subject, *inter alia*, to prior averaging over the *soft* configuration variable δ; arguments for explicit retention of the factor $2\pi/\gamma'(p)$ lose force as the barriers to pseudorotation increase.

Chain Stiffness and Flexibility.

The question of the stiffness or flexibility of a polysaccharide chain involves considerations both of the chain extension and of the inherent conformational freedom of the skeletal segments which constitute the chain. It is important to distinguish between these two factors when discussing chain stiffness or flexibility. Unperturbed chain extension is adequately measured by a characteristic ratio such as $C_x{}^{r}$ or $C_x{}^{s}$, either defined as above or in some other appropriate way which normalizes the linear chain dimensions by a factor proportional to the square root of the degree of polymerization. Thus, amylosic chains with $C_\infty{}^{r} \sim 6$[37] are significantly less extended than cellulosic chains for which $C_\infty{}^{r} \sim 35$,[14a,38] particularly when account is taken of the longer residue vector length which enters the denominator of $C_\infty{}^{r}$ for cellulosic (ca. 5.45 Å) as compared to amylosic (ca. 4.25 Å) chains. Thus, $\langle r^2\rangle_{0,cel}/\langle r^2\rangle_{0,amy} \sim 10$ for chains of sufficient length for $C_x{}^{r}$ to have converged to its asymptotic limit for large x.

It can be demonstrated on the other hand that despite the greater extension of cellulosic coils, the conformational freedom of the cellulosic skeletal segments exceeds that of amylosic segments.[39] The configuration partition function has been suggested as an appropriate measure of the conformational freedom of skeletal segments.[12c,40] We prefer to employ the configurational entropy per residue, ΔS_c, this quantity being independent of the choice of zero of configuration energy. In general

$$\Delta S_c = x^{-1}\left[R \ln Z + RT\left(\frac{\partial \ln Z}{\partial T}\right)\right] \qquad (38)$$

where Z must be evaluated from eq 16 with due attention to inclusion of factors $C_{\zeta;k}$ and $C_{\zeta\eta;k}$ in the statistical weight matrices $\mathbf{U}_k$. For homopolymeric chains composed of residues incapable of conformational isomerism $\Delta S_c = \Delta S_{\zeta(p)\eta(q)}$ with $p = q$ and $\zeta = \eta$. Assuming cellulosic and amylosic chains to conform to this description and employing methods described earlier,[13] we find $\Delta S_{c,cel} = -0.1$ cal/deg mol and $\Delta S_{c,amy} = -1.4$ cal/deg mol;[39] negative values reflect the reference state described above. The corresponding values calculated for $C_\infty{}^{r}$ are ~115 and ~7. Modification of the calculation for cellulosic chains to reproduce the observed characteristic ratio,[14a] i.e., ca. 35, increases $\Delta S_{c,cel}$ still further. Hence, although cellulosics are the "stiffer" chains in the sense of chain extension, they are the more "flexible" in regard to the conformational freedom of their skeletal segments. This example clearly illustrates the necessity for care in discussion of chain stiffness and flexibility.

Acknowledgment. Beneficial discussions with Charles Harvie and Max Wolfsberg are gratefully acknowledged.

References and Notes

(1) This work was supported in part by U.S. Public Health Service Research Grant GM 13908 and in part by National Science Foundation Research Grant GP 37792.
(2) Address correspondence to this author.
(3) H. Benoit, *J. Polym. Sci.*, **3**, 376 (1948).
(4) J. Eliezer and H. J. G. Hayman, *J. Polym. Sci.*, **23**, 387 (1957).
(5) W. Burchard, *Makromol. Chem.*, **42**, 151 (1960).
(6) R. L. Cleland, *Biopolymers*, **9**, 811 (1970).
(7) N. Yathindra and V. S. R. Rao, *J. Polym. Sci., Part A-2*, **9**, 1149 (1971).

(8) The term "conformation" is employed with reference to any of the several geometric arrangements of the atoms of a given molecule, or portion of a molecule, which are related to one another principally by internal rotations about the (single) chemical bonds of the molecule. A particular molecular conformation is sometimes referred to as a "conformer". The term "configuration" is used with reference to the configuration integral of classical statistical mechanics, to the configuration energy which enters the configuration integral, and to the configuration-dependent properties of the polymeric species which depend upon the average over all conformations and which may be computed using the classical configuration integral. We likewise speak of the configuration variables or configuration space over which the configuration integral is evaluated and in terms of which a given conformation of the entire polymer chain may be described in detail. With reference to specific representative skeletal segments of the chain the terms configuration energy and configuration space are sometimes replaced respectively by conformational energy and conformation space.
(9) (a) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969; (b) P. J. Flory, *Macromolecules*, **7**, 381 (1974).
(10) (a) V. S. R. Rao, N. Yathindra, and P. R. Sundararajan, *Biopolymers*, **8**, 325 (1969); (b) N. Yathindra and V. S. R. Rao, *ibid.*, **9**, 783 (1970); (c) N. Yathindra and V. S. R. Rao, *ibid.*, **10**, 1891 (1971).
(11) R. L. Cleland, *Biopolymers*, **10**, 1925 (1971).
(12) (a) S. G. Whittington, *Biopolymers*, **10**, 1481, 1617 (1971); (b) *Macromolecules*, **4**, 569 (1971); (c) S. G. Whittington and R. M. Glover, *ibid.*, **5**, 55 (1972); (d) G. M. Hallman and S. G. Whittington, *ibid.*, **6**, 386 (1973).
(13) (a) C. V. Goebel, W. L. Dimpfl, and D. A. Brant, *Macromolecules*, **3**, 644 (1970); (b) D. A. Brant and W. L. Dimpfl, *ibid.*, **3**, 655 (1970).
(14) (a) D. A. Brant and K. D. Goebel, *Macromolecules*, **5**, 536 (1972); (b) L. G. Dunfield and S. G. Whittington, *ibid.*, **7**, 946 (1974).
(15) D. A. Brant, *Annu. Rev. Biophys. Bioeng.*, **1**, 369 (1972).
(16) J. F. Stoddart, "Stereochemistry of Carbohydrates", Interscience, New York, N.Y., 1971.
(17) W. K. Olson and P. J. Flory, *Biopolymers*, **11**, 1, 25, 57 (1972).
(18) (a) W. L. Mattice, K. Nishikawa, and T. Ooi, *Macromolecules*, **6**, 443 (1973); (b) T. Ooi, D. S. Clark, and W. L. Mattice, *ibid.*, **7**, 337 (1974).
(19) J. D. Dunnitz, *J. Chem. Educ.*, **47**, 488 (1970).
(20) (a) N. Gō and H. A. Scheraga, *Macromolecules*, **6**, 273 (1973); (b) *ibid.*, **3**, 178 (1970).
(21) (a) H. M. Pickett and H. L. Strauss, *J. Chem. Phys.*, **53**, 376 (1970); (b) H. M. Pickett and H. L. Strauss, *J. Am. Chem. Soc.*, **92**, 7281 (1970); H. L. Strauss, *J. Chem. Educ.*, **48**, 221 (1971).
(22) The notation employed conforms in most respects to the unpublished proposals of the IUPAC-IUB Subcommittee on Polysaccharide Nomenclature (draft SCBN/P.S./6) under the chairmanship of Professor R. H. Marchessault, Department of Chemistry, University of Montreal, Montreal 101, Quebec.
(23) Let the terms "chemical and stereochemical composition" refer respectively to the frequency of occurrence in the chain of residues of each chemical type and the frequency of occurrence of the several ring conformations characteristic of residues of each chemical type. The term "chemical sequence" will refer to the sequential ordering in the chain or a chain segment of the residues of various chemical types. The "stereochemical sequence" is the order of occurrence of the several ring conformations for a chain or chain segment of given chemical sequence.
(24) In these definitions the generic symbol θ[A,B,C,D] represents a torsion or dihedral angle defined by bonds A–B, B–C, and C–D connecting the four atoms enclosed in the square brackets. The eclipsed (cis, synplanar) conformation in which the projections of bonds A–B and C–D coincide in a plane perpendicular to bond B–C corresponds to $\theta = 0°$. Torsion angles are measured in the range −180 to +180° relative to this origin. The positive sense of rotation about bond B–C is that which, when viewing along that bond toward the atom at its more distant terminus, corresponds to a clockwise motion of the atoms or groups bonded to that terminal atom. This convention is independent of the direction of viewing along bond B–C. Endocyclic and exocyclic torsion angles for residue i are designated respectively $\nu(m)_i$ and $\chi(m)_i$; detailed definitions of these angles are not required for present purposes but are presented here for pyranose rings for future reference: $\nu(1) = \theta$[O(5),C(1),C(2),C(3)], $\nu(2) = \theta$[C(1),C(2),C(3),C(4)], $\nu(3) = \theta$[C(2),C(3),C(4),C(5)], $\nu(4) = \theta$[C(3),C(4),C(5),O(5)], $\nu(5) = \theta$[C(4),C(5),O(5),C(1)], $\nu(6) = \theta$[C(5),O(5),C(1),C(2)] and $\chi(m) = \theta$[C(m − 1),C(m),O(m),HO(m)] for m = 2, 3, 4, and 6, $\chi(1) = \theta$[O(5),C(1),O(1),HO(1)], $\chi(5) = \theta$[O(5),C(5),C(6),O(6)].
(25) *Alternatively, the chain vector might be considered to originate at* $O(4)_1$ in which case the contribution from the preceding bond $H(4)_1$–$O(4)_1$ would be neglected.
(26) The quantity $V\{\varphi, \psi, \omega, \xi\}$ is a property of the chain as a whole. It is known as the "configuration" energy because of its role in the configuration integral (configuration partition function) of classical statistical mechanics. Contributions V_k, introduced in eq 4, to the configuration energy from the several residue pairs $k,k+1$ are known as "conformational" energies in keeping with common usage.[8]
(27) The factors $C_{\zeta\eta;k}$ which appear in eq 8–12 normally may be factored from the numerator and denominator of eq 7 and hence are of no consequence for the calculation of configuration dependent properties. They are retained for purposes of a discussion below with reference to evaluation of thermodynamic quantities from the configuration partition function. The factors $C_{\zeta;k}$ in general may not be factored from the summations. This matter will be examined further in the Discussion.
(28) P. J. Flory, *J. Polym. Sci., Polym. Phys. Ed.*, **11**, 621 (1973).
(29) The definitions of the characteristic ratios given in eq 22 are consistent with the usual definitions.[9a] It should be recognized, however, that the present definitions, which require a detailed knowledge of the chemical and stereochemical composition of the chain, will be suitable for comparison of calculated and experimental unperturbed chain dimensions only if the experimental values of $\langle r^2\rangle_0$ and $\langle s^2\rangle_0$ are also normalized using the function which appears in the denominator of eq 22. Evidently any consistent choice of normalizing function will render theoretical and experimental results comparable.
(30) For the present discussion and in considerations to follow the chemical identities as well as the ring conformations of the several sugars which constitute the polymer chain under consideration become matters of immediate concern. Consequently, the previous notation $\mathbf{L}_i$, $\mathbf{l}_i$, $\mathbf{b}_i$, V_i^s, V_i^m, $u_{\zeta;i}$, $u_{\zeta\eta;i}$, $z_{\zeta\eta;i}$, $C_{\zeta;i}$, and $C_{\zeta\eta;i}$ may be elaborated for these respective quantities to $\mathbf{L}_{\zeta(p);i}$, $\mathbf{l}_{\zeta(p);i}$, $\mathbf{b}_{\zeta(p);i}$, $V_{\zeta(p);i}{}^s$, $V_{\zeta(p)\eta(q);i}{}^m$, $u_{\zeta(p);i}$, $u_{\zeta(p)\eta(q);i}$, $z_{\zeta(p)\eta(q);i}$, $C_{\zeta(p);i}$, and $C_{\zeta(p)\eta(q);i}$ in order to designate explicitly, using identifiers p and q, the chemical identities, as well as the ring conformations, of residues i and $i+1$. In circumstances where the sequential position of the residues is of no importance, e.g., in eq 22–24, the serial index may be omitted.
(31) *When these frequencies are sought for copolymeric materials of given mean* chemical composition but which occur in samples containing molecules of varying composition and chemical sequence, Monte Carlo averaging is required over a suitable sample of chains characterized by the mean chemical composition in question. Chemical sequence distributions for the molecules of the Monte Carlo sample must, of course, be consistent with the copolymerization statistics of the material under investigation.
(32) M. Gō, N. Gō, and H. A. Scheraga, *J. Chem. Phys.*, **52**, 2060 (1970).
(33) The torsion angles $\nu(m)_i$ do not vary monotonically as the pseudorotation cycle is traversed for six-membered rings with C_2 symmetry.[20] It is consequently desirable to transform from the set of internal rotation angles $\nu(m)_i$ to an alternative set of coordinates μ_i in terms of which the conformations of six-membered rings are more conveniently expressed. Pseudorotation of flexible forms may be described in terms of variables from the set of μ_i by an azimuthal angle δ_i which varies monotonically from 0 to 2π as the ring executes one cycle of pseudorotation; in this regard δ_i is analogous to the pseudorotation coordinate φ of Pickett and Strauss.[21] All variables of the set μ_i, save δ_i, are hard for flexible six-membered rings with fixed bond angles and lengths. Interconversion of flexible and chair ring conformers proceeds through a large energy barrier[21] and corresponds to changes in variables of μ_i in addition to δ_i. All variables in μ_i, including δ_i, are hard for chair form conformers. The Jacobian of the transformation from the set $\nu(m)_i$ to μ_i depends only on the hard variables of μ_i. It can therefore be treated as a constant and removed from the configuration integral in the usual fashion.[9b,34] Because the ratio of this Jacobian for chair and flexible forms is approximately unity, its precise values are immaterial for present purposes.
(34) D. R. Herschbach, H. S. Johnston, and D. Rapp, *J. Chem. Phys.*, **31**, 1652 (1959).
(35) *The relative free energies of the chair forms of pyranose sugars can be* estimated following a procedure developed by Angyal.[16]
(36) The relative frequency of occurrence in the chain of flexible and chair form conformers cannot be deduced from a knowledge of $\Delta A_{\zeta(p)} - V_{\zeta^*(p)}{}^a$ inasmuch as the effect of neighbor interactions in the chain sequence is ignored in the present considerations. Such neighbor interactions are reflected in the quantity $\Delta A_{\zeta(p)}$ defined in eq 36.
(37) K. D. Goebel and D. A. Brant, *Macromolecules*, **3**, 634 (1970).
(38) D. W. Tanner and G. C. Berry, *J. Polym. Sci., Polym. Phys. Ed.*, **12**, 941 (1974).
(39) K. D. Goebel and D. A. Brant, unpublished results.
(40) P. J. Flory, *Pure Appl. Chem.*, **26**, 309 (1971).