Carbohydrate Research, 248 (1993) 55–80
Elsevier Science Publishers B.V., Amsterdam

The conformations of cyclic (1 → 2)-β-D-glucans: Application of multidimensional clustering analysis to conformational data sets obtained by Metropolis Monte Carlo calculations

William S. York *, Jan U. Thomsen and Bernd Meyer

Complex Carbohydrate Research Center, The University of Georgia, 220 Riverbend Road, Athens, GA 30602 (USA)

(Received March 17th, 1993; accepted May 3rd, 1993)

ABSTRACT

Sets containing up to 1.3×10^6 energetically accessible conformations of linear (1 → 2)-β-D-glucan oligosaccharides were obtained by Metropolis Monte Carlo (MMC) calculations performed with the GEGOP (GEometry of GlycOProteins) program. Quantitative analyses of the data sets (which were expressed in terms of the glycosidic dihedral angle coordinates) were obtained by two different clustering methods: (*i*) the three-distance hierarchical clustering method (3-DM), published by Jure Zupan, and (*ii*) a nonhierarchical clustering method (Population-Density Projection, PDP) which, through a segmentation analysis of two-dimensional projections of the population-density surface, establishes a partitioning of conformational space into a set of "cluster regions", followed by a clustering step where each conformation of the data set is assigned to one of these regions. Computer programs (MCLUST and PDPCLUST) were developed to perform the 3-DM and PDP analyses, respectively. The two types of analysis provided very similar sets of conformational families (clusters), which could be expressed as combinations of distinct conformations of the glycosidic torsional angles (ϕ, ψ) centered at (50°, 10°) for conformation A, (40°, 160°) for conformation B, (55°, −160°) for conformation B′, and (170°, 10°) for conformation C. The analysis provided the populations of the families, along with relative rates for transitions between families. Examination of the frequencies of the A, B, and C glycosidic bond conformations with respect to their relative positions in the sequence revealed the tendency of the (1 → 2)-β-D-glucan to adopt conformational repeating structures of the general form $[A_nB]$, where $n = 3$ or 6. These repeating structures combine in an energetically cooperative fashion to give low-energy cyclic conformations having, for example C_5 symmetry $[AAAB]_5$ for the eicosamer, and C_3 symmetry $[AAAAAAB]_3$ for the heneicosamer.

INTRODUCTION

Cyclic (1 → 2)-β-D-glucans produced by gram-negative bacteria are thought to play an important role in adapting the bacteria to changes in environmental osmotic pressure by regulating the osmotic balance between the cytoplasm and the

* Corresponding Author.

periplasmic space[1]. The importance of the cyclic structure[2,3] of these glucans in terms of their biological function is, as of yet, unknown. Nevertheless, these molecules present a unique challenge in understanding the relationships between the solution conformations of complex carbohydrates and their physical and biological properties. This paper describes one approach to this problem based on cluster analysis of large sets of molecular conformations obtained by Metropolis Monte Carlo (MMC) calculations performed with the GEGOP[4] (GEometry of GlycOProteins) program.

Palleschi and Crescenzi[5] have proposed a model for the cyclic (1 → 2)-β-D-glucans based on the idea of conformationally equivalent triglucosyl repeat units in which the torsional angles ϕ = (H-1–C-1–O-1–C2) and ψ = (C-1–O-1–C-2–H-2) for the glycosidic bonds are the same for every third glycosyl linkage (i.e., $\phi = \phi_{i+3}$, $\psi_i = \psi_{i+3}$). More highly symmetric conformations of the (1 → 2)-β-D-glucans (where the ϕ and ψ do not vary from one glycosyl residue to the next) are either acyclic or energetically unfavorable[5]. Examination of the conformational families obtained by cluster analysis of MMC calculations for the linear (1 → 2)-β-D-glucan heptasaccharide, undecasaccharide, and eicosasaccharide suggested that certain novel conformational repeating structures are adopted at significantly high frequencies, and these repeating structures give rise to low-energy cyclic forms that are distinct from those proposed by Palleschi and Crescenzi.

The use of computer programs utilizing semiempirical force fields in the analysis of the conformational dynamics of macromolecules is rapidly gaining importance. Recent developments of relatively low-cost powerful workstations with massive amounts of fast storage, now allows widespread access to perform molecular dynamics (MD)[6–9] and MMC[10–14] calculations on biopolymers such as oligonucleotides, proteins, glycoproteins, and polysaccharides. However, the motional freedom at each linkage between residues of a biopolymer can result in a large number of significantly populated backbone conformations. MMC or MD calculations must therefore simulate an extended time period in order to adequately span the conformational space available to a biomolecule. MMC calculations of large oligosaccharide or glycopeptide structures often produce data sets consisting of several million conformations[10–14]. Such vast numbers of conformations having multiple degrees of freedom cannot be analyzed manually without making simplifications or assumptions that may not be justifiable. However, cluster analysis is a powerful data reduction method that has the potential to partition large data sets of conformations into a smaller set of conformational families. Thus, the underlying assumption of cluster analysis is that a large collection of objects (conformers) can be subdivided into smaller sets of highly populated distinct families, each of which contain only closely similar objects that can be represented by an average representative object, and that the families are separated by sparsely populated regions.

Cluster analyses require that the molecular conformations to be analyzed be represented as vectors. Many clustering methods rely on a metric such as the

Euclidean distance between two vectors as a measure of the similarity between the objects of the dataset. Often, the first step is to create a distance matrix, which holds the distances between every pair of vectors of the set. For a collection of N objects, the distance matrix will have $N(N-1)/2$ elements, (i.e., on the order of 10^{12} elements for a set of 10^6 conformations). The computer memory requirements and the execution time to perform such a task is prohibitive even for a high-end workstation. The clustering methods employed in this study do not create a distance matrix, and consequently do not suffer from the same space limitations as methods that rely on a distance matrix.

MMC calculations may be used to explore the regions of conformational space that are energetically accessible to biopolymers[10–15]. The calculation is performed by taking a reasonable initial conformation and randomly varying conformational parameters in small increments[15]. The internal energy of the resulting conformation is estimated using a force field, and the new conformation is accepted or rejected on the basis of the Boltzmann probability at a given temperature. Each successive MMC step is performed using the last accepted conformation.

Assuming that the energetically favored conformations of a molecule are restricted to a set of distinct conformational families, a cluster analysis of the results of a MMC calculation will allow the conformational families and their populations to be determined. Furthermore, if the analyzed system is considered to be in thermodynamic equilibrium, the cluster analysis can provide a basis upon which the energy and the entropy of each conformational family may be calculated. Additionally, clustering of the complete data set allows transition rates between families to be calculated. This result can be used to evaluate the kinetic and thermodynamic parameters of the conformational transitions and to assess how faithfully the calculation simulates a condition of thermodynamic equilibrium in the accessed area of the multidimensional conformational space. A central theorem of statistical thermodynamics, the "principle of microscopic reversibility", states that at thermodynamic equilibrium the net flux through a transition between two states is zero. Consequently, a necessary (albeit not sufficient) condition for thermodynamic equilibrium of MMC data sets is that the number of transitions $T(i \rightarrow j)$ be nearly equal to the number of transitions $T(j \rightarrow i)$ for each pair (i, j) of conformational families. An additional requirement for thermodynamic equilibrium of MMC data sets is that all energetically favored conformations have been adequately visited. The use of an elevated temperature parameter increases the likelihood that both of these criteria are fulfilled. However, only by performing mutiple MMC calculations using different initial conformations, or alternatively by performing an extended MMC calculation and perform a "sliding window" analysis of the data set, can one ensure that the parameters obtained from the analysis converge to yield their thermodynamic equilibrium values. In this study all data sets showed near-equality of transitions between conformational families, but only a single MMC calculation was performed for each molecule and consequently thermodynamic equilibrium was not ascertained.

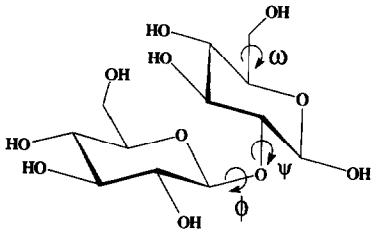
The application of cluster analysis is by no means limited to sets of molecular conformations generated by MMC calculations. Molecular dynamics trajectories calculated for fragments of human parathyroid hormone have been analyzed by "fuzzy" clustering[16], which, in contrast to the methods described here, allows a conformation to have partial membership in more than one conformational family.

METHODS

Metropolis Monte Carlo calculations.—MMC calculations were performed for linear oligomers of (1 → 2)-linked β-D-glucopyranosyl residues using the GEGOP program[4]. Conformational energy calculations performed by GEGOP are based on the assumption that each sugar ring may be considered a rigid entity. The conformation of an oligosaccharide is then uniquely specified by a vector, the elements of which are the torsional angles for the exocyclic chemical bonds, the most significant of which are those between ring carbons and glycosidic oxygens or hydroxymethyl groups (i.e., the angles ϕ = [H-1–C-1–O-1–C-2], ψ = [C-1–O-1–C-2–H-2], and ω = [O-5–C-5–C-6–O-6]). This approach also has the advantage that it embodies a much less complicated calculation of the conformational energy of an individual conformation than do schemes that allow flexibility of the sugar residues themselves, and therefore it allows extensive MMC calculations for large oligosaccharides to be performed. Conformational energies calculated by this method consist of terms involving van der Waals potentials, the exo-anomeric effect for acetals, a bond angle potential for the glycosidic oxygen with a minimum energy set to 117°, and a torsional term for the hydroxymethyl groups' rotation.

The simulated time span of MMC and MD calculations is limited by practical considerations to a very short period (usually in the picosecond to low nanosecond range). Furthermore, transitions between conformational states that are separated by energy barriers as small as 4 kcal/mol may not be observed in a simulation at 300 K because the lifetime of a given conformation (as defined by residence in an energy well) may be similar to or larger than the total simulation time. The pre-exponential factor in the Arrhenius expression for the rate constant for the free rotation of a methyl group has been estimated[17] to be $\sim 10^{13}$ s^{-1}. This may be considered as an upper limit for the pre-exponential factor for conformational transitions. Using a value of 10^{12} s^{-1} for the pre-exponential factor, a transition with an activation energy of 4 kcal/mol at 300 K gives a rate constant of 1.22×10^9 s^{-1}, corresponding to a natural lifetime of 821 ps. In addition, the GEGOP approach of keeping the sugar-ring structures rigid can cause steric effects in crowded regions of the molecule to be overemphasized, leading to the overestimation of transition energies. In order to compensate for these effects, the MMC temperature parameter was set to 1000 K for the disaccharide, heptasaccharide, and undecasaccharide and to 2000 K for the eicosasaccharide. It should be noted that the rigid-residue approximation of GEGOP will not significantly skew any derived macroscopic observables if the force field faithfully models the highly

populated conformations, since contributions from insignificantly populated high-energy conformations are minor. The step length for the maximum variation of the dihedral angles ϕ and ψ was adjusted to 70° for the disaccharide, to 16° for the heptasaccharide, to 12° for the undecasaccharide and to 10° for the eicosasaccharide. The calculations consisted of 10^5 MMC steps (47 423 accepted) for disaccharide **1**, 6.0×10^5 steps (245 853 accepted) for the heptasaccharide, 10^6 steps (373 791 accepted) for the undecasaccharide, and 1.3×10^6 steps (425 412 accepted) for the eicosasaccharide.



1

Modeling of cyclic glucans with the SYBYL software.—The current version of GEGOP does not support calculations of cyclic oligosaccharides. Therefore, no MMC calculations were performed on cyclic molecules, which were modeled using the SYBYL software (version 5.5, Tripos Associates, St. Louis MO). Glucans were generated using β-D-glucose from the standard SYBYL sugar library. Minimum-energy structures for the various cyclic glucan forms were achieved by MD calculations using a temperature ramping scheme. A typical energy minimization using this approach consisted of a series of 1-ps MD simulations between which the temperature parameter was decreased from 100 K to 1 K in steps of ~ 10 K. The step length during the MD simulations was 1 fs, and the standard Tripos force field was used. Electrical charges and hydrogen-bonding were not incorporated in the calculations, and no solvent molecules were included. In order to maintain the symmetry of cyclic molecules, the initial atomic velocities for the first interval were set to zero, and the final atomic velocities for each interval were used as the initial parameters for the next interval.

Three-distance method clustering (3-DM).—Our implementation of the three-distance clustering method of Zupan[18,19] (3-DM) is the program MCLUST, written in the ANSI "C" programming language to run under the ULTRIX (UNIX) operating system on DECstations from Digital Equipment Corporation.

The 3-DM algorithm is a hierarchical clustering method[18,19], and as such it relies on the creation of a hierarchical tree of interconnected vertices (Fig. 1). The origin of the tree is called the root vertex and each vertex of the tree has two subordinate vertices. Each vertex stores the average vector of its subtree and keeps a count of all its subordinate members. Finer and finer distinctions are made at each vertex as the hierarchical tree is traversed from the root towards the leaves.

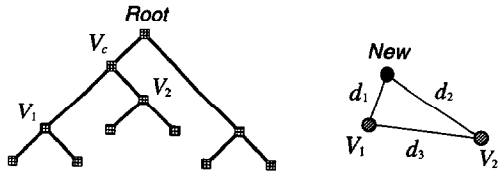


Fig. 1. The root vertex of the hierarchical tree (depicted on the left) represents all of the vectors in the set. Each vertex V_c has two subordinate vertices, V_1 and V_2, which may in turn represent a subset of "similar" vectors (i.e., a cluster). In order to position a new vector in the tree, the 3-DM algorithm starts at the root vertex and traverses the tree via a series of connected vertices. At each vertex (V_c) the basis for positioning the new vector is the 3-DM branch criterion, which depends on the relative magnitudes of the three distances d_1, d_2, and d_3 (depicted on the right).

The leaves may also be considered vertices, each representing only one object, namely one of the actual vectors being clustered. Each vertex is a potential candidate for a cluster, the members of which are all its subordinate leaves. However, a vertex adequately represents a homogenous cluster only if all of its subordinate leaves are "similar" according to some criterion.

A hierarchical tree for a dataset of N vectors consists of $2N-1$ vertices. In addition, most hierarchical clustering methods[20,21] utilize a distance matrix with at least $N(N-1)/2$ elements, and are thus impractical for large molecules due to limitations in the amount of data storage available to most computers. In contrast, the 3-DM does not require a distance matrix and, therefore, can be used to analyze MMC data sets for large oligomers.

The calculation of average angles and distances in angle space is complicated by the circular equivalence of angles that differ by any multiple of 360°. Thus, the true average value of an angle is not in general equal to its arithmetic mean value, and the difference between any two angles cannot be calculated by simply subtracting two numerical values. This problem was circumvented by expanding each interglycosidic angle into its component cosine and sine values. Thus, the $2(N-1)$ interglycosidic torsional angles in an oligosaccharide having N residues is represented by a conformational vector having $4(N-1)$ components. The Euclidean square distance was used throughout as the distance measure. Given vectors X and Y, the Euclidean square distance $(d_{xy})^2$ is defined by the equation:

$$(d_{xy})^2 = \sum_i (X_i - Y_i)^2 \qquad (1)$$

The *a priori* specification of a single objective clustering (similarity) criterion is not possible[21], as similarity depends on the choice of coordinate system (dihedral angles, cartesian coordinates, rms deviation of backbone atoms, etc.), and upon which aspect of a data set is at focus in the study (backbone conformation, proximity of selected groups, conformation of side chains, etc.). In this study, a vertex was chosen to represent a cluster only if the rms distance from its mean vector to its subordinate members was less than a specified threshold. Thus, clustering was performed by traversing the tree from the root towards the leaves

until a set of vertices that all fulfill the rms distance criterion were found. For a given rms value, the quality of the clusters produced by MCLUST was assessed through the calculation of statistical descriptors for all the dihedral angles of each cluster. These descriptors included the mean value, standard deviation, skew, and number of outliers at one, two, and three times the standard deviation for each dihedral angle of a cluster. The optimal value of the rms threshold was determined empirically for every MMC dataset by cutting the cluster tree at various levels of the rms value until a set of clusters was found that showed satisfactory statistics for all dihedral angles. A shortcoming of the three-distance method is that closely related clusters occasionally occur in different areas of the tree. To tackle this problem, MCLUST also calculates the fairly small intercluster distance matrix, which may be used for manually pooling closely similar clusters.

Optimal acceptance ratios for an MMC simulation are 30–50%, resulting in data sets containing short sequences of identical conformations. In order to reduce memory requirements, we have therefore modified the 3-DM algorithm to allow each leaf to represent several identical conformational vectors. The hierarchical tree created by the 3-DM depends[18,19] on the presentation sequence of the vectors. We have therefore implemented an optional random presentation sequence, which was used throughout this work to minimize the effects of sequence dependencies.

Initially, MCLUST reads in all the vectors to be clustered. Two vectors are chosen at random to create an "embryonic" tree having only two leaves. The remaining vectors are added to the growing tree in a random order. Each new vector is positioned by traversing the tree from the root towards the leaves and calculating *three distances* at each vertex: the distances d_1 and d_2 between the new vector and the two subordinate vertices of the current vertex, and the distance d_3 between the two subordinate vertices themselves (Fig. 1). If the distance d_3 between the two subordinate vertices is shorter than either of the distances d_1 and d_2, a vertex carrying the new vector as a leaf is inserted at a position immediately preceding the current vertex. Otherwise, the tree is traversed further in the direction of that subordinate vertex which represents the shorter of the distances d_1 and d_2 until the 3-DM branch criterion is met, or alternatively until a terminating leaf is reached, in which case a vertex carrying the new vector as a leaf is inserted in a position immediately preceding the terminating leaf. Since the 3-DM branch criterion is local in nature, the initial tree will be dependent on the presentation sequence, and it will typically have many misplaced vectors. Therefore, an iterative relocation procedure is adopted which, using a random sequence, traverses the tree to find the closest matching vector for each of the vectors comprising the tree. If a vector does not correctly find "itself" as the closest match, it is removed from the tree and subsequently positioned anew using the 3-DM algorithm. This iterative relocation process is continued until all vectors consistently find "themselves" as their closest match, or until a maximum number of iterations is reached. The use of a random sequence does not only ensure that the organization of the resulting hierarchical tree will be less influenced by the

presentation order, but it will also counter tendencies towards infinite loops of cyclic vector relocations.

Occasionally the distance between the average vector of a vertex and one of its subordinate members may become quite large. In such cases it is entirely possible that d_3 is larger than some individual distance between the new vector and a subordinate member of the current vertex, even though d_3 is shorter than either d_1 or d_2. In such instances it would be more appropriate to join the new vector to one of the two subordinate vertices than to create a new leaf. To address this issue a so called safety wall[18,19], which effectively increases the value of d_3, is employed. The use of a safety wall improves the quality of the clusters while also serving to reduce the depth of the tree[18,19]. The safety wall employed throughout this study is defined in the following equation:

$$d_3' = d_3(1 + N_v/N_w) \quad (2)$$

where d_3' is used in place of d_3 for the 3-DM criterion. N_v is the number of subordinate vectors for the current vertex, and N_w is an empirically determined safety wall parameter which may be interpreted as the number of subordinate vectors for which d_3' becomes twice the value of d_3. The effect of the safety wall is to make the criterion for adding new vectors to a vertex increasingly permissive as the vertex membership count increases. Values of N_w in the range 70–90 produced acceptable results for all the data sets of this study. It should be mentioned that this type of safety wall has a tendency to favor hyperspherical clusters[18,19].

Population-density projection clustering (PDP).—As previously mentioned, "natural clusters" are characterized by the presence of sparsely populated intervening regions. The process of obtaining such natural clusters was simplified by the observation that the dependence of population density on the glycosidic dihedral angles ϕ and ψ appeared to be dominated by local interactions (i.e., between adjacent residues) in the (1 → 2)-β-D-glucan oligomers, as the overall appearance of population density plots showed little dependence on the size of the oligomer. Distal interactions (between nonadjacent residues) further restricted the energetically available conformational space, in effect adding to the conformational restrictions imposed by local interactions. Thus, the locally defined energy barriers provide an easily accessible means to begin the process of separating conformational families since they exist whether or not distal interactions are operative; therefore, these barriers should be observable in two-dimensional projections of the population-density surface. Minima in these population projections can be used to define sparsely populated boundary regions (energy barriers) between conformational families. Pairs of distinct conformational subpopulations that are separated by a minimum in at least one such projection are distinguishable by a simple analysis of a set of these two-dimensional projections. Therefore, assuming the predominance of local interactions in determining energy barriers allows an N-dimensional clustering problem to be reduced to a set of $N/2$ two-dimensional clustering problems, followed by a classification procedure that combines the two-dimensional cluster identities yielded by each projection.

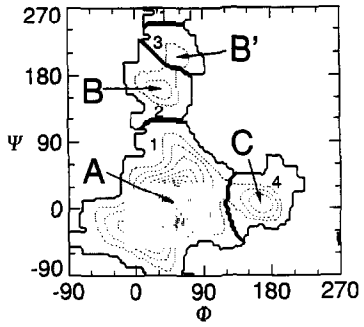


Fig. 2. Population-density surface of β-sophorose (i.e., 2-*O*-β-D-glucopyranosyl-β-D-glucopyranose, (**1**)] obtained by MMC calculations. Contours (dotted lines) correspond to population densities of 1, 4, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 96% of the maximum value. Four indexed regions (solid lines, numerals) of the ϕ, ψ space were obtained by partitioning with the PDP program. For oligosaccharides with more than two residues, the ϕ, ψ projection of the population-density surface corresponding to each glycosidic bond is partitioned in a similar fashion, providing a basis for generation of an "index vector" representation for each conformation (see text.)

Prototype software (PDPCLUST) using the PV-WAVE™ macro language was developed to carry out segmentation analysis for sets of two-dimensional population-density projections and to perform the subsequent classification of MMC data sets.

The procedure for population-density projection analysis was as follows. Population densities (as a function of the dihedral angles ϕ and ψ) were determined by simple counting, using a bin size of $3° \times 3°$ to yield a 120×120 element matrix for each glycosidic linkage. Each matrix was smoothed by applying a moving average "window" (of width 3 to 5 points) to each row and column. Edge effects at the 0° and 360° boundaries were eliminated by wrapping, using modulo 360 arithmetic in the calculations. The matrices were partitioned into regions of significant populations separated by regions of low population, and each significantly populated region was assigned a region index (Fig. 2). A (region) index vector representation (of dimension $N/2$) was then created for each conformation in the MMC data set by mapping each glycosidic linkage in the conformation to an indexed region defined during the matrix partitioning procedure. Conformations which gave rise to identical index vectors were considered "similar" and were therefore assigned membership to the same region of conformational space (a cluster). Thus, each glycosidic linkage was partitioned into only a few significantly populated conformational regions, and consequently the number of unique index vectors (potential clusters) was much smaller than the number of unique conformations. Furthermore, many of the possible index-vectors corresponded to regions of conformational space that were not populated, so the total number of clusters found by this method was very small compared to the number of conformations.

The initial partitioning of the smoothed population-density matrices was accomplished as follows: rows, columns, and diagonals of the population-density matrices

were convoluted with 5- or 7-element vectors of the form [−3, −2, −1, 0, 1, 2, 3], effectively calculating the average of the first derivative across several points, and minima were identified as points where this average derivative ascended through zero. All elements of the matrix having populations less than a specified threshold value were also considered to be minima. The locations of all minima were recorded, and adjacent minima were connected, thus defining for each (ϕ, ψ) pair a set of bounded discrete regions, each of which was assigned a unique region index (Fig. 2).

RESULTS AND DISCUSSION

MMC calculations were performed for linear fragments of a (1 → 2)-β-D-glucan containing 2, 7, 11, and 20 sugar residues. The MMC calculations for the disaccharide were intended merely to explore the conformational freedom of the (1 → 2)-β-D-glycosidic bond in the absence of any effects that may arise from the interaction of nonadjacent sugars. Three major conformational energy minima for this glycosidic bond can be directly visualized in the population-density map for this calculation (Fig. 2), having centroids at (ϕ, ψ) values of approximately (55°, 10°), (40°, 160°) and (170°, 10°).

Clustering of MMC data sets for the heptasaccharide using the 3-DM algorithm.— The consequences of contacts between nonadjacent residues were initially investigated by performing extended MMC calculations for a linear (1 → 2)-β-D-glucan heptasaccharide. The conformations of the glycosidic linkages were very similar to those calculated for the disaccharide, having centroids at (ϕ, ψ) values of approximately (50°, 10°) for conformation A, (50°, 160°) for conformation B, and (160°, 10°) for conformation C.

Cluster analysis of the MMC data set from the (1 → 2)-β-D-glucan heptasaccharide was perfomed using MCLUST (described in the experimental section). A cluster selection criterion using an rms distance threshold of 1.8 was employed, resulting in a partitioning of the data set into 24 clusters. The standard deviations of the dihedral angles for all clusters had values in the range 15–30° with distributions that were close to a normal distribution and without significant skew, indicating that the clusters found were symmetrical. These 24 clusters were further combined (assisted by the intercluster distance matrix) into 11 conformational families that could all be expressed in terms of their distribution of A, B, and C conformations along the heptasaccharide backbone (summarized in Table I) and that collectively accounted for over 99% of the conformations of the MMC data set.

The most prevalent glycosidic linkage conformation of the heptasaccharide was the A conformation, but a significant percentage of the bonds also adopted the B conformation. The temperature used in the MMC calculation was 1000 K, and yet very few bonds adopted the C conformation, which tended to occur only at the ends of the molecule. Therefore, only A and B conformations are likely to be

TABLE I

Major conformers of (1 → 2)-β-D-glucan heptasaccharide in A, B, C space [a], as distinguished by 3-DM analysis

Sequence [b]	Population (%)
AAAAAA	69.0
AAAAAB	10.9
CAAAAA	6.8
ABAAAA	6.1
AAABAA	2.5
AABAAA	1.3
AAAAAC	1.0
CAAAAB	0.7
AACAAA	0.7
ABAAAB	0.2
BCAAAA	0.2
Total	99.4

[a] Glycosidic bond conformations are defined for (ϕ, ψ) centered at; (50, 10) = A; (50, 160) = B; (160, 10) = C. [b] Sequence of glycosidic bond conformations, starting from the nonreducing end.

significantly represented at 300 K in a cyclic (1 → 2)-β-D-glucan having no end residues to accommodate the C conformation. These results suggest that cyclic (1 → 2)-β-D-glucan molecules may be composed of conformational repeat structures containing both A and B conformations.

The populations of the 11 clusters of the heptasaccharide (Table I) were analyzed in order to quantify the influence that the conformation of one glycosidic linkage might impose on the conformation of other glycosidic linkages in the (1 → 2)-β-D-glucan chain. In this context, conditional frequencies of two simultaneous non-A conformations occurring in the same heptasaccharide were calculated. The frequency $F(\alpha \mid \beta)$ of event a *conditional* on the simultaneous occurrence of event β may be expressed as the frequency $F(\alpha \cap \beta)$ of simultaneous α and β events divided by the total frequency $F(\beta)$ of event β. That is, $F(\alpha \mid \beta) = F(\alpha \cap \beta)/F(\beta)$. $F(\alpha \mid \beta)$ may be compared to the total frequency $F(\alpha)$ of event α. If $F(\alpha \mid \beta) \neq F(\alpha)$, then the frequency of event α depends on the simultaneous occurrence of event β, which either increases or decreases the frequency of event α.

Conditional frequencies of glycosidic linkage conformations for the MMC data set of the heptasaccharide are depicted in Fig. 3. F(B2 | B6) represents the frequency of a B conformation in position 2 conditional on the presence of a B conformation at position 6. That is, F(B2 | B6) = F(xBxxxx | xxxxxB) = F(xBxxxx ∩ xxxxxB)/F(xxxxxB), where the sequence of glycosidic bonds is written starting from the nonreducing terminus and "x" indicates that the glycosidic bond can be in any conformation (A, B, or C).

Conditional frequencies were used to examine whether conformational repeat structures consisting of A and B conformations were likely to form. The occurrence of the ABAAAB conformation with a small but significant population of

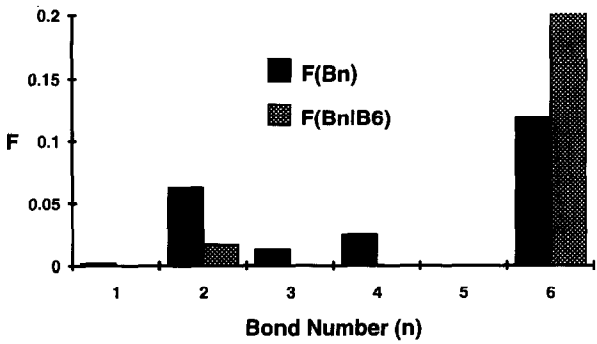


Fig. 3. Frequencies of glycosidic bond conformation B as a function of location in the (1 → 2)-β-glucan heptasaccharide. F(Bn) corresponds to the global frequency of a B conformation at bond n. F(Bn|B6) corresponds to the frequency of a B conformation at bond n, conditional upon the presence of a B conformation at bond 6. F(B6|B6) = 1, by definition, and thus is off-scale.

0.2% indicated that a tetrasaccharide fragment having the [AAAB] conformation was a potential candidate for such a repeat structure. The conditional frequency F(B2|B6) for the ABAAAB conformation was found to be 0.018 (Fig. 3), which is *less* than the value 0.063 for F(B2), indicating that the occurrence of a B conformation at position 6 disfavors a simultaneous occurrence of a B conformation at position 2 by a factor of 3.5. However, the occurrence of a B conformation at position 6 appears to totally preclude another B conformation at positions 3, 4, and 5. Thus, while not highly favored, the [AAAB] conformational repeat structure is accessible to the heptasaccharide. Furthermore, the tendency of (1 → 2)-β-D-glucans to adopt the [AAAB] conformational repeat structure increases with growing chain length (see below).

A cyclic (1 → 2)-β-D-glucan eicosamer consisting of five AAAB repeat structures could be readily constructed with the SYBYL™ molecular modeling software by adjusting the (ϕ, ψ) values for the A and B conformations to (58°, 10°) and (58°, 161°), respectively. This brought O-1 of residue 20 within a distance of 1.46 Å from the C-2 of residue 1. Ligation of the molecule after removal of OH-2 of residue 1 and H-1 of residue 20 resulted in a cyclic molecule having C_5 symmetry. Using the SYBYL force field to perform an energy minimization, the cyclic molecule adopted the conformation, depicted in Fig. 4A, having a calculated energy of 69.5 kcal/mol. The dihedral angles of this conformation are presented in Table II. As a comparison, the minimized energy of the linear eicosamer consisting entirely of A-type glycosidic bond conformations was 38.4 kcal/mol. However, these values clearly do not represent the global minima. Thus, molecular dynamics calculations in which the temperature parameter was gradually decreased to 1 K yielded conformational energies of 4.3 kcal/mol and 20.0 kcal/mol for the cyclic $[AAAB]_5$ and linear $[A]_{20}$ eicosamers, respectively, while retaining the symmetry of the molecules. It should be cautioned that the energies for the linear and cyclic eicosamers may only be used for a rough comparison since the

two molecules have different numbers of atoms and bonds due to the lack of a ring closure in the linear molecule. Thus, the [AAAB] repeat structure represents an energetically favorable motif for cyclic (1 → 2)-β-D-glucan oligosaccharides due to the presence of cooperative effects originating from van der Waals interactions between nonadjacent residues. Conversely, the all-A conformation represents the lowest energy form for linear oligosaccharides with fewer than six sugar residues.

3-DM analysis of the MMC data set for the undecasaccharide.—MMC calculations of a β-(1 → 2)-D-glucan undecamer were carried out in order to test the conclusions stated in the preceding paragraph. The large size of the MMC data set made it necessary to modify our approach to the clustering analysis. 3-DM clustering was performed on a 155 000-vector subset of the MMC data. To avoid

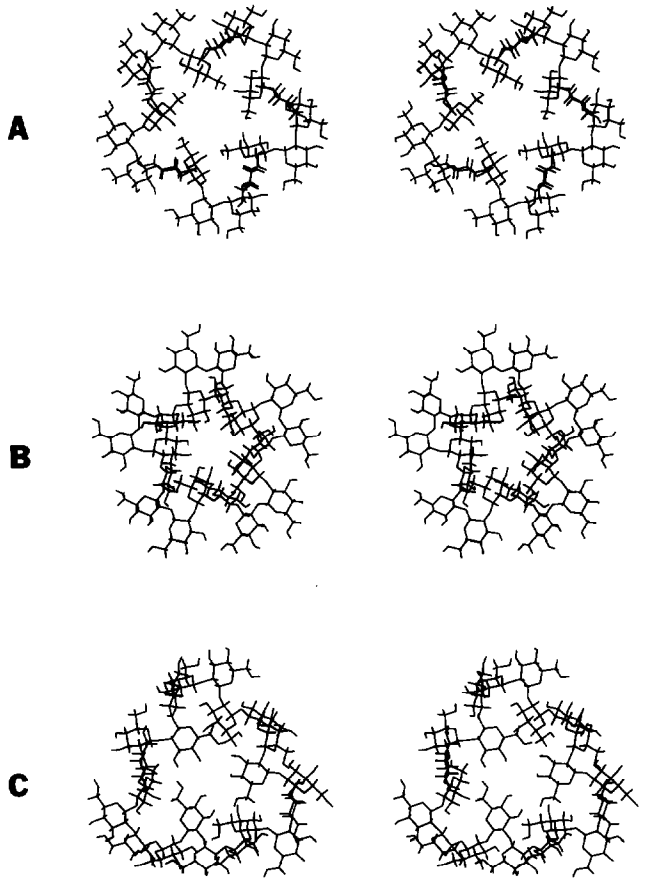


Fig. 4. Stereoviews of cyclic (1 → 2)-β-D-glucan models. (A) The $[AAAB]_5$ eicosamer with C_5 symmetry (E = 69.5 kcal/mol after minimization). (B) The $[AAAB]_5$ eicosamer after annealing (4.3 kcal/mol). (C) The $[AAAAAAB]_3$ heneicosamer with C_3 symmetry (E = 40.1 kcal/mol, after minimization). (D) The $[AAAAAAB]_3$ heheicosamer after annealing (−1.6 kcal/mol. (E) The $[AAA]_7$ eicosamer based on the model of Palleschi and Crescenzi, after minimization (45.0 kcal/mol). (F) The annealed $[AAA]_7$ eicosamer (31.2 kcal/mol).

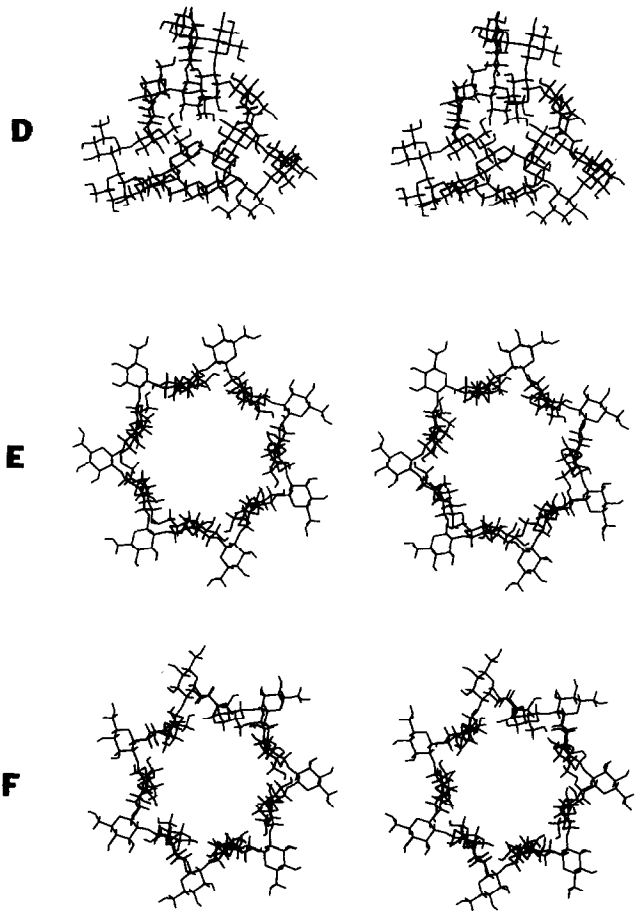


Fig. 4. (continued)

starting effects from the arbitrarily chosen start conformation the first 30 000 conformations were discarded, and every other vector starting with the 30 001st and up to the 340 000th were included in the cluster analysis. A cluster selection criteria using an rms threshold value of 3.0 resulted in 12 clusters, while an rms threshold of 2.5 produced 57 clusters. The glycosidic bonds for each of the 12 clusters obtained using the rms threshold value of 3.0 were assigned as having conformations A, B, or C as described above for the heptasaccharide, and the clusters were combined and classified on this basis (Table III). A similar set of conformational families (Table III) was obtained by combination of the 57 clusters obtained using a cutoff value of 2.5. However, examination of the 57 clusters revealed that the B conformation was separated into two distinct conformations, designated B and B′, for which the ψ angle values of the population centroids were approximately +160 and −160 degrees, respectively.

The frequency of occurrence for the glycosidic conformations of the undecasaccharide were similar to that found for the heptasaccharide [i.e., F(A) ≫ F(B) ≫

TABLE II

Torsional repeat sequence of the cyclic (1 → 2)-β-D-glucans

Bond [c]	Conformation [a]			
	Initial	Minimized [b]	Annealed [b]	
Eicosamer with C_5 symmetry: $(BAAA)_5$				
		(69.5 kcal/mol)	(4.3 kcal/mol)	
i	58, 161	61.2, 170.4	52.2, 170.8	B
i+1	58, 10	57.5, 40.4	57.1, 39.5	A
i+2	58, 10	57.3, 21.0	63.5, 55.3	A
i+3	58, 10	42.7, 25.2	63.3, −30.7	A
Heneicosamer with C_3 symmetry: $(BAAAAAA)_3$				
		(40.1 kcal/mol)	(−1.6 kcal/mol)	
i	55, −160	69.0, −162.5	61.0, −177.3	B
i+1	55, 10	49.6, 24.0	66.0, 45.8	A
i+2	55, −20	58.0, −26.9	71.0, −10.2	A
i+3	70, 12	60.2, 27.0	66.6, −15.4	A
i+4	50, 6	48.2, 13.5	28.4, 31.8	A
i+5	55, 10	58.9, 30.3	72.9, 68.4	A
i+6	55, 10	56.5, 25.0	68.8, 58.2	A
Palleschi and Crescenzi heneicosamer model with C_7 symmetry: $(AAA)_7$				
		(45.0 kcal/mol)	(31.3 kcal/mol)	
i	75, 32	67.8, 50.2	68.7, 58.7	A
i+1	54, 22	57.4, 46.2	59.5, 51.7	A
i+2	22, 5	26.5, 23.5	8.8, 42.5	A

[a] Expressed as the dihedral angles (ϕ, ψ) for the glycosidic bond. [b] See text for details. The conformational energy is in parentheses. [c] The glycosidic bond index increases toward the potential reducing end.

F(C) for linkages not involving terminal residues], and the C conformation occurred most frequently at the ends of the oligosaccharide.

PDP analysis of MMC data sets for the undecasaccharide.—The simplicity of the PDP method allowed it to be applied to the entire 373 791-vector data set obtained by the MMC calculation for the (1 → 2)-β-D-glucan undecasaccharide. The results of the PDP analysis of the undecamer MMC data set is included in Table III.

A series of two-dimensional (ϕ, ψ) population-density maps for the full MMC data set for the undecasaccharide is shown in Fig. 5A. Significant population of the C conformation is apparent only near the ends of the molecule, and distinct populations within the region corresponding to conformation B can be seen. PDP analysis of this data resulted in 52 conformational families, which were distinguished in part on the basis of the slightly different B and B′ conformations. The population-density maps of the dominant "all-A" family and two conformational families containing the [AAAB] repeat structure are presented in Figs. 5B–5D. The criterion for separation of the latter two families was the detection of minima in the global (ϕ, ψ) population-density map for the 5th glycosidic bond, allowing the B and B′ conformations to be distinguished at this position. This particular distinction was not readily obtained using the 3-DM analysis unless a small rms

TABLE III

Populations (in percent) for the major conformers of the (1 → 2)-β-D-glucan undecasaccharide in A, B, C space [a], as distinguished by 3-DM and PDP analyses

Sequence [b]	3-DM		PDP
	(2.5) [c]	(3.0) [c]	
AAAAAAAAAA	66.10	64.70	69.81
AAAABAAAAA	8.45	8.45	6.98
CAAAAAAAAA	6.61	6.61	5.05
AAAABAAABA	5.68	5.68	4.58
AAAAAAAABA	5.18	4.36	4.35
AAAAAAABAA	2.59	2.59	2.15
CABAAAAAAA	2.21	2.21	1.56
AAABAAAAAA	2.14	2.38	2.27
AAAAAAAACA	1.03	0.69	1.06
ABAAAAAAAA	0.00	0.68	0.80
AAAAAAAAAB	0.00	0.00	0.66
AABAAAAAAA	0.00	0.00	0.46
CAAAAAAAAB	0.00	0.00	0.19
AAAAAAAABC	0.00	0.83	0.06
AATAAAAAAA	0.00	0.21 [d]	0.00
ABTAAAAAAA	0.00	0.28 [d]	0.00

[a,b] See Table I. In order to compare the clustering methods, conformational families were combined such that conformation B was not distinguished from conformation B′, and conformation A was not distinguished from conformation A′. [c] The distance cutoffs of 2.5 and 3.0 for 3-DM analysis resulted in 12 and 57 clusters, respectively. PDP analysis produced 52 clusters. [d] The conformation at bond 3 for this cluster, as generated by 3-DM analysis, was centered at the transition point between conformations A and C, and hence labelled T.

cutoff value was employed. However, an undesired side effect of using a small rms cutoff value was the partitioning of otherwise homogenous conformational families into multiple closely related families. For example, an rms cutoff value of 2.5 produced a set of 57 3-DM clusters for this data, 26 of which (64.7% of the total population) were comprised totally of conformation A. Conversely, PDP analysis of the data clearly resolved the two closely related families depicted in Fig. 5C and 5D, yet yielded only five "all-A" clusters representing 69.8% of the total population. Two of the "all-A" clusters were not significantly populated (< 0.1% of total). Two others, comprising 9.9% the total population, were distinguished from the most populated "all-A" cluster (59.9% of the total) by the presence of an A′ bond conformation. The A′ conformation corresponds to the region centered at (ϕ, ψ) coordinates (−10°, −30°), which is separated from the bulk of the A region by a shallow minimum in the population-density projections of the global data for bonds 1 and 2.

In order to compare the populations obtained by PDP with those obtained by 3-DM, B′ conformations were combined with the B conformations and A′ conformations were combined with the A conformations. Comparison of the results (Table III) for the 3-DM and PDP methods found that they were in good agreement.

Conditional frequencies for the simultaneous occurrence of two non-A conformations in the undecasaccharide were calculated (Fig. 6), revealing that when two B conformations occurred simultaneously, they were most often separated by three

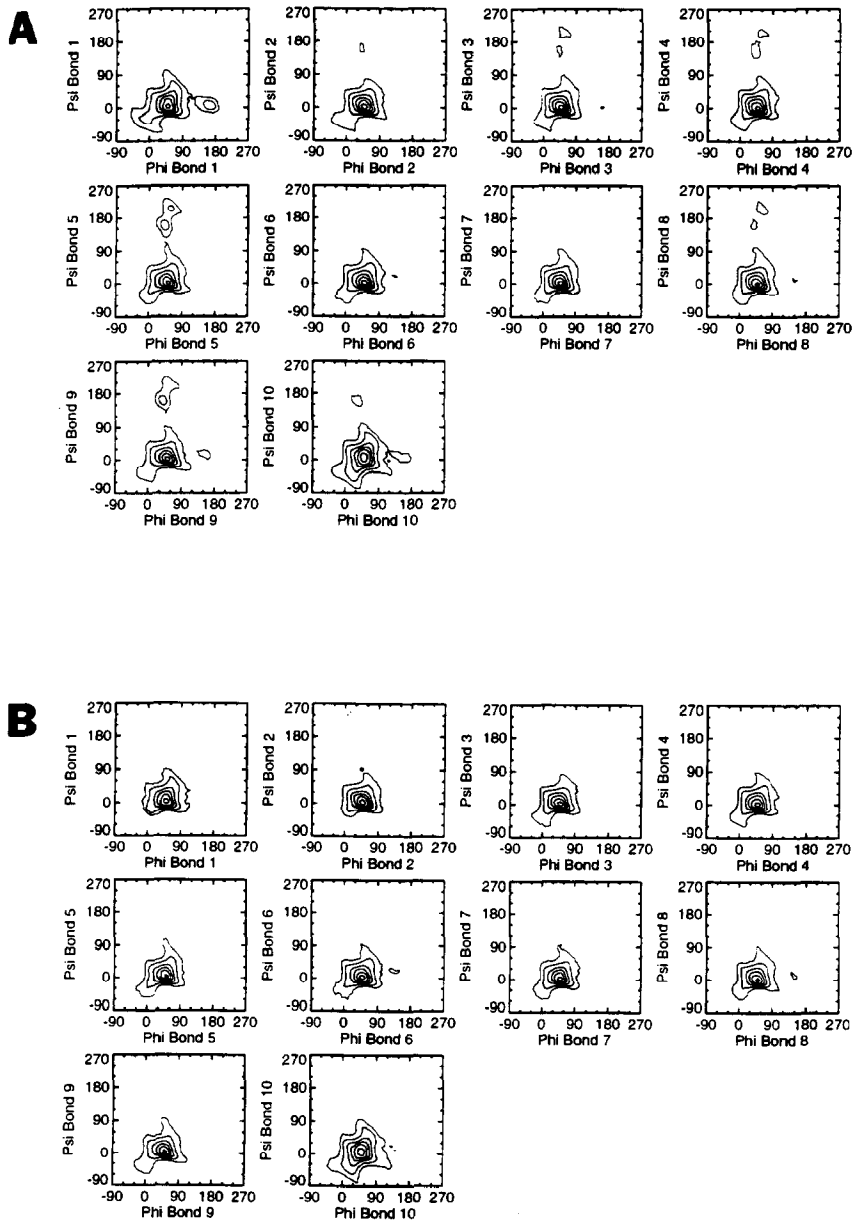


Fig. 5. Population-density surfaces for the (1 → 2)-β-D-glucan undecasaccharide MMC calculation. (A) The complete set of 1 000 000 MMC steps. (B) The dominant "all-A" conformational family, comprising 59.9% of the total data. (C) The sequence [AAABAAABA], comprising 1.71% of the total. (D) The sequence [AAAB′AAABA], comprising 1.03% of the total.

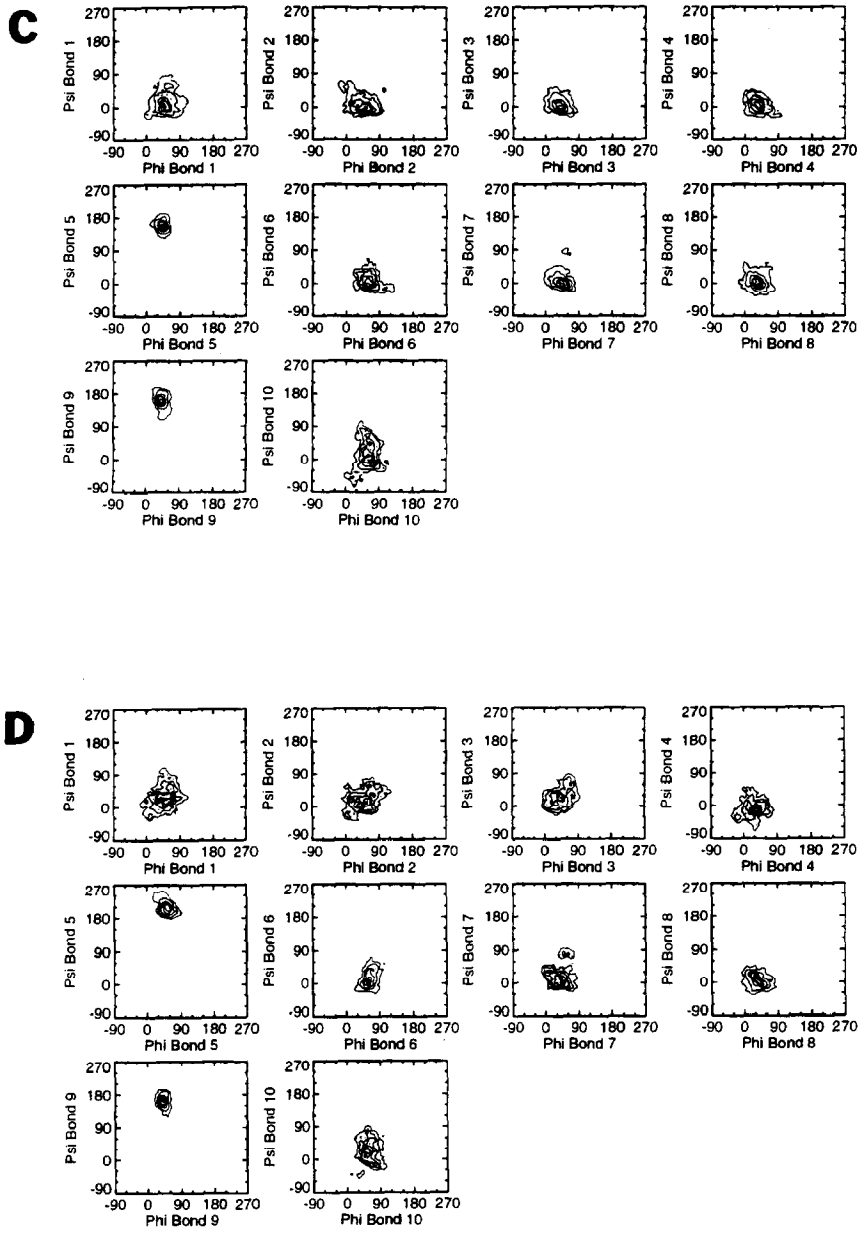


Fig. 5. (continued)

intervening A conformations, as was the case for the heptasaccharide. However, the conditional frequencies corresponding to an [AAAB] repeating unit were significantly different for the undecasaccharide (Fig. 6) than for the heptasaccharide (Fig. 3). For the undecasaccharide, F(B5|B9) and F(B5) were 0.52 and 0.143, respectively, indicating that the presence of a B conformation at position 9 made it

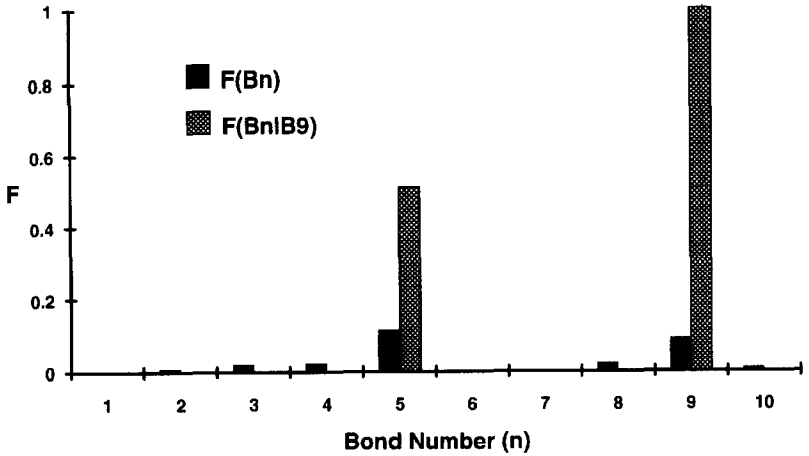


Fig. 6. Frequencies of glycosidic bond conformation B as a function of location in the (1 → 2)-β-D-glucan undecasaccharide. See Fig. 3 for nomenclature.

about 3.5 times *more* probable that a B conformation would also occur at position 5 than would otherwise be the case. This result is in stark contrast to the heptasaccharide, where a similar arrangement of two simultanously ocurring B conformations was about 3.5 times *less* probable than that of a single B conformation. These results suggest that favorable van der Waals contact between nonadjacent residues stabilizes the [AAAB] conformational repeat structure, and that this effect seems to increase with the length of the (1 → 2)-β-D-glucan chain. However, a different conformational repeat structure containing both A and B conformations is significantly represented in (1 → 2)-β-D-glucans with 20 residues (see below).

In addition to the population analysis, the PDPCLUST software also generates a transition matrix T containing the number of transitions in each direction for every pair of conformational families. Combination of the population count (N_i) for family i and the number of interfamily transitions (i.e., T_{ij} of the transition matrix T) provides a pseudo-rate constant (k_{ij}) for the transition i → j, having the form $k_{ij} = T_{ij}/N_i$. The rate constant k_{ij} corresponds to the probability that an individual MMC step initiated in conformation i will result in a transition to the j conformation. Inappropriate subclassification of conformations by PDP may be detected by examining these transition rate constants. Thus, pairs of PDP clusters that exhibit very high rates of interconversion due to the lack of a sparsely populated intervening region of conformational space (i.e., an energy barrier) may be more properly classified as a single cluster. For example, the transition rates between the three significantly populated "all-A" PDP families were very high, indicating that they should be combined.

The transition matrix defined above may be used to asses asymmetry of the

transition flux between pairs of conformational families, as an indication of how closely the MMC calculation adheres to the "principle of microscopic reversibility" (see above). Transition counts between zero and 985 were obtained for the undecamer. Transition asymmetries were zero for 38 of the 77 interconverting pairs of families, with 37 pairs having single-transition asymmetries, one pair having a two-transition asymmetry, and one pair having a three-transition asymmetry, suggesting that the MMC calculation did indeed approach microscopic reversibility. However, thermodynamic equilibrium may not have been achieved since there is no guarantee that the entire energetically accessible space had been visited. The [AAAB] conformational motif is possible at several locations in the undecasaccharide. In fact, the minimized conformational energies of the [AABAAABAAA] and [AAABAAABAA] conformations were similar to that of the [AAAABAAABA] conformation. Of these three, only the [AAAABAAABA] conformation was visited during the calculation, suggesting that a limited portion of the available conformational space was explored. Of course, the near equality of the energy minima for the three conformations does not necessarily indicate that an infinitely long MMC calculation would produce equal populations for their conformational families, since the width of the energy well plays an important role (entropy) in determining the frequency with which a conformation is visited during a MMC calculation.

The PDP procedure can occasionally fail to distinguish pairs of families whose demarcation does not correspond to a minimum in the total population distribution *in any projection*. Regions of low population that were apparently due to interactions of distal glycosyl residues were frequently observed in two-dimensional population-density projections of PDP clusters, indicating that these clusters contained two or more conformational families. These regions of low population were obscured in the population-density projections of the global data by the vast number of conformations that did not embody the distal contact. An example is shown for the cluster in Fig. 5D, where the population-density map for bond 7 exhibits two local maxima. The conformational family was readily sub-clustered by a second PDP iteration, yielding two new "natural" clusters.

An iterative procedure, in which all pairs of PDP clusters exhibiting high interconversion rates are combined (see above) and the resulting set of clusters is sub-clustered by PDP analysis, should resolve all natural clusters which are observable in the population-density projections. Nevertheless, there may be cases where an iterative PDP clustering will fail to separate all of the conformational families because intervening regions of low population density are not observable in any of the projections chosen. These families would be distinguishable only in projections obtained via orthogonal transformations (rotation) of the coordinate system or by application of a different (sub)clustering method to the PDP results.

PDP analysis of MMC data sets for the eicosasaccharide.—The capability of PDP analysis to classify extremely large data sets made it feasible to analyze the results of MMC calculations for a linear eicosamer, which is large enough to form a

covalent circle. In order to examine as much of the conformational space as possible, 1.3×10^6 MMC steps were executed using a temperature parameter of 2000 K to increase transition probabilities. PDP analysis of the MMC data resulted in the generation of 666 clusters, which were combined by the A, B, C conformational criteria for simplified presentation in Table IV. Some of the data is summarized in Fig. 7, which compares the global frequencies of B conformations (solid bars) to selected conditional frequencies of B conformations (crosshatched bars). This analysis indicates that the presence of a B conformation strongly favors the simultaneous occurrence of another B conformation several bonds away. Furthermore, the B conformation strongly disfavors the occurrence of another B conformation in a few of the adjacent glycosidic bonds. For example, a B conformation at bond 3 (Fig. 7A) completely abolished B conformation in positions 2, 4, 5, 9, 10, 11, 15, 16, 17, and 19 while increasing the B conformation frequency at positions 7, 8, 12, and 14. However, contiguous repeats of the [AAAB] conformational repeating structure was not evident in this data, as this would correspond to increased B conformation frequencies at bonds 7 and 11 when a B conformation was present at bond 3.

An alternative conformational repeat [AAAAAAB] that leads to a low-energy cyclic structure is apparent in the data presented in Fig. 7B. The B conformation at bond 9 reduces the B conformation frequency at bonds 3, 4, 6, 7, 8, 10, 12, 13, 15, and 19 while significantly increasing it at bonds 2 and 16. The resulting [ABAAAAAABAAAAAABAAA] conformational family is the most highly populated family that contains more than one non-A bond conformation (Table IV). This family was composed of eight subfamilies (which were distinguished by the presence of either B or B′ conformations) which comprised 1.14% of the total population.

A circular (1 → 2)-β-D-glucan heneicosamer consisting of 3 [AAAAAAB] conformational repeats was constructed with the SYBYL software by adjusting the ϕ and ψ angles to the values listed in Table II. This brought O-1 of residue 21 to within 1.25 Å of C-2 of residue 1 and resulted in a conformation with C_3 symmetry. Atoms were deleted and bonds were formed to make a covalently closed circle and an energy minimization was performed (Fig. 4C), bringing the energy to 40.1 kcal/mol. Further annealing of the molecule (Fig. 4D) was accomplished by performing a series of molecular dynamics calculations with the temperature approaching 0 K, as described above for the $[AAAB]_5$ eicosamer with C_5 symmetry. This resulted in a conformational energy of −1.6 kcal/mol for the $[AAAAAAB]_3$ heneicosamer retaining C_3 symmetry.

It is interesting to note that, because consecutive stretches of conformation A form helices that repeat every three residues in the (1 → 2)-β-D-glucans, both the [AAAB] and [AAAAAAB] conformational repeat structures embody an integral number of helix turns. Thus, molecular models based on these conformational sequences consist of short helical domains that are covalently attached to each other at the corners by "kinks" of glycosidic bonds in the B conformation. The

helical domains have favorable van der Waals contacts with each other leading to very stable structures.

The conformational energy of the $[AAAAAAB]_3$ heneicosamer can be directly compared with that of an heneicosamer with C_7 symmetry based on the trisaccharide conformational repeat suggested by Palleschi and Crescenzi. A cyclic $[AAA]_7$ $(1 \rightarrow 2)$-β-D-glucan was constructed using the ϕ and ψ angles listed in Table II, which brought O-1 of residue 21 to within 1.94 Å of C-2 of residue 1. Minimization of the covalently closed molecule (Fig. 4E) resulted in a conformational energy of 45.0 kcal/mol and annealing (Fig. 4F) further reduced the energy to 31.3 kcal/mol. Use of prolonged MD simulations with temperature parameters above 200 K resulted in collapsed conformations with energies of less than 20 kcal/mol, but these contained non-A conformations and bore little resemblance the model of Palleschi and Crescenzi[5]. Thus, the cyclic $[AAAAAAB]_3$ conformation was found to be significantly more stable than the cyclic $[AAA]_7$ model.

The exo-anomeric effect, which is commonly accepted as being important in determining favored conformations of the glycosidic bond[22,23], results in stabilization (for glycopyranosidic bonds with equatorial glycosidic oxygens) of conformations that embody a ϕ angle of 60° or 180° and de-stabilization of conformations having a ϕ angle of $-60°$. Conformational energies calculated by the SYBYL force field do not include a term accounting for the exo-anomeric effect. Nevertheless, the ϕ angles embodied in the minimum energy structures calculated by SYBYL for the cyclic $(AAAB)_5$ eicosamer and the cyclic $[AAAAAAB]_3$ heneicosamer are almost exclusively in the highly favored 40° to 70° range. Conversely, one out of three ϕ angles in the energy minimized cyclic $[AAA]_7$ heneicosamer suggested by Palleschi and Crescenzi is constrained to the less favorable range of 8° to 27°. Thus, inclusion of the exo-anomeric effect would make the calculated stability of the $[AAAAAAB]_3$ model even greater compared to that of the $[AAA]_7$ model.

Cyclic $(1 \rightarrow 2)$-β-D-glucan models derived from the [AAAB] and [AAAAAAB] conformational repeat structures suggested in this paper have relatively small, polar central cavities (Figs. 4A–4D). Conversely, the relatively nonpolar cavity in the Palleschi and Crescenzi model (Figs. 4E–4F) of the $(1 \rightarrow 2)$-β-D-glucan is similar to that described for the cyclodextrins (cyclic $(1 \rightarrow 4)$-α-D-glucans), which have been widely reported to form complexes with a variety of nonpolar guest molecules[24]. However, only one preliminary report[25] is available regarding the putative interaction of cyclic $(1 \rightarrow 2)$-β-D-glucans with nonpolar guest molecules. The polar central cavity characteristic of the new models proposed here would be more likely to chelate or coordinate ions or form complexes with polar guest molecules. More experimental evidence is clearly necessary to distinguish between these possibilities.

The model of Palleschi and Crescenzi[5] does not provide a good description $(1 \rightarrow 2)$-β-D-glucans with degrees of polymerization that are not evenly divisible by 3. However, cyclic $(1 \rightarrow 2)$-β-D-glucans with between 17 and 24 residues are widely

TABLE IV

Populations (in percent) for the major conformers (90.4% of the total population) for the (1 → 2)-β-D-glucan eicosasaccharide in A, B, C space [a], as distinguished by PDP analysis

Sequence	Population
AAAAAAAAAAAAAAAAAAA	33.82
AAAAAAAAAAAAAAAAABA	6.14
CAAAAAAAAAAAAAAAAAA	4.67
BAAAAAAAAAAAAAAAAAA	3.43
AAAAAAAABAAAAAAAAAA	2.96
AAAAAAAAABAAAAAAAAA	2.83
AAAAAAAAABAAAAAAAAA	2.66
ABAAAAAAAAAAAAAAAAA	2.27
AABAAAAAAAAAAAAAAAA	2.26
AAAAAAAAAAAABAAAAAA	2.02
AAAAAAAAAAAAABAAAAA	1.76
AAAAABAAAAAAAAAAAAA	1.35
AAAAAAAAAAAAAAAAAAB	1.32
AAAAAAAAAAABAAAAAAA	1.29
ABAAAAAABAAAAABAAA	1.14
AAAAAAABAAAAAAAAAAA	1.08
AAAAAAAAAAAAACAAAA	1.07
BAAAAAAAAAAABAAAAAA	0.98
AAAAAABAAAAAAAAAAAA	0.83
AAAAAAAAABABAAABAAA	0.80
AABAAAAAAAABAAAAAAA	0.74
AAAAAAAABABAAABABA	0.72
AABAAAABAAAAABAAAAA	0.72
BBAAAAAABAAAAAABAAC	0.69
AABAAAAAAAAAABAAAAA	0.69
AABAAABAAAAAAAAAAAA	0.63
BAAAAAAAAAAAAAAAAAB	0.58
ABAAAAACBAAAAAAAAAA	0.57
AAAAAAAAABAAAAAAABA	0.53
AAABAAAAAAAABAAAAAC	0.52
CAAAAAAABAAAAAAAAAA	0.50
AAABAAAAAAAABAAAAAA	0.47
BCAAAAAAAAAAAAAAABB	0.47
AAAABAAAAAAAAAAAAAA	0.46
AAAACAAAAAAAAAAAAAA	0.45
AAAAACAAAAAAAAAAAAA	0.45
AAAAAAAAAABAAAAAABA	0.43
AAAAAAAAAAAAAAABAAA	0.41
AAABAAAAAAAAAAAAAAA	0.41
ABAAAAAABABAAAAABA	0.40
AAAAAAABAAAAABAAAAA	0.39
CAAAAAAAAAAAAAAAABA	0.38
BAAAAAAABAAAABAAAAA	0.38
CAABAAAAAAAAAAAAAAA	0.37
AABAAAAAAAABAAAAABA	0.35
CAAABAAAAAAAAAAAAAA	0.34
ABAAAAAAAABAAAAAABA	0.34
BBAAAAAAAAAAAAAAAAA	0.34
AAAAAAAAACAAAAAAAAA	0.33
AAAAAAAAAAABAAAAAAB	0.31
CAAAAAAAAAAAAAAAAAB	0.31
AAAAABAAAAAAAAAAAAC	0.27
AAAABCAABAAAAAABAAA	0.27
AAABCAAAAAAAAAAAAAA	0.25

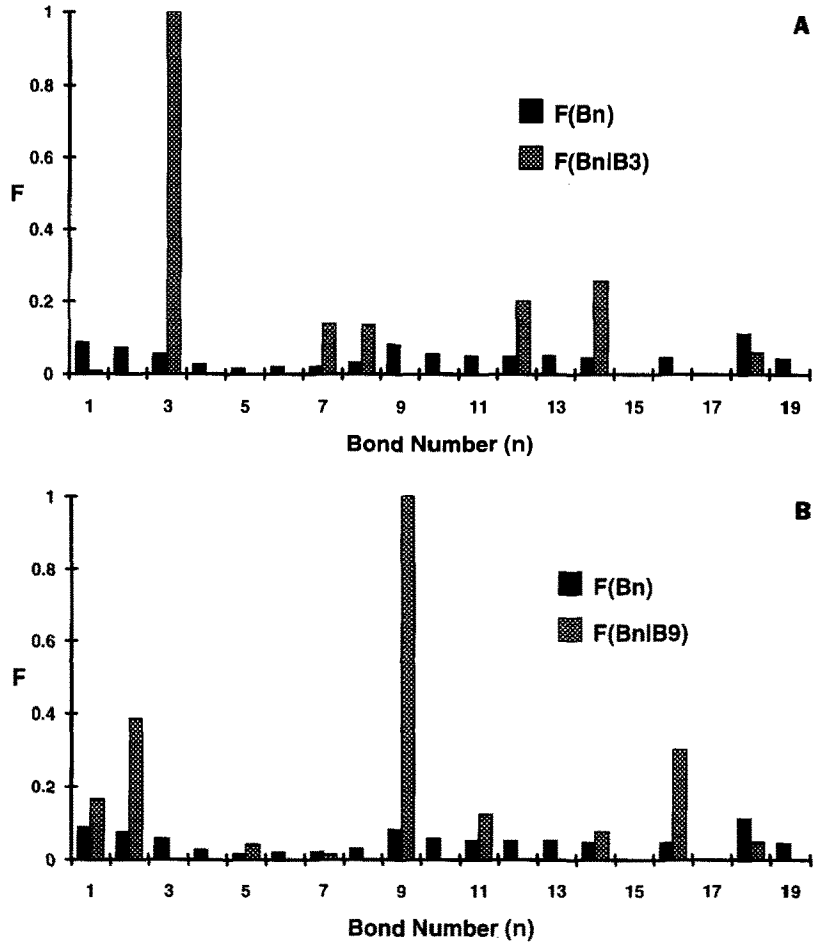


Fig. 7. Frequencies of glycosidic bond conformation B as a function of location in the (1 → 2)-β-D-glucan eicosasaccharide. See Fig. 3 for nomenclature.

reported[2,3,26]. Our model, based on the presence of the [AAAB] and [AAAAAAB] conformational structures may account for this diversity in size. Conformations of the form $[AAAB]_n[AAAAAAB]_m$ have the potential to form low-energy cyclic structures for glucans with 18 to 24 residues. The heptadecasscharide, which is typically found at very low levels, may represent an energetically strained species of the glucan. No cyclic (1 → 2)-β-D-glucans with fewer than 17 residues have been reported.

The different conformational models for the cyclic (1 → 2)-β-D-glucans are not necessarily exclusive. Only six resonances are present in the ^{13}C and ^{1}H NMR spectra of size homogeneous preparations of the (1 → 2)-β-D-glucans, indicating that rapid conformational interconversion of individual glucose residues is occurring in solution. The time-averaged molecular features observed during NMR

spectroscopy might represent the rapid interconversion of several conformational forms.

The algorithm employed in this study for partitioning the two-dimensional population-density matrices in the PDP program performed adequately on the data sets described herein. However, we have recently implemented an improved algorithm which employs a "hill climbing" strategy to detect local maxima (modes) in the population-density matrices and a "hill sliding" strategy to define the surrounding regions. This algorithm is faster, more robust, and uses fewer empirical parameters than the procedure outlined in the methods section (above).

CONCLUSIONS

Cluster analysis of large sets of molecular conformations that result from Metropolis Monte Carlo calculations for oligosaccharides can be exploited to provide insight into the general features of these molecules, including their conformational states and thermodynamic and kinetic behavior. This approach suggests that linear (1 → 2)-β-D-glucans have a tendency to adopt [AAAB] and/or [AAAAAAB] conformational repeat structures that give rise to energetically stable cyclic forms.

ACKNOWLEDGMENTS

This research is supported by United States Department of Energy (DOE) grant DE-FG09-85ER13426, by the DOE-funded (DE-FG09-87ER13810 and DE-FG09-93ER20097) Center for Plant and Microbial Complex Carbohydrates, and by a grant from the National Institutes of Health (1-P41-RR05351-01). The authors thank Drs. Alan G. Darvill and Peter Albersheim, Directors of the Complex Carbohydrate Research Center, for continued support of this research.

REFERENCES

1 K.J. Miller, E.P. Kennedy, and V.N. Reinhold, *Science*, 231 (1986) 48–51.
2 A. Dell, W.S. York, M. McNeil, A.G. Darvill, and P. Albersheim, *Carbohydr. Res.*, 117 (1983) 185–200.
3 M. Hisamatsu, A. Amemura, T. Matsuo, H. Matsuda, and T. Harada, *J. Gen. Microbiol.*, 128 (1982) 1873–1879.
4 R. Stuike-Prill and B. Meyer, *Eur. J. Biochem.*, 194 (1990) 903–919.
5 A. Palleschi and V. Crescenzi, *Gazz. Chim. Ital.*, 115 (1985) 243–245.
6 W.F. van Gunsteren and J.C. Berendsen, *Angew. Chem.*, 102 (1990) 1020–1055.
7 C.J. Edge, U.C. Singh, R. Bazzo, G.L. Taylor, R.A. Dwek, and T.W. Rademacher, *Biochemistry*, 29 (1990) 1971–1974.
8 S.N. Ha, L.J. Madsen, and J.W. Brady, *Biopolymers*, 27 (1988) 1927–1952.
9 S.W. Homans, *Biochemistry*, 29 (1990) 9110–9118.
10 S. Levy, W.S. York, R. Stuike-Prill, B. Meyer, and L.A. Staehelin, *Plant*, 1 (1991) 195–215.
11 L. Poppe, R. Stuike-Prill, B. Meyer, and H. van Halbeek, *J. Biomol. NMR*, 2 (1992) 109–136.

12 B. Meyer, M. Zsiska and R. Stuike-Prill, in D.P. Landau, K.K. Mon, and H.B. Schuttler (Eds.), *Computer Simulation Studies in Condensed Matter Physics IV*, 1993, in press.
13 T. Peters, B. Meyer, R. Stuike-Prill, R. Somorjai, and J.R. Brisson, *Carbohydr. Res.*, 238 (1993) 49–73.
14 A. Pollex-Krüger, B. Meyer, R. Stuike-Prill, V. Sinnwell, K.L. Matta, and I. Brockhausen, *Glycoconj. J.*, in press.
15 N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *J. Chem. Phys.*, 21 (1953) 1087–1092.
16 H.L. Gordon and R.L. Somorjai, *Proteins: Struct. Funct. Genet.*, 14 (1992) 249–264.
17 J.B. Lambert, R.J. Niehuis, and J.W. Keepers, *Angew. Chem., Int. Ed. Eng.*, 20 (1981) 487–500.
18 J. Zupan, *Clustering of Large Data Sets*, Research Studies Press, New York, NY, 1982.
19 J. Zupan and M.E. Munk, *Anal. Chem.*, 57 (1985) 1609–1616.
20 A.K. Jain, in T.Y. Young, and K.-S. Fu, (Eds.), *Handbook of Pattern Recognition and Image Processing*, Academic Press, 1986, 33–57.
21 R. Dubes and A.K. Jain, *Pattern Recog.*, 8 (1976) 247–260.
22 R.U. Lemieux, K. Bock, L.T.J. Delbaere, S. Koto, and V.S. Rao, *Can. J. Chem.*, 58 (1980) 631–653.
23 K. Bock, *Pure Appl. Chem.*, 55 (1983) 605–622.
24 S.P. van Helden, M.J. van Drooge, A.J. Claessens, A.C.A. Jansen, and L.H.M. Janssen, *Carbohydr. Res.*, 215 (1991) 251–260.
25 J. Koizumi, Y. Okada, S. Horiyama, T. Utamura, T. Higashiura, and M. Ikeda, *Incl. Phenomena*, 2 (1984) 891–899.
26 M.W. Breedveld, L.P.T.M. Zevenhuizen, and A.J.B. Zehnder, *Appl. Environ. Microbiol.*, 56 (1990) 2080–2086.