Influence of Aqueous Solvation on Side Chain–Backbone Interaction in Comblike Branched Bacterial Polysaccharides

Bjørn T. Stokke,† Todd A. Talashek,‡ and David A. Brant*

Department of Chemistry, University of California, Irvine, California 92717-2025

Received August 5, 1993; Revised Manuscript Received November 15, 1993®

ABSTRACT: The solution behavior of regularly repeating branched bacterial copolysaccharides has been modeled using an approximate statistical mechanical treatment that accounts for the effects of side chain–backbone interactions on the conformational freedom of the backbone glycosidic linkages. The probability distribution for glycosidic linkage orientations is based on pseudoindependent nearest-neighbor conformational energy surfaces for each backbone linkage, perturbed by the influence of the side chain–backbone interaction. The influence of aqueous solvation has been investigated by using conformational potential energy functions that incorporate strong intrapolymeric hydrogen bonding to simulate weak solvation; stronger aqueous solvation is modeled by reducing the strength of the intramolecular hydrogen bonds. This approach recognizes the geometric specificities of aqueous hydrogen-bonded solvation more realistically than do continuum approaches to solvation but is less computationally intensive than procedures (e.g., molecular dynamics) that incorporate the solvent molecules explicitly. It is shown that variations of the strength of aqueous solvation over an energetically reasonable range can have a marked influence of the solution configuration of the polysaccharide chain. Poor solvation is predicted for one paradigmatic comblike polysaccharide chain to induce an intrinsic bending anisotropy reminiscent of the intrinsic supercoiling of regularly repeating polynucleotides.

I. Introduction

Realistic polymer chain models have been used in conjunction with the statistical mechanical theory of polymer chain configuration to calculate the solution characteristics of several relatively simple polysaccharide chains. Most applications of these techniques have been to linear polysaccharides.[1-5] Treatments of branched carbohydrate polymers, like the regularly repeating, comblike copolysaccharides widely produced by bacteria and fungi,[6] are made more difficult by the need to account for interactions between the side chains and the main chain. Consequently, fewer branched polysaccharides have been modeled, and the models have generally been more primitive.[7,8]

Much current interest focuses on bacterial polysaccharides on the gellan family.[9-19] These polymers share the regularly repeating tetrasaccharide structural unit of the linear polymer, gellan (Figure 1). Except for gellan they display comblike branching with side chains containing one or two sugars. Marked differences in physical behavior, and the potential for commercial exploitation of several members of this group,[20] make these materials interesting for studies of polysaccharide structure–property relationships. Earlier computer modeling of the solution configuration of single-stranded gellan, welan, and rhamsan was based on simple conformational energy functions that included only intramolecular van der Waals and coulombic interactions.[7] The absence in earlier work of potential function terms to account for intrapolymeric hydrogen bonding is equivalent to assuming that such interactions are swamped by strong aqueous solvation of the macromolecule.[1,21,22]

It is evident, however, that interactions of a polysaccharide backbone with relatively flexible side chains will be strongly modulated by the extent to which the backbone and side chain are solvated. In the present work we undertake to describe the role of aqueous solvation by assuming that its dominant component is hydrogen bonding: To the extent that aqueous hydrogen-bonded solvation is effective, intrapolymeric hydrogen-bonded interactions will be suppressed. This simple approach recognizes the geometric and functional group specificities of aqueous solvation more realistically than do continuum approaches,[23-26] but it is much less computationally intensive than would be efforts to account in detail for solvation using a molecular dynamics approach.[27,28] The effects of hydrogen bonding on the mutual interaction energies of components of biopolymers have been incorporated into molecular mechanics potential functions using a variety of functional forms.[22] Here we employ a widely used hydrogen bond potential function of intermediate complexity.[29]

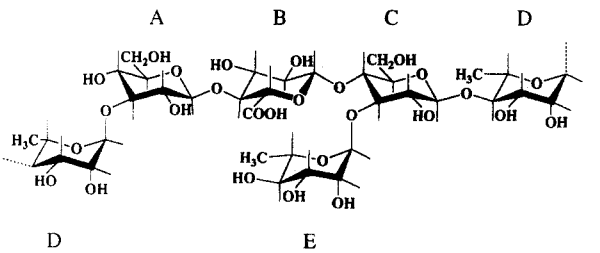

Figure 1. Pentasaccharide repeat unit of welan consisting of residues A (β-D-glucose), B (β-D-glucuronic acid), C (β-D-glucose), D (α-L-rhamnose), and E (α-L-rhamnose (67%) or α-L-mannose (33%), not shown). Gellan lacks the side chain residue E.

We restrict ourselves to a consideration of single-stranded, regularly repeating copolysaccharides, linear or with comblike side chains containing at most one sugar. Specifically we compare the solution characteristics of gellan and its branched congener, welan (Figure 1). Extension of these methods to branched polysaccharides with longer side chains is conceptually straightforward. We make no effort here to model the solution behavior of the multiple-stranded native forms proposed for gellan,[11,12] welan,[17] and other microbial polysaccharides.[30,31] Precedent for the approach adopted can be found not only in our earlier treatments of bacterial polysaccharides[7,8] but also in Cleland's modeling of the effect of the 2′-*N*-acetyl group in hyaluronic acid[32] and a treatment of the influence

† Department of Physics, University of Trondheim, NTH, N-7034 Trondheim, Norway.

‡ Present address: Kelco Division of Merck, 8355 Aero Drive, San Diego, California 92123.

® Abstract published in *Advance ACS Abstracts*, January 15, 1994.

0024-9297/94/2227-1124$04.50/0

of fluctuations in the valence angle at the glycosidic bridge in cellulosic chains.[33] General discussions of the theoretical foundations of the methods used here have also been presented earlier.[22,34]

II. Computational Methods

A. Computation of Mean Properties. We have earlier described methods for computing the configurationally averaged mean-square end-to-end distance,[1] persistence vector,[2] and end-bond correlation function[4] for homopolysaccharides. Extensions of these methods to regularly repeating copolysaccharides[35] and to copolysaccharides of arbitrary[36] or random sequence[37] are also available. These methods are used without further elaboration in calculations pertinent to gellan and welan reported below. The structural notation and conventions employed here are those described earlier.[22]

Requisite for all such calculations is an appropriate description of the configuration energy of the chain,[3,5,22] that is, an expression for the energy of the chain in each of its accessible conformations as a function of the spatial variables that describe those conformations. The configuration energy is used to apply conventional Boltzmann weighting to each of the several conformations contributing to any averaged property. An approximation to the configuration energy appropriate for linear and comblike branched polysaccharides is described in the next section. We use the term *configuration* here to refer to molecular properties averaged over all chain conformations and to the energy of the chain as a whole, required to compute configurational averages.[38] The term *conformation* is used to refer to a particular shape adopted by the chain or by a chain segment. The *conformational energy* is the energy of a segment of the polymer chain, usually a small representative segment such as a disaccharide or the repeating unit of a regularly repeating copolysaccharide.[22,36]

B. Mean-Field Potentials. Most theoretical treatments of the solution configuration of random-coil polysaccharides carried out to data have assumed that the configuration energy of the polymer chain can be expressed as a sum of *independent* terms, one for each independent subunit in the chain, usually taken to be a disaccharide unit. This approximation cannot be justified in all instances.[3,5,22,36] An important case of nonseparability arises for branched polysaccharides. For branched chains, interaction of the side chain with backbone residues closer both to the reducing and nonreducing ends of the main chain serves to couple the contributions to the configuration energy from backbone residues on either side of the side chain.

Shown in Figure 2 is a schematic segment of a comblike, branched polysaccharide. Let this represent the typical structure of a regularly repeating bacterial copolysaccharide. Capital letters designate the sequential order of the backbone residues. Primed letters refer to side-chain residues attached to the main-chain residue designated by the same letter. Only the first sugar of a given side chain is shown, in keeping with the restrictions specified above. There is no presumption concerning the size of the repeating oligosaccharide unit. Depending on the size and composition of the repeating unit, several different letters may refer to the same kind of sugar in Figure 2. Backbone glycosylation is chosen for illustration to occur on every second main-chain residue. Glycosidic torsion angles $\theta_K = (\phi_K, \psi_K)$ govern the mutual orientations of backbone residues K and L. The mutual orientations of backbone residues K and M are controlled by glycosidic

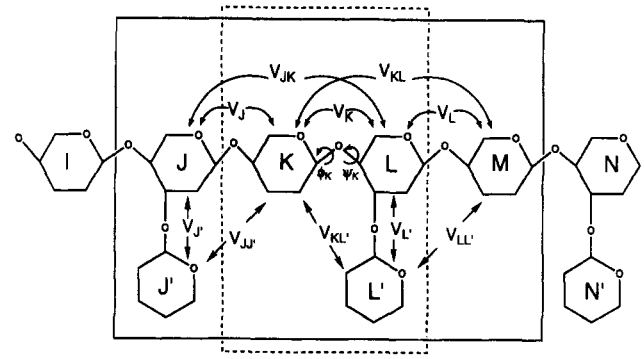

Figure 2. Schematic view of a comblike, branched polysaccharide showing first (V_i) and second (V_{ij}) neighbor interactions making contributions to the configuration energy of the chain. Notation and conventions are described in section II.B of the text.

torsion angles θ_K and $\theta_L = (\phi_L, \psi_L)$. Orientation of the substituent group L′ with respect to backbone residue L is described by $\theta_{L'} = (\phi_{L'}, \psi_{L'})$.

Let $V(\{\theta\})$ represent the conformational energy of the chain segment contained within the window in Figure 2 designated by the solid rectangle. This is the effective potential energy for the hexasaccharide chain segment expressed in terms of the set of glycosidic torsion angles $\{\theta\} = (\theta_J, \theta_K, \theta_L, \theta_{J'}, \theta_{L'})$. In general, the symbol $\{\theta\}$ denotes the complete set of glycosidic torsions for the chain or chain segment under consideration. Description of the segment conformation in terms of this highly contracted set of spatial coordinates may be seen as a consequence of the rigid residue approximation adopted below. More generally, suppression of the remaining variables describing the configuration of the system, i.e., carbohydrate bond lengths and bond angles, other carbohydrate torsion angles, and the coordinates specifying the locations of the solvent atoms, is viewed as a result of having averaged over the generally higher frequency motions of these coordinates for each choice of the principal conformational coordinates $\{\theta\}$.[3,5,22] From either point of view $V(\{\theta\})$ may be chosen, as in the present case, to include a more or less explicit representation of contributions to the conformational energy of the chain segment from the solvent.

Without further approximation, other than pairwise additivity of the contributions to $V(\{\theta\})$, we can write

$$V(\{\theta\}) = V_J + V_K + V_L + V_{J'} + V_{L'} + V_{JK} + V_{KL} + V_{JJ'} + V_{KL'} + V_{LL'} \quad (1)$$

where $V_K(\theta_K)$ represents the energy of interaction of residues K and L, which depends only on θ_K, $V_{KL}(\theta_K,\theta_L)$ represents the energy of interaction of residues K and M, which depends on both θ_K and θ_L, $V_{LL'}(\theta_L,\theta_{L'})$ represents the energy of interaction of residues L′ and M, which depends on both θ_L and $\theta_{L'}$, etc. Singly subscripted *first neighbor* interaction energies depend on the torsion angles of a single glycosidic linkage (or disaccharide unit) and are mutually independent. If these were the only contributions to $V(\{\theta\})$, the configuration energy $V(\{\theta\}) = \sum_i V_i$ of a chain of any length would be fully separable. Doubly subscripted *second neighbor* interaction energies depend simultaneously on the torsion angles of two consecutive glycosidic linkages. They render the conformational energies of two disaccharide units interdependent and, in general, destroy the separability. Interactions of longer range, not included in the present analysis, compound the interdependence.

Note that interdependence arises through V_{KL}, even if the side chain is absent, because this term couples the glycosidic torsions θ_K and θ_L. Moreover, if intrachain

hydrogen bonding is strong and competition exists between interresidue and intraresidue hydrogen bonds, then first neighbor hydrogen-bonded interactions alone can serve to couple successive backbone glycosidic torsions, even if no significant second neighbor interactions occur.[36] In the latter case it is no longer proper to describe the conformation of the chain segment exclusively in terms of glycosidic torsion angles; variables describing the arrangement of hydrogen bonds must also be included. It is thus clear that earlier treatments of the configurational statistics of carbohydrate polymers in the approximation of configuration energy separability require for their validity the absence of significant second neighbor interactions and the absence of strong intrachain hydrogen bonding.[36]

Computation of observable mean properties of oligo- and polysaccharide chains in solution depends on making a properly weighted average over all accessible molecular conformations.[3,5,22,36,38] Typically, averaged products of functions of several glycosidic torsions, e.g., $\langle \cdots f(\theta_J)\, g(\theta_K)\, h(\theta_L) \cdots \rangle$, are required to compute such properties as the mean distance between two atoms of interest, which may be, for example, at the ends of the chain or linked by some scalar or dipolar interaction detectable in an NMR experiment. Here, the angle brackets imply the statistical mechanical average, $f(\theta_J)$ is some function of θ_J, $g(\theta_K)$ is some function of θ_K, etc. When the configuration energy of the chain is separable, the averaged product simplifies immediately to the product of averages over the independent glycosidic linkages:[36,38]

$$\langle \cdots f(\theta_J)\, g(\theta_K)\, h(\theta_L) \cdots \rangle = Z^{-1} \sum_{\{\theta\}} \cdots f(\theta_J)\, g(\theta_K)\, h(\theta_L) \cdots \exp[-\beta \sum_{\{\theta\}} V(\{\theta\})] \quad (2)$$

$$\langle \cdots f(\theta_J)\, g(\theta_K)\, h(\theta_L) \cdots \rangle = \cdots z_J^{-1} \sum_{\theta_J} f(\theta_J) \exp[-\beta V(\theta_J)] z_K^{-1} \sum_{\theta_K} g(\theta_K) \exp[-\beta V(\theta_K)] z_L^{-1} \sum_{\theta_L} h(\theta_L) \exp[-\beta V(\theta_L)] \cdots \quad (3)$$

$$\langle \cdots f(\theta_J)\, g(\theta_K)\, h(\theta_L) \cdots \rangle = \cdots \langle f(\theta_J) \rangle \langle g(\theta_K) \rangle \langle h(\theta_L) \rangle \cdots \quad (4)$$

where $\beta = (RT)^{-1}$, $Z = \sum_{\{\theta\}} \exp[-\beta V(\{\theta\})]$ is the configuration integral for the chain as a whole, and $z_K = \sum_{\theta_K} [-\beta V(\theta_K)]$ is the configuration integral for residue K, etc. Integrals are typically approximated by summations on a regular 10 or 20° grid of rotational isomeric states.[36]

In what follows we employ eq 4 to compute mean properties of branched polysaccharides on the assumption that the doubly subscripted terms in eq 1 represent only a small perturbation on contributions from the singly subscripted terms, so that the configuration energy is "nearly separable". In this approximation the appropriate mean-field potential for averaging functions of θ_K, e.g., $g(\theta_K)$, is given by[32-34,39]

$$\langle V_K(\theta_K) \rangle = -\beta^{-1} \ln \sum_{\substack{\{\theta\} \\ \theta_i \neq \theta_K}} \exp[-\beta V(\{\theta\})] \quad (5)$$

$$\langle V_K(\theta_K) \rangle = V_K(\theta_K) - \beta^{-1} \ln \sum_{\substack{\{\theta\} \\ \theta_i \neq \theta_K}} \exp[-\beta(V_J + V_L + V_{J'} + V_{L'} + V_{JK} + V_{KL} + V_{JJ'} + V_{KL'} + V_{LL'})] \quad (6)$$

$$\langle V_K(\theta_K) \rangle = V_K(\theta_K) + A_{JLJ'L'}(\theta_K) \quad (7)$$

where $A_{JLJ'L'}(\theta_K)$ is a temperature-dependent perturbation of $V_K(\theta_K)$ arising from summation over a discrete set of rotational isomeric states for the torsion angles θ_J, θ_L, $\theta_{J'}$, and $\theta_{L'}$ for a given θ_K.

The potential of mean force $\langle V_K(\theta_K) \rangle$ can be understood as an adjusted version of the first neighbor interaction potential $V_K(\theta_K)$ that includes the effects of other local first and second neighbor interactions, averaged with appropriate statistical weighting over the several rotational isomeric states of θ_J, θ_L, $\theta_{J'}$, and $\theta_{L'}$.[22] This mean-field potential can be used in conjunction with eq 4 in the "nearly separable" formulation to average functions of θ_K with approximate accounting for the perturbations due to nonseparability of $V(\{\theta\})$ imposed by the doubly subscripted contributions $V_{ij}(\theta_i, \theta_j)$.[39] The mean-field potentials for rotations about linkages other than K are calculable in a similar fashion.

The development in eqs 5–7 incorporates the mean field for rotations θ_K due to chain components within the window selected in Figure 2 to be symmetric about θ_K. The size of the window needed for convergence of the perturbation of V_K due to longer range interactions is subject to computational investigation. In our experience a window extending one backbone residue beyond the central disaccharide in each direction is usually sufficient for approximate convergence of the perturbation, and smaller windows are often satisfactory (see below). Notice that the perturbation free energy $A_{JLJ'L'}(\theta_K)$ will grow in magnitude as the window widens due to inclusion of increasing numbers of interactions. It is only the perturbation of the *shape* of the energy surface $V_K(\theta_K)$ that is important, and this should be kept in mind in comparisons of $\langle V_K(\theta_K) \rangle$ based on different perturbation windows. Such comparisons are conveniently made using the normalized conformational probability distributions in θ_K space or, as done here, by referring all energies on a potential energy surface to the energy of the least energetic conformation.

Other well-developed methods exist for dealing with the propagating interdependence of conformational states along the backbone of a branched or (1→2)-linked polysaccharide,[36,38] but these methods fall outside the spirit of the nearly separable formulation under consideration here.

C. Conformational Energy Functions. Torsional, van der Waals, and coulombic contributions to the conformational energy functions employed here are those developed in this laboratory for use with polysaccharides in the rigid residue approximation.[21,22] This approximation acknowledges the dominant role of the "soft" torsions of the glycosidic linkage in determining the shapes of carbohydrate macromolecules.[3,22] Thus, conformational excursions along the stiffer internal coordinates, i.e., bond lengths and bond angles, have only a small amplitude influence on the macromolecular conformation in comparison with glycosidic torsional excursions involving a similar change in energy. Application of these potentials to another copolysaccharide has recently been described in detail.[35]

Hydroxyl hydrogens, HO(p), not involved in hydrogen bonds are treated in a united atom approach, and HO(p) is collapsed into O(p). When contributions to the conformational energies from intramolecular hydrogen bonding are under consideration, explicit representations of the hydroxyl hydrogen positions are required, and the energy depends explicitly on torsion angles in addition to $\{\theta\}$. For example, when hydrogen bonding of a secondary

hydroxyl group HO(3) is considered, the potential function must be formulated to depend explicitly on the exocyclic torsion angle $\chi(3)$ for rotation about the corresponding C(3)–O(3) bond.[21] When the position of hydroxyl hydrogen HO(p) is required, we choose b[O(p)–HO(p)] = 1.02 Å and τ[C(p)–O(p)–H(p)] = 109.43°.

All torsional barriers have been set equal to zero for the present calculations. These are typically on the order of 1 kcal/mol for the C–O bond rotations in question here. This small barrier has a minor effect on the overall conformational energy surface, and eliminating the barrier simplifies computation of the hydrogen-bonding energy (see below). Torsion angles $\chi(5)$ for rotations about a C(5)–C(6) bond are restricted here to the staggered conformations, so the corresponding torsional terms contribute nothing to the conformational energy. No other C–C torsions are permitted within the context of the rigid residue approximation.

For atoms determined to be participants in a hydrogen-bonded interaction the usual van der Waals potential terms are replaced by a special hydrogen bond potential function.[22] Here we employ a hydrogen bond potential function with an adjustable inherent bond strength. A large intrapolymeric hydrogen bond energy is used to simulate an environment that provides weak hydrogen-bonded solvation of the carbohydrate hydroxyl, ether, and carboxyl oxygens. To simulate strong hydrogen-bonded solvation, the intramolecular hydrogen bond energy is diminished or turned off altogether.

The widely used 10–12 potential of McGuire et al.,[29] which involves no explicit dependence on variables describing the linearity of the hydrogen bond, was selected after comparisons (unpublished) showed that it gave results almost indistinguishable from an alignment-dependent hydrogen bond potential function.[21] We have recast the McGuire potential in the usual way to express the energy in terms of the equilibrium O···H distance, ρ, and the depth, ξ, or the potential well.

$$V_{HB} = \xi\left(5\left(\frac{\rho}{r}\right)^{12} - 6\left(\frac{\rho}{r}\right)^{10}\right) \tag{8}$$

Here r is the actual O···H distance. Use of the McGuire potential in the form of eq 8 permits convenient scaling of the hydrogen bond strength and the equilibrium length. We have taken ρ = 1.6 Å to yield an equilibrium O···O distance of 2.62 Å. Potential well depths ξ = −1, −2, and −4 kcal/mol were investigated with the McGuire potential. The choice ξ = −4 kcal/mol yields energies of about −5 kcal/mol for an optimal hydrogen bond once the coulombic contribution is included (see below).

D. Structural Geometry. The deacylated polysaccharides gellan and welan were constructed from monosaccharides following procedures already described.[7] The valence angle at the glycosidic oxygen, $\beta = \tau$[C(1)–O(1)–C(m')], where m' is the number of the carbon at which the glycosidic linkage occurs in the ring closer to the reducing end of the chain, was invariably taken as 117°. For residue B of gellan and welan the geometry of the carboxylate $CO_1O_2^-$ was adapted from the geometry of the formate ion using bond lengths b[C(6)–$O_{1,2}$(6)] = 1.25 Å and valence angles τ[C(5)–C(6)–$O_{1,2}$(6)] = 125.0°.[40]

Conformational energy functions used in the previous work[7] did not differentiate between CH_3, CH_2OH, or COOH at the C(6) position, and explicit representation of the structure at C(6) was therefore not incorporated. Because the hydrogen-bonding ability of these three groups differs, they are represented here as follows: The $C(6)H_3$ groups of α-L-Rha$p$ residues D and E are invariably treated as structureless methyl groups (united atom representation) centered at the location of C(6). The CH_2OH groups are represented as a structureless methylene group localized at C(6) with the position of O(6) specified by the torsional angle $\chi(5) = \theta$[C(4)–C(5)–C(6)–O(6)] = −60°, 60°, and 180°. The two oxygens of the $CO_1O_2^-$ groups are located using the torsional angles $\chi(5_1)$ = −60°, 60°, and 180° and $\chi(5_2) = \chi(5_1) + 180°$.

The position of the HO(6) of a primary hydroxyl was calculated when O(6) was in positions favorable for hydrogen bonding as described below. When O(6) was not in a position for hydrogen bonding, the primary hydroxyl was represented as a structureless hydroxyl group located at O(6). The hydrogen position of a secondary hydroxyl group was computed, as described below, only when that hydroxyl was in a position to hydrogen bond.

E. Treatment of Hydrogen Bonds. Explicit consideration of hydrogen-bonded interactions using the potential of McGuire et al.[29] greatly increases the dimensionality of the rigid residue treatment by adding to $\{\theta\}$ the set of exocyclic torsion angles $\{\chi\}$ that control the hydroxyl orientations. Computation of effective mean-field potentials $\langle V_i(\theta_i)\rangle$ for the several backbone rotations using the formalism of eq 5 requires, in principle, summations over appropriate coordinates from the set $\{\chi\}$ as well as those from the set $\{\theta\}$ already explicitly indicated in eq 5. We adopt here an alternative approach. This involves, in effect, substitution for the averaged energy of the least energetic hydrogen-bonded arrangement for each choice of the coordinates $\{\theta\}$; from the set of variables $\{\chi\}$ we average only over the allowed values of $\chi(5)$ that govern the orientations of the primary hydroxyls and carboxylate oxygens. The requisite search for the preferred hydrogen-bonding patterns yields additional detailed information that would otherwise be lost in the averaging process. In general, in applications of eq 5 to carbohydrates with significant intramolecular hydrogen bonding, one can avoid the need to average over $\{\chi\}$ by using a hydrogen bond potential function that depends only on the O···O distance.[41,42]

For gellan and welan it is easy to show by exploratory conformational energy calculations that only those interactions within the dashed window in Figure 2 contribute significantly to $\langle V_i(\theta)_i\rangle$. Thus, second neighbor interactions along the backbone need not be taken into account. These calculations show, moreover, that the side-chain residue L′, for example, experiences significant second neighbor interactions only with the second neighbor backbone residue to which it is physically closer, namely, K. Choice of the smaller window within which to compute $\langle V_i(\theta)_i\rangle$ appears to be legitimate for many comblike branched polysaccharide structures, and this circumstance greatly simplifies consideration of the possible hydrogen-bonding patterns in the present analysis.

Hydrogen-bonded interactions between a pair of hydroxyl groups involve two possible donor–acceptor choices.[21] Both arrangements were considered for every hydroxyl–hydroxyl interaction. Hydrogen bond energies were computed after rotating the hydrogen of the donor hydroxyl, HO(p), into the plane defined by the donor oxygen, O(p), the acceptor oxygen, O(q), and carbon C(p). Of the two hydrogen positions thus defined, that which placed HO(p) closest to the O(p)···O(q) line was selected.[21] Hydrogen bonds with ether or carboxylate oxygens as the acceptor were constructed with donor hydrogens placed in the same manner. Because all torsional barriers in the potential functions were zero, there was no change in

torsional energy associated with placement of the donor hydrogen in the C(p)–O(p)···O(q) plane. In all instances hydrogen-bonded interactions were considered only when the relevant O(p)···O(q) distance fell in the range 2.1–5.0 Å. Only interresidue hydrogen bonds were considered, because these affect directly the stability of a given conformational state specified by $\{\theta\}$. The possible effects of competition between interresidue and intraresidue hydrogen bonds were thus neglected.

Several hydrogen-bonding arrangements are often possible for a given set of torsion angles $\{\theta\}$ defining the conformation of the molecular segment under consideration. We rank the possible hydrogen bonds in the order of decreasing stability and select from these the least energetic, self-consistent set for a given conformation of the polymer segment. The identity of this polymer segment is dictated by the structure of the polymer and the range of interactions to be included. That is, one must select the window (Figure 2) that contains the linkage in question. After the possible hydrogen bonds have been listed, they are considered in order of decreasing stability against the following criteria until the lowest energy self-consistent set is obtained:

(a) A hydroxyl hydrogen is permitted to participate in no more than one hydrogen bond.

(b) A hydroxyl oxygen is permitted to participate in no more than two hydrogen bonds, once as a donor and once as an acceptor.

(c) Ether and carboxylate oxygens are permitted to participate in only one hydrogen bond, and then as an acceptor.

The combined energies of two or more hydrogen bonds were not considered in competition with those selected by application of these rules. Once selection of the least energetic, self-consistent set of hydrogen bonds is completed for a given $\{\theta\}$, the hydrogen bond energy computed from the McGuire potential is substituted in $V_i(\theta_i)$ or $V_{ij}(\theta_i, \theta_j)$ for the normal van der Waals terms for the atoms involved in the hydrogen bond. The normal coulomb potential, computed on the basis of a united atom approximation for the hydroxyl groups, is replaced by a new coulomb potential computed using the all-atom representation of the hydroxyls employed in the McGuire potential. Because of the large size of the primary hydroxyl and carboxyl groups and their potential for interactions with other parts of the polysaccharide chain, we carry out a statistical mechanical average over the set of torsion angles $\{\chi(5)\}$ for rotations about the C(5)–C(6) bonds of the residues within the window rather than simply selecting the values of $\{\chi(5)\}$ that correspond to the lowest energy hydrogen-bonding arrangement.

III. Conformational Energies

A. Gellan B–C Conformational Energy Surfaces. Conformational energy contour diagrams for the B–C disaccharide of gellan are shown in Figure 3. Energies were computed at 10° intervals from eq 7, using a window that encompasses only residues B and C, i.e., a window analogous to the dashed window in Figure 2. Consequently, no perturbations from second neighbor interactions are included, and $\langle V_B(\theta_B)\rangle = V_B(\theta_B)$. (All gellan energy surfaces are treated here with windows that include just two backbone residues, so in all cases $\langle V_i(\theta_i)\rangle = V_i(\theta_i)$.) The intramolecular hydrogen bond potential is turned off in Figure 3A to simulate strong aqueous solvation. The progression from Figure 3B to 3D corresponds to increasing the strength of the intramolecular hydrogen bonds or, correspondingly, to decreasing the strength of aqueous solvation. As the strength of intramolecular hydrogen bonding increases, the minimum on the potential surface moves from $(\phi_B, \psi_B) = (50°, 0°)$ to (30°, −30°), the absolute energy at the minimum decreases, and the accessible range of conformation space is reduced dramatically.

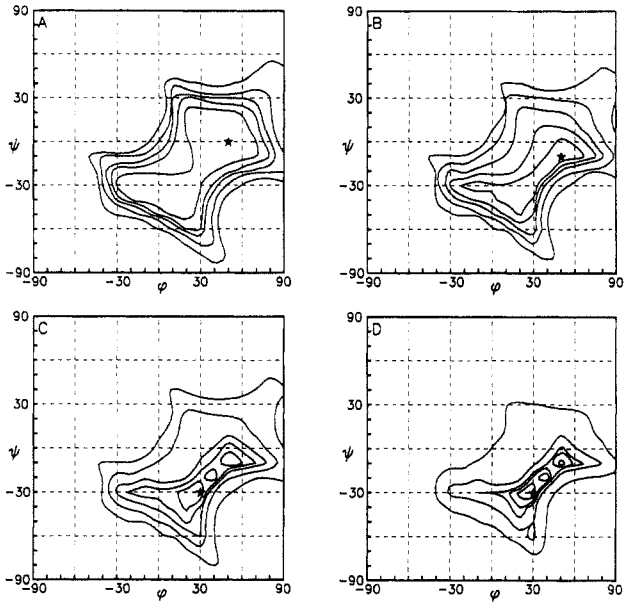

Figure 3. Contour diagrams for the conformational energy $\langle V_B(\theta_B)\rangle = V_B(\theta_B)$ of the B–C disaccharide of gellan without intramolecular hydrogen bonds (A) and with intramolecular hydrogen bond strength $\xi = -1$ (B), −2 (C), and −4 (D) kcal/mol included in the potential functions. Contour lines are drawn at 1, 2, 3, 5, and 10 kcal/mol above the minimum, denoted by a star on each map.

Table 1. The Most Important Hydrogen Bonds in the B–C Disaccharide β-D-Glc*p*A-(1→4)-β-D-Glc*p* of Gellan

				energy gain in kcal/mol for $\xi = -4$ with $\chi(5)_C =$		
ϕ_B	ψ_B	hydrogen bond	length/Å	−60	60	180
30	−30	$O(6)_C \rightarrow O(2)_B$	2.70	−4.76		
30	−30	$O(6)_C \leftarrow O(2)_B$	3.99		−0.43	
30	−30	$O(3)_C \rightarrow O(5)_B$	2.58	−5.21	−5.21	−5.21
40	−20	$O(6)_C \leftarrow O(2)_B$	2.66	−4.95		
40	−20	$O(6)_C \leftarrow O(2)_B$	4.26		−0.33	
40	−20	$O(3)_C \rightarrow O(5)_B$	2.56	−5.02	−5.02	−5.02
50	−10	$O(6)_C \leftarrow O(2)_B$	2.79	−3.72		
50	−10	$O(6)_C \leftarrow O(2)_B$	4.57		−0.23	
50	−10	$O(3)_C \rightarrow O(5)_B$	2.67	−4.94	−4.94	−4.94

Detailed characteristics of the important interresidue B–C hydrogen bonds corresponding to Figure 3D are summarized as a function of (ϕ_B, ψ_B) in Table 1. An arrow indicates the donor → acceptor relationship for each hydrogen bond.[21] The contour diagrams in Figure 3 are averaged with appropriate weighting over $\chi(5)_C$, whereas the individual hydrogen bonds corresponding to the three stable rotational isomeric states of $\chi(5)_C$ are shown in Table 1. The table contains no reference to $\chi(5)_B$ because neither carboxylate oxygen in residue B is involved in significant hydrogen-bonded interactions with residue C. Note, as observed above, that the strongest hydrogen bonds have energies ≈−5 kcal/mol, when $\xi = -4$ kcal/mol in eq 8. The hydrogen bond energies reported in Table 1 represent the decrease in energy computed from the hydrogen bond potential function and revised coulombic potential relative to that computed from the standard van der Waals and coulombic terms. The three local minima evident in the trough in Figure 3D are thought to be artifacts of using a 10° grid.

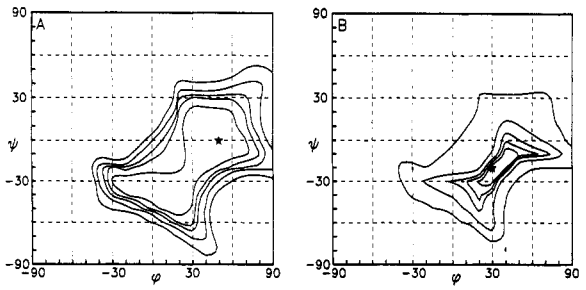

Figure 4. Contour diagrams for the conformational energy $\langle V_A(\theta_A)\rangle = V_A(\theta_A)$ of the A–B disaccharide of gellan without intramolecular hydrogen bonds (A) and with intramolecular hydrogen bond strength $\xi = -4$ kcal/mol (B) included in the potential functions. Contour lines are drawn at 1, 2, 3, 5, and 10 kcal/mol above the minimum, denoted by a star on each map.

Table 2. The Most Important Hydrogen Bonds in the A–B Disaccharide β-D-Glc*p*-(1→4)-β-D-Glc*p*A of Gellan

				energy gain in kcal/mol for $\xi = -4$ with $\chi(5)_B =$		
ϕ_A	ψ_A	hydrogen bond	length/Å	−60°	60°	180°
30	−30	$O_1(6)_B \leftarrow O(2)_A$	2.87	−2.59		
30	−30	$O_1(6)_B \leftarrow O(2)_A$	3.93		−0.60	
30	−30	$O_2(6)_B \leftarrow O(2)_A$	2.87			−2.66
30	−30	$O(3)_B \rightarrow O(5)_A$	2.58	−5.21	−5.21	−5.21
30	−20	$O_1(6)_B \leftarrow O(2)_A$	2.63	−4.14		
30	−20	$O_1(6)_B \leftarrow O(2)_A$	3.64		−0.73	
30	−20	$O_2(6)_B \leftarrow O(2)_A$	2.75			−3.59
30	−20	$O(3)_B \rightarrow O(5)_A$	2.71	−4.63	−4.63	−4.63
40	−20	$O_1(6)_B \leftarrow O(2)_A$	2.86	−2.65		
40	−20	$O_2(6)_B \leftarrow O(2)_A$	3.84		−0.63	
40	−20	$O_2(6)_B \leftarrow O(2)_A$	3.07			−1.71
40	−20	$O(3)_B \rightarrow O(5)_A$	2.56	−5.02	−5.02	−5.02
40	−10	$O_1(6)_B \leftarrow O(2)_A$	2.71	−3.76		
40	−10	$O_2(6)_B \leftarrow O(2)_A$	3.64		−0.74	
40	−10	$O_2(6)_B \leftarrow O(2)_A$	3.03			−1.83
40	−10	$O(3)_B \rightarrow O(5)_A$	2.76	−4.08	−4.08	−4.08

The strongest hydrogen bonds between residues B and C are $O(3)_C \rightarrow O(5)_B$ and $O(6)_C \leftrightarrow O(2)_B$. The former is strongest when $30° \le \phi_B \le 50°$ and $-30° \le \Psi_B \le -10°$, and the latter is particularly important for $\chi(5)_C = -60°$. Both of these hydrogen bonds have been identified in the crystal structure of cellulose II,[43] whereas analysis of the X-ray fiber diffraction data for gellan finds only the $O(6)_C \leftrightarrow O(2)_B$ interaction at distances favorable for hydrogen bonding.[11,12]

B. Gellan A–B Conformational Energy Surfaces. Interactions within the gellan A–B disaccharide differ from those in B–C only because a carboxylate and a primary hydroxyl group at C(6) exchange places relative to the glycosidic linkage (see Figure 1). Accordingly, the contour diagrams for the A–B disaccharide shown in Figure 4 differ very little from the corresponding diagrams in Figure 3. For the A–B disaccharide only the case (Figure 4A) without intramolecular hydrogen bonds (strong solvation) and that (Figure 4B) with strong intramolecular hydrogen bonds (weak solvation) are shown. The negative charge on the carboxylate arising from deprotonation of the carboxylic acid is ignored in calculating the reported conformational energies of all molecular fragments involving residue B. Inclusion of this charge results principally in a lowering of the absolute energy of a surface and produces insignificant perturbations of the shape.

Details of the hydrogen bonding between residues A and B for $\xi = -4$ kcal/mol are shown in Table 2, where the dependence of hydrogen bond strength on the orientation of the carboxylate group ($\chi(5)_B$) is given. The $O(3)_B \rightarrow O(5)_A$ hydrogen bond is strong, as is the analogous $O(3)_C \rightarrow O(5)_B$ bond in the B–C disaccharide, throughout

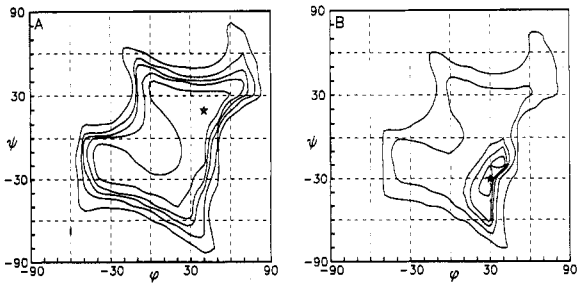

Figure 5. Contour diagrams for the conformational energy $\langle V_D(\theta_D)\rangle = V_D(\theta_D)$ of the D–A disaccharide of gellan drawn as in Figure 4.

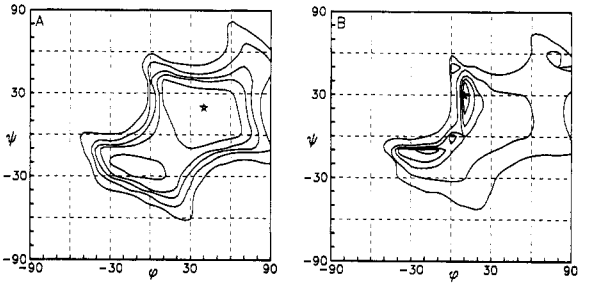

Figure 6. Contour diagrams for the conformational energy $\langle V_C(\theta_C)\rangle = V_C(\theta_C)$ of the C–D disaccharide of gellan drawn as in Figure 4.

the low-energy trough of (ϕ_A, ψ_A) space, and the carboxylate serves as the receptor of a hydrogen bond from $O(2)_A$ within the same domain. Chandrasekaran et al.[12] report the occurrence of both of these hydrogen bonds in fibrous solid-state gellan. The carboxylate orientations $\chi(5)_B = -60°$ and 180° both generate effective receptor positions, whereas in the B–C dimer it is only the primary hydroxyl orientation $\chi(5)_C = -60°$ that is important.

C. Gellan D–A Conformational Energy Surfaces. Iso-energy contour diagrams for the gellan D–A disaccharide are shown for the case of strong solvation in Figure 5A and for weak solvation in Figure 5B. The (1→3) D–A linkage has glycosidic and aglycone bonds oriented respectively *axially* (*a*) and *equatorially* (*e*) with respect to the sugar rings to which they are attached. All other disaccharide units in the gellan backbone repeating unit are linked (1→4) with *e–e* orientation; they consequently have conformational energy surfaces with common features. The gross differences apparent in a comparison of the D–A surfaces in Figure 5 with those of Figures 3, 4, and 6 arise primarily from the difference in linkage orientation. Taken together, the (1→3)-linkage position and the *a–e* linkage orientation of the D–A linkage serve to introduce an abrupt change in the chain direction that can be described effectively using the twist angle defined by Rees and Scott[3,44] (see Figure 1).

Introduction of strong intramolecular hydrogen bonding causes the preferred region of (ϕ_D, ψ_D) conformation space to shift from (40°, 20°) in Figure 5A to (30°, −30°) in Figure 5B. The calculated absolute energies become more negative as a reflection of the hydrogen bond formation, and the minimum energy conformer occurs in a much deeper and more conformationally confined well in Figure 5B. A single hydrogen bond, $O(4)_A \rightarrow O(5)_D$, accounts for the significant perturbation of the D–A energy surface upon introduction of strong intramolecular hydrogen bonding. This bond has a strength of >5 kcal/mol in the deepest parts of the energy well in Figure 5B, as shown in Table 3. The importance of the $O(4)_A \rightarrow O(5)_D$ hydrogen bond in (1→3) *a–e* linkages has been noted earlier.[12,45]

D. Gellan C–D Conformational Energy Surfaces. The conformational energy surfaces for the C–D disac-

Table 3. The Most Important Hydrogen Bonds in the D–A Disaccharide α-L-Rhap-(1→3)-β-D-Glcp of Gellan

ϕ_D	ψ_D	hydrogen bond	length/Å	energy gain in kcal/mol for ξ = −1	−2	−4
20	−40	$O(4)_A \rightarrow O(5)_D$	2.74	−1.89	−2.64	−4.14
30	−40	$O(4)_A \rightarrow O(5)_D$	2.53	−2.21	−3.04	−4.71
20	−30	$O(4)_A \rightarrow O(5)_D$	2.62	−2.29	−3.29	−5.29
30	−20	$O(4)_A \rightarrow O(5)_D$	2.76	−1.90	−2.66	−4.16
40	−20	$O(4)_A \rightarrow O(5)_D$	2.61	−2.31	−3.31	−5.31
40	−10	$O(4)_A \rightarrow O(5)_D$	2.81	−1.72	−2.34	−3.58

Table 4. The Most Important Hydrogen Bonds in the C–D Disaccharide β-D-Glcp-(1→4)-α-L-Rhap of Gellan

ϕ_C	ψ_C	hydrogen bond	length/Å	energy gain in kcal/mol for ξ = −1	−2	−4
−20	−10	$O(3)_D \leftarrow O(2)_C$	2.57	−2.43	−3.40	−5.32
−10	−10	$O(3)_D \rightarrow O(2)_C$	2.67	−2.29	−3.23	−5.11
0	0	$O(3)_D \rightarrow O(2)_C$	2.61	−2.42	−3.42	−5.43
10	10	$O(3)_D \rightarrow O(2)_C$	2.73	−2.10	−2.92	−4.57
10	20	$O(3)_D \rightarrow O(2)_C$	2.67	−2.32	−3.28	−5.19
10	30	$O(3)_D \leftarrow O(2)_C$	2.67	−2.32	−3.28	−5.18

charide are shown for strong and weak aqueous solvation respectively in parts A and B of Figure 6. Although the general features of these maps resemble those in Figures 3 and 4, the perturbation of the strong solvation surface (Figure 6A) due to the introduction of strong intramolecular hydrogen bonds (Figure 6B) is quite different from the corresponding effects seen in Figures 3 and 4. As shown in Table 4, it is the $O(3)_D \leftrightarrow O(2)_c$ hydrogen bond that is strong within the crescent-shaped minimum in Figure 6B. This hydrogen bond does not appear among those that stabilize the preferred conformers on the B–C (Figure 3D) or A–B (Figure 4B) weak solvation maps, and the preferred conformers in Figure 6B occur in a new region of conformation space, i.e., in the range $-20° \leq \phi_C \leq 10°$, $-10° \leq \psi_C \leq 30°$. These differences arise because residue D is L-rhamnose, whereas residues A, B, and C are all D-sugars, and because rhamnose, lacking an OH(6), presents fewer sites for hydrogen bonding to the adjacent residue, C.

E. Welan. Preliminary calculations using both the larger and smaller windows of Figure 2 revealed that potentials of mean force $\langle V_i(\theta_i) \rangle$ for welan are satisfactorily converged even when the smaller window is used (see Section II.E). The conformational energy surfaces for the D–A and A–B disaccharides in welan and gellan are identical in this approximation. The B–C surface is, on the other hand, significantly perturbed by interactions between the side-chain residue E and the backbone. The C–D surface of gellan is potentially subject to perturbation by the side chain E, but effective convergence of the several $\langle V_i(\theta_i) \rangle$ for welan within the smaller window of Figure 2 implies that second neighbor interaction energies of the type $V_{LL'}$ (or $V_{JJ'}$) are small compared to $V_{KL'}$. (Note that $V_{KL'}$ represents the interaction of the side-chain residue L′ with the closer of two second neighbor backbone residues (K), whereas $V_{LL'}$ corresponds to its interaction with the more distant second neighbor backbone residue (M). Clearly, unless $V_{LL'} \ll V_{KL'}$ or $V_{LL'} \gg V_{KL'}$, $\langle V_K(\theta_K) \rangle$ and $\langle V_L(\theta_L) \rangle$ will not be converged using the smaller window.) Indeed, our preliminary calculations show that the interaction of the type $V_{LL'}$ can be neglected altogether in computing $\langle V_C(\theta_C) \rangle$ for welan. Hence, in this approximation, the conformational energy map for the C–D disaccharide segment in welan is identical to that in gellan. This conclusion can perhaps be made more palatable by noting that the trisaccharide segment E–C–D of welan is effectively identical to D–A–B; if second neighbor interactions between backbone residues D and B are negligible,

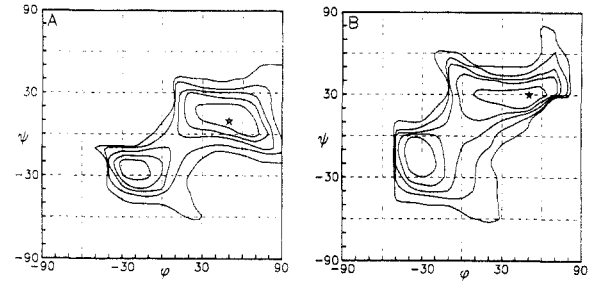

Figure 7. Contour diagrams for the conformational energies $\langle V_B(\theta_B) \rangle$ for the B–C disaccharide (A) and $\langle V_E(\theta_E) \rangle$ for the E–C disaccharide (B) in welan. No intramolecular hydrogen bonds are included, and contour lines are drawn at 1, 2, 3, 5, and 10 kcal/mol above the minimum, denoted by a star on each map.

then second neighbor interactions between sidechain residue E and backbone residue D are also negligible.

The conformational energy surface $\langle V_B(\theta_B) \rangle$ for the B–C disaccharide segment in welan is shown in Figure 7A for the case of strong solvation (no intramolecular hydrogen bonds). Comparison of Figure 7A with the corresponding B–C surface for gellan (Figure 3A) shows the steric intrusion of the side chain E in welan into the otherwise accessible conformational space of the backbone B–C disaccharide in the vicinity of $(\phi_B, \psi_B) = (40°, -20°)$. This is just the region where, in gellan, intramolecular hydrogen bonds $O(2)_B \leftrightarrow O(6)_C$ and $O(3)_C \rightarrow O(5)_B$ are formed with the greatest gain in stability (Figure 2D and Table 1). Interference with the latter hydrogen bond is evidently a direct consequence of glycosylation of gellan at $O(3)_C$ to form welan. The B–C surface of gellan (Figure 3A) is seen in Figure 7A to be split by the presence of the side chain in welan into two predominant minima. This effect reflects the *favorable* interactions of the backbone with the side chain in the vicinity of $(\phi_B, \psi_B) = (-20°, -20°)$ and $(50°, 10°)$. In these regions the absolute energies of the welan surface $\langle V_B(\theta_B) \rangle$ are significantly lower than the corresponding gellan surface.[7]

In a similar fashion, the strong solvation conformational energy surface $\langle V_E(\theta_E) \rangle$ for the disaccharide E–C (Figure 7B) is perturbed by the presence of the residue B in welan in comparison to the identical D–A linkage in gellan and welan (Figure 5A). Steric effects due to backbone residue B exclude the region near $(\phi_E, \psi_E) = (30°, -30°)$ on the E–C map, which is accessible to the D–A linkage. This conformational domain, where the strong hydrogen bond $O(4)_A \rightarrow O(5)_D$ can form if solvation of the polysaccharide is weak (Figure 5B and Table 3), is preempted by glycosylation of residue C at $O(4)_C$, so the analogue of this hydrogen bond is not expected for the E–C linkage. Two predominant minima develop on the E–C map (Figure 7B) at $(\phi_E, \psi_E) = (-30°, -15°)$ and $(50°, 30°)$ as a consequence of favorable interactions between the sidechain residue and the backbone.

The effects on the conformational energy surfaces of the B–C and E–C linkages of welan of allowing strong intramolecular hydrogen bond formation are displayed respectively in parts A and B of Figure 8, which should be compared with the corresponding maps in Figure 7. As just noted, many of the hydrogen bonds between first neighboring residues found in gellan disaccharides B–C and D–A are precluded by glycosylation of residue C at the adjacent sites, $O(3)_C$ and $O(4)_C$. Thus, the deep conformational energy wells near $(\phi_B, \psi_B) = (40°, -20°)$ in Figure 3D and $(\phi_D, \psi_D) = (30°, -30°)$ in Figure 5B do not appear in parts A and B of Figure 8, respectively. The principal hydrogen bonds present in the trisaccharide segment B–C–E are detailed in Table 5. Here it is seen

Table 5. The Most Important Hydrogen Bonds in the B–C–E Trisaccharide of Welan

residue pair	ϕ_B	ψ_B	ϕ_E	ψ_E	hydrogen bond	length in Å with $\chi(5)_C = 60°$ and with $\chi(5)_B =$ −60°	60°	180°	energy gain in kcal/mol for $\xi = -4$
B–C	−20	−20			$O(6)_C \leftarrow O(2)_B$	2.45	2.45	2.45	−4.84
B–E	−20	−20	60	50	$O_2(6)_B \leftarrow O(4)_E$		2.58		−4.46
B–E	−20	−20	70	60	$O_1(6)_B \leftarrow O(4)_E$			2.74	−3.12
B–C	−10	−20			$O(6)_C \leftarrow O(2)_B$	2.65	2.65	2.65	−3.56
C–E			70	50	$O(2)_C \rightarrow O(5)_E$		3.07		−1.63
B–E	−10	−20	70	50	$O_2(6)_B \leftarrow O(3)_E$		2.80		−3.64
C–E			70	60	$O(2)_C \rightarrow O(5)_E$	2.78			−3.68
B–E	−10	−20	70	60	$O_2(6)_B \leftarrow O(4)_E$	2.53			−4.21
B–E	−10	−20	70	60	$O_1(6)_B \leftarrow O(4)_E$			2.53	−4.30
B–E	0	0	70	40	$O_1(6)_B \leftarrow O(4)_E$		2.65		−4.76
B–E	0	0	70	40	$O_2(6)_B \leftarrow O(3)_E$		3.34		−1.03
B–E	20	10	10	30	$O_1(6)_B \leftarrow O(4)_E$			2.57	−4.79
B–E	20	20	20	30	$O_2(6)_B \leftarrow O(4)_E$	2.61			−4.85
B–E	30	20	20	30	$O_2(6)_B \leftarrow O(4)_E$	2.67			−3.00
B–E	30	20	40	30	$O_1(6)_B \leftarrow O(4)_E$		2.65		−3.36

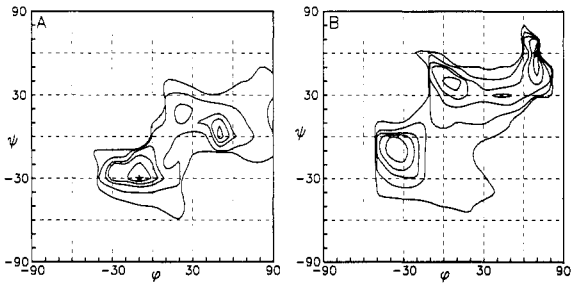

Figure 8. Same as Figure 7 except that strong intramolecular hydrogen bonding ($\xi = -4$ kcal/mol) is included in the potential energy functions.

that the majority of the strong hydrogen bonds form between second neighbor residues B and E, although the bonds $O(2)_B \rightarrow O(6)_C$ and $O(2)_C \rightarrow O(5)_E$ are strong over limited ranges of the low-energy conformational domain. Among the B···E hydrogen bonds, the carboxylic acid oxygens $O_1(6)_B$ and $O_2(6)_B$ accept hydrogen bonds from $O(3)_E$ and $O(4)_E$. Calculations relevant to the B–C–E segment of welan treated residue E exclusively as L-rhamnose. Consequently, the maps presented in parts A and B of Figure 8 are averaged only over $\chi(5)_B$ and $\chi(5)_C$, but not $\chi(5)_E$, and Table 5 contains no entries involving $O(6)_E$, which might be expected to occur when the side chain is L-mannose.

The carboxylic acid functionality of welan is reported to be "screened" or otherwise suppressed in aqueous solution.[10,14,15,18] We find, if the solvation is assumed to be relatively weak, that strong hydrogen bonds can form simultaneously between the carboxylate of residue B and residues A ($O(2)_A \rightarrow O(6)_B$, Table 2) and E ($O(3)_E \rightarrow O(6)_B$ and $O(4)_E \rightarrow O(6)_B$, Table 5). The last of these is also suggested to occur in fibrous, solid-state welan.[17] A propensity of this carboxylate to interact with neighboring residues may be involved in suppression of the polyelectrolyte character of the welan chain. A different set of simultaneous hydrogen bonds involving the carboxylate group has previously been suggested by Crescenzi et al.[46]

Figure 9 shows several low-energy conformations of the welan chain repeating unit, which have been selected to illustrate the simultaneous hydrogen bonding of the carboxylate of residue B to hydroxyls of residues A and E. The glycosidic torsion angles corresponding to the conformers shown are given in the figure legend.

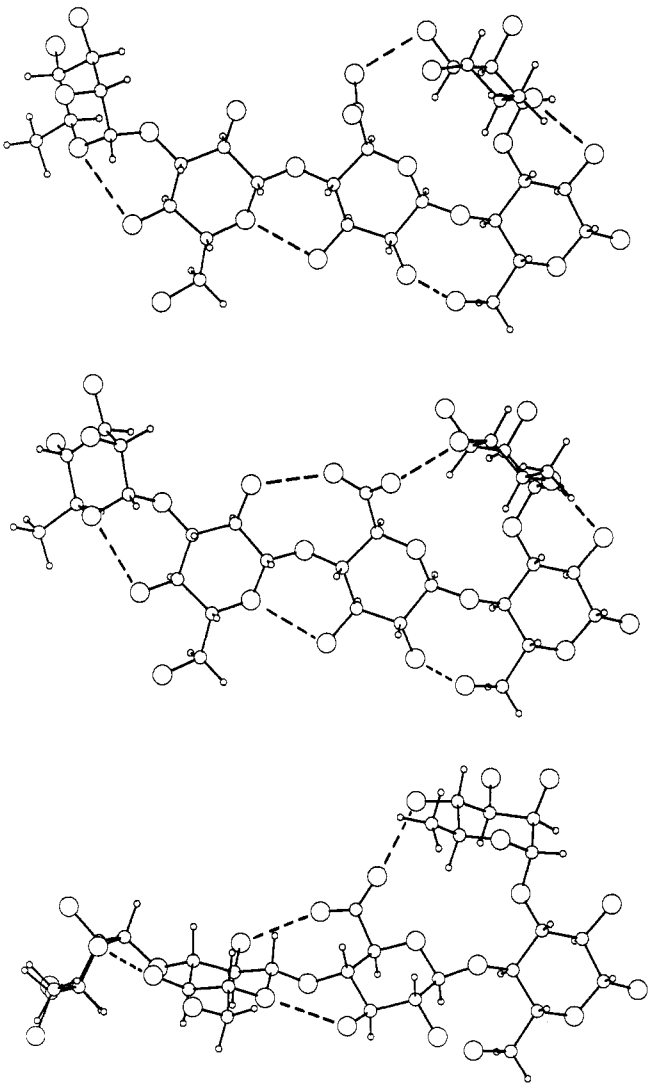

Figure 9. Projections of three low-energy conformations of the welan repeating unit which display simultaneous hydrogen bonding between the carboxylate of residue B and hydroxyls in residues A and E. Residues appear in the order D–A–B–C–E, reading left to right. Hydrogen bonds are shown as dashed lines. The conformations of these repeating units are as follows:

	ϕ_D	ψ_D	ϕ_A	ψ_A	ϕ_B	ψ_B	ϕ_E	ψ_E	$\chi(5)_B$
top	10	−40	30	−30	−10	−20	70	50	60
middle	20	−30	30	−20	−10	−20	70	60	−60
bottom	40	−20	40	−10	20	20	20	30	−60

IV. Chain Configuration

A. Characteristic Ratios. The influence of aqueous solvation on the solution configuration of gellan and welan chains can be investigated by examining the calculated mean-square end-to-end distances of these polymers as a

Table 6. Characteristic Ratio C_n As a Function of the Number of Backbone Residues n for Gellan and Welan

	C_n for gellan with ξ =				C_n for welan with ξ =			
n	0	−1	−2	−4	0	−1	−2	−4
4	3.6	3.6	3.6	3.7	3.6	3.6	3.6	3.6
8	6.1	6.1	6.0	6.0	6.1	6.2	6.2	6.1
16	9.5	8.2	7.7	7.3	9.8	8.9	7.9	7.1
32	13.4	10.1	9.6	9.8	14.4	11.1	7.6	5.7
64	16.5	11.7	11.6	13.3	18.3	12.6	7.8	5.7
128	18.3	12.7	12.9	16.4	20.7	13.5	7.8	5.6
256	19.2	13.1	13.6	18.3	21.9	13.9	7.9	5.6
512	19.6	13.4	13.9	19.2	22.5	14.1	7.9	5.6
1024	19.8	13.5	14.1	19.7	22.8	14.2	7.9	5.6
2048	20.0	13.5	14.2	19.9	22.9	14.3	7.9	5.6
4096	20.0	13.5	14.2	20.0	23.0	14.3	7.9	5.6

function of the assumed strength of the solvation. Here mean-square end-to-end distance will be represented by the dimensionless characteristic ratio, $C_n = \langle r^2 \rangle_0 / nL^2$, where the unperturbed mean-square end-to-end distance, $\langle r^2 \rangle_0$, is normalized by the product of the number of backbone sugar residues, n, and the mean-squared length, L^2, of the vector spanning each residue. We have computed C_n as described earlier.[35,36] Computed values of C_n are presented in Table 6 as a function of n for gellan and welan, assuming strong aqueous solvation (no intramolecular hydrogen bonds, $\xi = 0$) and for various levels of diminished solvation ($\xi = -1, -2$, and -4 kcal/mol). It is apparent from Table 6 that C_n has effectively converged to its high chain-length asymptote for $n = 4096$; the notation C_∞ refers to this asymptotic behavior.

We observe first, as reported previously,[7] that introduction of the welan side chain under conditions of strong solvation ($\xi = 0$), i.e., substitution of the conformational energy map in Figure 7A for that in Figure 3A, is predicted to have little effect on the chain dimensions. Thus, the well-solvated side chain appears not to disturb significantly the configuration of the well-solvated backbone, as measured by C_n. It is not correct to conclude, however, that there is no interaction of side chain and backbone. The side chain intrudes significantly into the conformation space of the B–C linkage, and the potential functions produce, furthermore, attractive interactions of side chain with backbone in the regions of the two dominant minima on the B–C surface in Figure 7A. Absence of influence of the side chain on C_n is, in this case, essentially accidental and cannot be assumed to occur generally in branched polysaccharides when the chain is strongly solvated.[8]

As the potential functions are modified to represent weaker solvation of the chain, the calculated C_∞ passes through a minimum for gellan and declines monotonically for welan. If the strength of the intramolecular hydrogen bonds were permitted to increase without limit, the conformational energy surface for each linkage could be expected to develop one dominant minimum. In this situation the chain would be locked effectively into a single, helical conformation for which C_n would be directly proportional to n. We thus anticipate that for both gellan and welan C_∞ should ultimately increase as the solvation becomes progressively poorer. Only the improbable occurrence of a least energetic helical form with zero pitch or the occurrence of two or more deep minima of equal or nearly equal energy on one or more of the backbone linkage energy surfaces would alter these expectations for C_∞ in a poorly solvating medium. It may be assumed therefore that C_∞ for gellan would continue to increase and that C_∞ for welan would pass through a minimum, if ξ were allowed to become yet more negative. More negative values of ξ, however, would seem quite unrealistic for aqueous media.

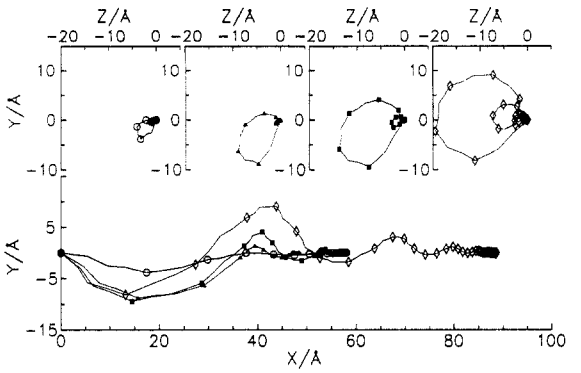

Figure 10. Projections into the xy (lower panel) and yz (upper panel) of the terminus of the persistence vector $\langle \mathbf{r}_n \rangle_0$ of gellan as a function of increasing numbers of backbone residues n. These points are connected to depict the mean chain trajectory. Oxygen $O(4)_C$ at the nonreducing chain end is arbitrarily placed at the origin of coordinates; the coordinate system is described in section IV.B. of the text. Symbols are attached to every $O(4)_C$ to define the chain repeating unit and to designate calculations without intramolecular hydrogen bonds (○) and with intramolecular hydrogen bond strength $\xi = -1$ (▲), -2 (■), and -4 (◇) kcal/mol included in the potential functions.

B. Persistence Vectors. The initial reduction in C_∞ calculated for both gellan and welan with increasing magnitude of ξ may be understood to result either from (1) a generally increased conformational freedom of one or more of the backbone linkages or (2) an increased propensity of the chain to adopt nonrectilinear trajectories. The first of these possibilities is inconsistent with the general effects of allowing for increasingly strong intramolecular hydrogen bonding, which invariably generates deeper and more precipitous minima on the conformational energy surface than are present on the corresponding surfaces in the absence of intramolecular hydrogen bonding.

The consequence of intramolecular hydrogen bonding for the character of the preferred chain trajectories can be investigated by examining the persistence vector, defined for a chain of length n as the unperturbed mean vector, $\langle \mathbf{r}_n \rangle_0$, from one end of the chain to the other. Its magnitude a_n is the persistence length of the chain, a measure of chain stiffness most commonly associated with the infinite chain (a_∞). The persistence vector is usually expressed in a coordinate system attached rigidly to one end of the chain in question. For present purposes we choose to express $\langle \mathbf{r}_n \rangle_0$ in a right-handed Cartesian coordinate system with its x axis lying along the persistence vector of the chain of *infinite* length and with its origin at the nonreducing chain terminus. The persistence vector has been calculated here using methods already described.[2,38]

If we plot projections of the terminus of $\langle \mathbf{r}_n \rangle_0$ as a function of n and connect the resulting points, projections of the mean trajectory of the infinitely long chain result. Despite its character as the mean over a distribution that may be rather broad, a plot of the mean trajectory can convey some sense of the shape adopted by the polymer chain. Such plots appear in Figure 10 for gellan and in Figure 11 for welan. The lower panel displays projections into the xy plane for $\xi = 0, -1, -2$, and -4 kcal/mol, while the upper panels display projections into the yz plane separately for the several values of ξ. Data corresponding to the calculated persistence vectors for chains consisting of a few repeating units are presented in Table 7 along with the persistence vectors for the infinitely long chains.

The mean chain trajectory must converge to a point as n increases, because of the loss of directional correlation with respect to a fixed vector at the origin of coordinates.

Table 7. Persistence Length a_n and Components of the Persistence Vector as a Function of Degree of Polymerization[a] for Gellan and Welan

		gellan with ξ =				welan with ξ =			
DP	b	0	−1	−2	−4	0	−1	−2	−4
1	a_n	18.115	18.575	18.852	19.142	18.115	18.575	18.852	19.142
	$\langle x \rangle$	17.401	15.670	14.458	13.173	17.518	16.314	14.520	12.866
	$\langle y \rangle$	−3.848	−8.821	−9.452	−8.189	−2.519	−6.860	−10.054	−11.023
	$\langle z \rangle$	−3.246	−4.656	−7.549	−11.218	−3.865	−5.641	−6.593	−8.909
2	a_n	30.090	31.075	32.275	33.656	30.161	31.642	32.746	34.029
	$\langle x \rangle$	29.789	29.087	28.581	27.341	29.990	30.321	29.943	29.474
	$\langle y \rangle$	−1.345	−6.328	−5.917	−2.276	0.083	−4.162	−8.297	−9.218
	$\langle z \rangle$	−4.030	−8.920	−13.778	−19.495	−3.204	−8.034	−10.259	−14.292
3	a_n	37.675	37.178	39.064	41.802	38.810	38.825	39.647	41.716
	$\langle x \rangle$	37.615	36.422	37.249	37.833	38.780	38.374	38.618	39.768
	$\langle y \rangle$	−0.021	−0.896	1.341	6.885	−0.018	−0.490	−2.322	−1.214
	$\langle z \rangle$	−2.133	−7.404	−11.693	−16.392	−1.511	−5.879	−8.670	−12.539
4	a_n	43.229	39.602	41.486	45.318	45.614	42.177	40.411	41.792
	$\langle x \rangle$	43.213	39.425	40.922	43.789	45.593	42.072	40.106	40.969
	$\langle y \rangle$	−0.337	1.358	4.084	9.086	−0.384	0.763	2.173	5.293
	$\langle z \rangle$	−1.106	−3.487	−5.462	−7.327	−1.322	−2.882	−4.453	−6.336
5	a_n	47.479	41.221	43.066	48.080	50.701	44.022	38.108	37.214
	$\langle x \rangle$	47.467	41.204	42.999	47.863	50.689	44.003	37.966	36.617
	$\langle y \rangle$	−0.471	0.583	1.962	4.219	−0.159	0.420	3.190	6.626
	$\langle z \rangle$	−0.964	−1.039	−1.362	−1.733	−1.102	−1.195	−0.813	−0.417
10	a_n	56.036	47.639	53.733	69.895	62.019	48.485	37.274	36.525
	$\langle x \rangle$	56.036	47.638	53.723	69.803	62.019	48.485	37.270	36.476
	$\langle y \rangle$	−0.063	−0.025	0.624	2.736	−0.036	0.009	−0.281	−0.580
	$\langle z \rangle$	−0.193	−0.148	−0.860	−2.320	−0.233	−0.206	−0.508	−1.803
15	a_n	57.634	48.351	56.393	79.856	64.622	49.022	36.984	34.851
	$\langle x \rangle$	57.634	48.351	56.392	79.830	64.622	49.022	36.984	34.849
	$\langle y \rangle$	−0.011	−0.000	0.115	1.161	−0.008	−0.002	−0.042	−0.353
	$\langle z \rangle$	−0.035	−0.033	−0.330	−1.700	−0.054	−0.026	−0.028	0.079
∞	a_∞	58.003	48.486	57.332	88.712	65.396	49.098	37.083	35.422
	$\langle x \rangle$	58.003	48.486	57.332	88.712	65.396	49.098	37.083	35.422
	$\langle y \rangle$	0	0	0	0	0	0	0	0
	$\langle z \rangle$	0	0	0	0	0	0	0	0

[a] Degree of polymerization (DP) refers in this table to the number of repeating units in the chain. [b] All persistence lengths and components expressed in angstrom units.

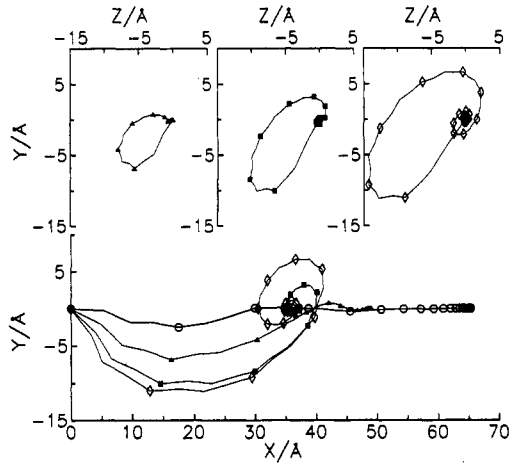

Figure 11. Same as Figure 10 but for welan. The yz projection for the chain in the absence of hydrogen bonds is omitted, because it closely resembles the corresponding projection for gellan.

This effect is clearly seen in Table 7 and in Figures 10 and 11, where the symbols that delimit the four-sugar backbone repeating units are observed to converge as n increases. Calculations (not presented here) of the end-bond correlation functions[4,35] for gellan and welan show that for the case of strong aqueous solvation ($\xi = 0$) the respective persistence vectors have effectively converged after 40 backbone residues or 10 repeating units. The corresponding persistence lengths a_∞ are about 60 Å as may be seen from the lower panels of Figures 10 and 11 and from Table 7. Convergence is somewhat slower as the magnitude of ξ increases.

For gellan a_∞ contracts somewhat for $\xi = -1$ and then increases as ξ becomes more negative (Figure 10 and Table 7). Positive and negative excursions in the y coordinate become more pronounced as the magnitude of ξ increases (Figure 10, bottom panel). This reflects an increased tendency of the chain to propagate in a helical trajectory, the mean radius of which increases as ξ becomes more negative. The amplitude of these oscillations in y necessarily declines with increasing contour distance from the nonreducing terminus of the chain, where the coordinate system is anchored, due to the growing loss of directional correlation. The periodicity of the helicity predicted to develop in poorly solvating media is about six repeating units or 24 backbone residues. This is to be compared with the 12 residue repeat characteristic of the gellan[11,12] and welan[17] chains in the fibrous double helix.

The increased helical tendencies predicted to develop with diminished solvation are readily seen for gellan in the yz projections at the top of Figure 10. Because the x axis lies along the persistance vector for the infinite chain, both ends of the mean infinite chain trajectory lie at $y = z = 0$ Å in these projections. Portions of the projected mean trajectories nearer the origin or nonreducing chain end are characterized by a larger spacing between the symbols marking the repeating unit structure. It is interesting to observe that the cross sections of helical gellan mean trajectories are roughly circular. Note, however, that essentially all of the excursion in the arbitrarily chosen z direction is toward negative values of z. Thus, the calculated mean chain trajectory of the single-stranded gellan chain is not disposed symmetrically with respect to the mean end-to-end vector of the infinite chain, and in this sense it may be said that the mean trajectory is not only helical but curved. This curvature bespeaks a "bending anisotropy" of the chain reflected in the absence

of symmetry about the x axis seen in the yz projection of $\langle \mathbf{r}_n \rangle_0$. Such curvature is a property of the flexible gellan chain. If the torsion angles at each glycosidic linkage are restricted to their least energetic conformations, a rigid, regular helix with a rectilinear helical axis is generated.

It is also useful to notice that whereas $C_\infty = 20$ both for $\xi = 0$ and -4 kcal/mol (Table 6), the value of a_∞ is considerably larger for $\xi = -4$ kcal/mol than it is for $\xi = 0$ (Table 7). Elementary considerations make clear that these two measures of chain extension are related in a complicated way through the details of the mean spatial distribution of the polymer chain.[47] In the present instance a_∞ is defined to depend only on the mean x component of the infinite chain terminus, whereas C_∞ depends on all three mean-square components. Consequently, C_∞ is invariably augmented by increases in the breadth of the spatial distribution. That C_∞ is as large for gellan with $\xi = 0$ as when $\xi = -4$ kcal/mol thus suggests a broader spatial distribution for the well-solvated polymer coil ($\xi = 0$), and this is clearly consistent with the greater conformational freedom revealed in the corresponding conformational energy surfaces.

The welan chain is predicted to respond to diminished solvation in a similar way, but the maximum amplitude and disposition of oscillations in the xy and yz planes are different from those for gellan (Figure 11 and Table 7). We observe initially that, for the range of ξ investigated, a_∞ declines monotonically as the magnitude of ξ increases. In the limit of a perfectly rigid helical chain, which would presumably result from further increases in the strength of the intramolecular hydrogen bonds, both C_∞ and a_∞ must approach infinity. Whereas both C_∞ and a_∞ pass through a minimum in the range $0 \leq \xi \leq -4$ kcal/mol for gellan, this minimum does not appear to have been reached for welan in the same range of ξ.

In poorly solvating media ($\xi = -2$ and -4 kcal/mol) the welan chain is predicted to be effectively more compact than that of gellan (Tables 6 and 7), not because the amplitudes of oscillations in the yz plane are larger than for gellan, but because the mean trajectory of welan also displays oscillations in the x direction which dramatically reduce the net chain extension that is achieved before directional correlations are lost. Oscillations in yz describe for welan a mean helical cross section that is more ellipsoidal than circular. (The yz projection for welan with $\xi = 0$ is omitted from the upper panel of Figure 11, because it so closely resembles the corresponding panel for gelan.) These excursions are again not symmetric with respect to the persistence vector and thus tend to impart net curvature as well as helicity to the mean trajectory. For welan this curvature is predicted in poorly solvating media to have a looping character that is occasioned by the oscillations in x, i.e., in the mean end-to-end direction, seen in the lower panel of Figure 11. These oscillations in x for $\xi = -2$ and -4 kcal/mol are sufficiently large to produce oscillations in a_n with increasing n (Table 7). This implies that for certain ranges of n the mean chain trajectory terminates closer to the origin as n increases. The same effect appears in muted form in C_n (Table 6).

The curvature predicted for the mean trajectories of gellan and welan chains dissolved in poorly solvating media is closely akin to that described for certain regularly repeating polynucleotides.[48] The oscillatory behavior of a_n as a function of n predicted for welan in poorly solvating media has been described as intrinsic supercoiling in the DNA literature.[48] The nearly separable approximation used here to model the aqueous solution behavior of gellan and welan is equivalent to the use of locally independent energy surfaces in modeling the intrinsic supercoiling of the regularly repeating polynucleotides.[49]

V. Summary

It is shown here that the configurational statistics of carbohydrate polymers, especially branched bacterial polysaccharides, can be expected to depend significantly on the strength of aqueous solvation in aqueous solution. This finding has important consequences for efforts to predict oligo- and polysaccharide conformation using typical molecular mechanics potential functions. These functions generally represent aqueous solvation in a very rudimentary way, if at all. They are frequently used without much attention to the important effects that solvation may have on the resulting conclusions. Aqueous solvation is modeled here in terms of the strength attributed to hydrogen bonds within the carbohydrate polymer. This approach has the advantage over continuum treatments of solvation in that the geometric specificities of hydrogen-bonded water–sugar interactions can be taken approximately into account. An approximate statistical mechanical method is developed for dealing with the interdependence of rotations at successive glycosidic linkages that results from the presence of short side chains on many polysaccharides of bacterial origin. It is shown, by including in the potential energy functions intramolecular hydrogen bonding with a strength quite plausible for aqueous solution, that one regularly repeating, comblike branched bacterial polysaccharide might be expected to display curved, supercoiling mean chain trajectories that reflect an intrinsic bending anisotropy for the chain. That intramolecular hydrogen bonding may play an important role in determining the aqueous solution conformation of carbohydrates is suggested by the marked propensity of many carbohydrate polymers, including gellan and welan, to occur in nature in multiple-stranded forms.

Acknowledgment. This work has been supported by NIH Research Grant GM 33062 to D.A.B. and by Royal Norwegian Council for Scientific and Industrial Research Grant 2823 to B.T.S.

References and Notes

(1) Brant, D. A.; Dimpfl, W. L. *Macromolecules* **1970**, *3*, 644–654.
(2) Jordan, R. C.; Brant, D. A.; Cesàro, A. *Biopolymers* **1978**, *17*, 2617–2632.
(3) Brant, D. A. In *The Biochemistry of Plants*; Preiss, J., Ed.; Academic Press: New York, 1981; Vol. 3; pp 425–472.
(4) Burton, B. A.; Brant, D. A. *Biopolymers* **1983**, *22*, 1769–1792.
(5) Brant, D. A.; Christ, M. D. In *Computer Modeling of Carbohydrate Molecules*; French, A. D., Brady, J. W., Eds.; American Chemical Society: Washington, DC, 1990; Vol. 430; pp 42–68.
(6) Kenne, L.; Lindberg, B. In *The Polysaccharides*; Aspinall, G., Ed.; Academic Press: New York, 1983; Vol. 2; pp 287–363.
(7) Talashek, T. A.; Brant, D. A. *Carbohydr. Res.* **1987**, *160*, 303–316.
(8) Brant, D. A.; Talashek, T. A. In *Industrial Polysaccharides: The Impact of Biotechnology and Advanced Methodologies*; Stivala, S. S., Crescenzi, V., Dea, I. C. M., Eds.; Gordon and Breach Science Publishers: New York, 1987; pp 229–253.
(9) Grasdalen, H.; Smidsrød, O. *Carbohydr. Polym.* **1987**, *7*, 371–393.
(10) Crescenzi, V.; Dentini, M.; Dea, I. C. M. *Carbohydr. Res.* **1987**, *160*, 283–302.
(11) Chandrasekaran, R.; Puigjaner, L. C.; Joyce, K. L.; Arnott, S. *Carbohydr. Res.* **1988**, *181*, 23–40.
(12) Chandrasekaran, R.; Millane, R. P.; Arnott, S.; Atkins, E. D. T. *Carbohydr. Res.* **1988**, *175*, 1–15.
(13) Talashek, T. A.; Brant, D. A. In *Frontiers in Carbohydrate Research-1: Food Applications*; Millane, R. P., BeMiller, J. N., Chandrasekaran, R., Eds.; Elsevier Applied Science Publishers: London, 1989; pp 271–288.

(14) Urbani, R.; Brant, D. A. *Carbohydr. Polym.* **1989**, *11*, 169–191.
(15) Campana, S.; Andrade, C.; Milas, M.; Rinaudo, M. *Int. J. Biol. Macromol.* **1990**, *12*, 379–384.
(16) Milas, M.; Shi, X.; Rinaudo, M. *Biopolymers* **1990**, *30*, 451–464.
(17) Lee, E. J.; Chandrasekaran, R. *Carbohydr. Res.* **1991**, *214*, 11–24.
(18) Campana, S.; Ganter, J.; Milas, M.; Rinaudo, M. *Carbohydr. Res.* **1992**, *231*, 31–38.
(19) Crescenzi, V.; Dentini, M.; Coviello, T. In *Frontiers in Carbohydrate Research*—2; Chandrasekaran, R., Ed.; Elsevier Applied Science Publishers: London, 1992; pp 100–114.
(20) Moorhouse, R. *Prog. Biotechnol.* **1987**, *3*, 187–206.
(21) Goebel, C. V.; Dimpfl, W. L.; Brant, D. A. *Macromolecules* **1970**, *3*, 644–654.
(22) Brant, D. A. *Quart. Rev. Biophys.* **1976**, *9*, 527–596.
(23) Tvaroska, I.; Kozar, T. *J. Am. Chem. Soc.* **1980**, *102*, 6929–6936.
(24) Tvaroska, I. *Biopolymers* **1982**, *21*, 1887–1897.
(25) Tvaroska, I.; Imberty, A.; Perez, S. *Biopolymers* **1990**, *30*, 369–379.
(26) Urbani, R.; Cesàro, A. *Polymer* **1991**, *32*, 3013–3020.
(27) Brady, J. W. *J. Am. Chem. Soc.* **1989**, *111*, 5155–5165.
(28) Ha, S.; Gao, J.; Tidor, B.; Brady, J. W.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 1553–1557.
(29) McGuire, R. F.; Momany, F. A.; Scheraga, H. A. *J. Phys. Chem.* **1972**, *76*, 375–393.
(30) Norisuye, T.; Yanaki, T.; Fujita, H. *J. Polym. Sci., Polym. Phys. Ed.* **1980**, *18*, 547–558.
(31) Stokke, B. T.; Smidsrød, O.; Elgsaeter, A. *Biopolymers* **1989**, *28*, 617–637.
(32) Cleland, R. L. *Biopolymers* **1971**, *10*, 1925–1948.
(33) Goebel, K. D.; Harvie, C. E.; Brant, D. A. *Appl. Polym. Symp.* **1976**, *28*, 671–691.
(34) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 1391–1436.
(35) Buliga, G. S.; Brant, D. A. *Int. J. Biol. Macromol.* **1987**, *9*, 77–86.
(36) Brant, D. A.; Goebel, K. D. *Macromolecules* **1975**, *8*, 522–530.
(37) Buliga, G. S.; Brant, D. A.; Fincher, G. B. *Carbohydr. Res.* **1986**, *157*, 139–156.
(38) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley-Interscience: New York, 1969.
(39) Talashek, T. A. Ph.D. Thesis, University of California, Irvine, 1988.
(40) *Tables of Interatomic Distances and Configuration in Molecules and Ions*; Bowen, H. J. M., Donahue, J., Jenkin, D. G., Kennard, O., Wheatley, P. J., Whiffen, D. A., Eds.; The Chemical Society: London, 1958.
(41) Blackwell, J.; Sarko, A.; Marchessault, R. H. *J. Mol. Biol.* **1969**, *42*, 379–383.
(42) Pérez, S.; Vergelati, C. *Int. J. Biol. Macromol.* **1987**, *9*, 211–218.
(43) Stipanovic, A. J.; Sarko, A. *Macromolecules* **1976**, *9*, 851–857.
(44) Rees, D. A.; Scott, W. E. *J. Chem. Soc. B* **1971**, 469–479.
(45) Sundaralingam, M. *Biopolymers* **1968**, *6*, 189–213.
(46) Crescenzi, V.; Dentini, M.; Coviello, T. In *Industrial Polysaccharides: The Impact of Biotechnology and Advanced Methodologies*; Stivala, S. S., Crescenzi, V., Dea, I. C. M., Eds.; Gordon and Breach Science Publishers: New York, 1987; pp 69–97.
(47) Yevich, R.; Olson, W. K. *Biopolymers* **1979**, *18*, 113–145.
(48) Maroun, R. C.; Olson, W. K. *Biopolymers* **1988**, *27*, 585–603.
(49) Maroun, R. C.; Olson, W. K. *Biopolymers* **1988**, *27*, 561–584.