ELSEVIER

Carbohydrate Research 276 (1995) 1–29

CARBOHYDRATE RESEARCH

Travelling on the potential energy surfaces of carbohydrates: comparative application of an exhaustive systematic conformational search with an heuristic search

Søren B. Engelsen [a], Jaroslav Koca [b], Isabelle Braccini [c], Catherine Hervé du Penhoat [c], Serge Pérez [a,*]

[a] *Ingénierie Moléculaire, Institut National de la Recherche Agronomique, BP527, 44026 Nantes Cedex 03, France*

[b] *Department of Organic Chemistry, Masaryk University, 611 37 Brno, Czech Republic*

[c] *Laboratoire de Chimie de l'Ecole Normale Supérieure, 24 rue Lhomond, 7523 Paris Cedex 05, France*

Received 19 December 1994; accepted 13 February 1995

Abstract

The calculated ensembles found by a heuristic conformational search algorithm, CICADA, for three small carbohydrates, ethyl β-lactoside, methyl α-D-galactoside, and methyl β-D-galactoside, are evaluated in terms of their ability to reproduce time-averaged optical rotation and NMR data. A unique dynamic model for methyl β-D-galactoside has been obtained by fitting experimental NOESY volumes to the theoretical ones elaborated from the CICADA ensemble internuclear distances with the model-free formalism. In the case of ethyl β-lactoside, the CICADA ensemble is compared to that of an exhaustive systematic grid-search method. The CICADA algorithm proved to be a very efficient method to find most of the important minima on even very complex potential energy surfaces, and the spectral quality of the CICADA ensemble was found to be of equal quality, if not superior, to that of the exhaustive systematic grid-search method. The CICADA algorithm has several advantages over other conformational search algorithms: (1) It has polynomial dependence of dimensions on computer time in contrast to the grid search, which has exponential dependence, (2) the conformations found are free of artificial harmonic constraint potentials, (3) it passes all barriers amongst families of conformations on conformational hypersurface but spends almost all its time in the essential highly populated areas, (4) the inherent properties of the algorithm make rigorous minimization criteria superfluous and provide good

* Corresponding author.

0008-6215/95/$09.50 © 1995 Elsevier Science Ltd. All rights reserved
SSDI 0008-6215(95)00152-2

convergence behavior, and (5) as an important spin-off, it provides low-energy interconversion pathways that can, amongst others, be used for estimating adiabatic rotational barriers.

Keywords: Molecular mechanics; Conformational search; Optical rotation; NMR; Ethyl β-lactoside; Methyl α-D-galactoside; Methyl β-D-galactoside

1. Introduction

"The potential energy surface (PES) is the meeting ground of chemistry with molecular physics and spectroscopy. It determines the molecular structure, the rotational and vibrational motions and the kinetics and dynamics of nuclear motions, and hence it controls the path of chemical reaction" [1]. The PES of a given molecule can be determined by ab initio quantum chemistry calculations or by backcalculations from the molecular spectra, although both methods are limited by the size of the molecular system to be determined. The empirical molecular mechanics method is able to combine the information provided by the two methods and extrapolate the results to much larger systems typical of biologically active molecules. The result of such an effort is called a molecular mechanics force field or a set of potential energy functions describing interatomic forces as a function of the position of the nuclei. The multivariate nature of the molecular mechanics force field makes it nontrivial to describe the PES of a given molecule. The lack of a generalized analytical solution to functions of multiple dimensions, called the multiple minima problem, is the most serious complication in describing the PES of large flexible molecules [2–5]. In the case of the carbohydrates, which are characterized by an exceptionally high ratio of hydroxyl groups per residue, the multi-conformational problem is always a major difficulty. The conformational flexibility is further increased by the number of low-energy conformations that occur at any glycosidic junction. Assuming rigid ring geometry, the typical monomer has 3^7 or 729 possible idealized staggered conformations, and the typical dimer has 3^{12} or 531,441. For polymers the exponential dependence on dimensions makes the use of grid-search methods unfeasible.

As the complexity of the PES increases exponentially with the dimension of the molecular system [6], a variety of heuristic search methods have been proposed to solve this problem: the tree branch search [7–9], the build-up method [10], quenched molecular dynamics [11], the cartesian stochastic search [12,13], the internal coordinate Monte Carlo search [14,15] and the increase of dimensionality [16]. Last but not least two biological approaches should be mentioned: simulated annealing [4,17,18], named by the analogy with the process of cooling physical substances (and how the state of the system depends on the rate at which the temperature is dropped), and the genetic search [19,20] based on the principles of Darwin's evolution theory (the survival of the fittest). In general the search algorithms exist in several variations and hybrid combinations.

Recently one of the authors has developed a conformational search method *CICADA* (Channels In Conformational space Analyzed by Driver Approach) [21], which best can be characterized as an optimized tree branch search with full memory. The method selects the lowest energy conformer of all (not yet fully explored conformations) to

perform new mutations, and it uses graph theory to avoid redundant calculations and to obtain interconversion pathways on the PES. This is important as the classical partition function of a given type of conformation is proportional to the volume of the local energy well on the PES and thus to the shape of the PES. To calculate spectral (time-averaged) properties, it is therefore necessary to sample the low-energy domains on the PES as completely as possible. The CICADA algorithm has recently been applied to carbohydrate molecules in the context of establishing flexibility indexes [22]. Before applying the method to larger and more biologically relevant systems, it was necessary to undertake the comparison of the quality of the spectral sampling and efficiency of the CICADA search ensemble with that of a fine grid-search ensemble [23] using the MM3 [24,25] PES of a small, model carbohydrate molecule.

In the present study the spectral quality of the ensembles is tested against experimental NMR and optical rotation data obtained in solution. NMR-determined conformations of molecules in solution rely on the internuclear distances extracted from relaxation data and the dihedral angles established from vicinal coupling constants. The latter parameters reflect a time-averaged structure and can be expressed with empirical Karplus-type relationships averaged over all microstates of conformational space. In contrast, experimental NOESY volumes or cross-relaxation rates can only be directly related to internuclear distances in the absence of internal motion [26]. However, it has been shown recently that fitting of theoretical homonuclear relaxation data elaborated from time-averaged distance matrices with the model-free formalism leads to reasonable motional models for small carbohydrates [23,27–29].

The carbohydrate molecules under investigation are ethyl β-lactoside and the methylated α and β anomers of one of its constituents: methyl α-D-galactoside and methyl β-D-galactoside. With its (1 → 4) linkage between β-D-galactopyranose and D-glucopyranose, ethyl β-lactoside serves as a general conceptual model for equatorial–equatorial (1 → 4) linked disaccharides. Besides their industrial importance, lactose and its derivatives have been the subject of many experimental investigations from which a wealth of information is available.

2. Methods

Nomenclature.—Ethyl β-lactoside. A schematic representation of ethyl β-lactoside (ethyl β-D-galactopyranosyl-(1 → 4)-β-D-glucopyranoside), along with the labeling of

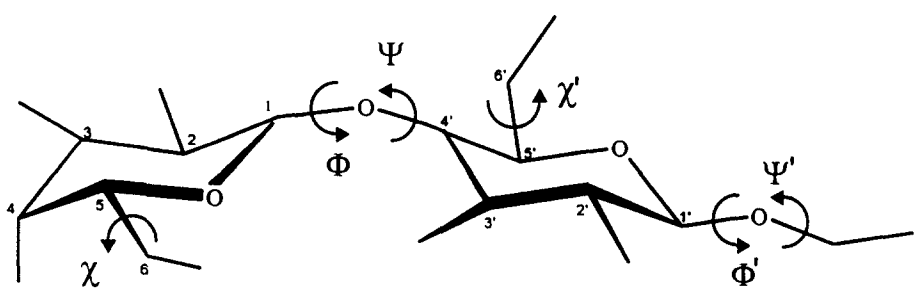


Fig. 1. Important torsional angles and atomic numbering of ethyl β-lactoside.

the atoms and the torsional angles of interest, is shown in Fig. 1. The sign of the torsion angles is defined in agreement with the IUPAC–IUB Commission of Biochemical Nomenclature [30]. The conformation about the glycosidic linkage bonds are described by the following torsion angles:

Φ = O-5–C-1–O-1–C-4′

Ψ = C-1–O-1–C-4′–C-5′

The orientations of the hydroxymethyl groups are described by the torsional angles:

χ = O-5–C-5–C-6–O-6

χ' = O-5′–C-5′–C-6′–O-6′

The orientation of the ethyl substituent at the reducing end of the disaccharide is described by the three torsional angles:

Φ' = O-5′–C-1′–O-1′–C-7′

Ψ' = C-1′–O-1′–C-7′–C-8′

ω' = O-1′–C-7′–C-8′–H-8′

Methyl α-D-galactopyranoside and methyl β-D-galactopyranoside.—For the α and β anomers of the methyl galactosides, practically the same nomenclature is applied. Two main exocyclic dihedrals, Φ and χ, determine, respectively, the orientation of the methoxy group and the primary hydroxyl group:

Φ = O-5–C-1–O-1–CMe

χ = O-5–C-5–C-6–O-6

The orientations of the primary hydroxyl groups are also referred to as either *gauche–gauche* (GG), *gauche–trans* (GT) or *trans–gauche* (TG). In this terminology, the dihedral O-5–C-5–C-6–O-6 is considered first, and the dihedral C-4–C-5–C-6–O-6 is considered second. To describe the orientation of all other torsional angles, we use the letter code definition given in Fig. 2.

Experimental.—The NMR data of ethyl β-lactoside was previously reported [23]. ^{1}H (400.13 MHz) and ^{13}C NMR (100.6 MHz) spectra of a 50 mM solution of methyl β-D-galactoside (Sigma Chemical Co.) in D_2O (99.96%) were recorded with a Bruker

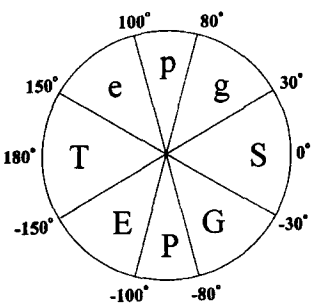


Fig. 2. The conformational wheel. Letter code for significant intervals of the dihedral angles.

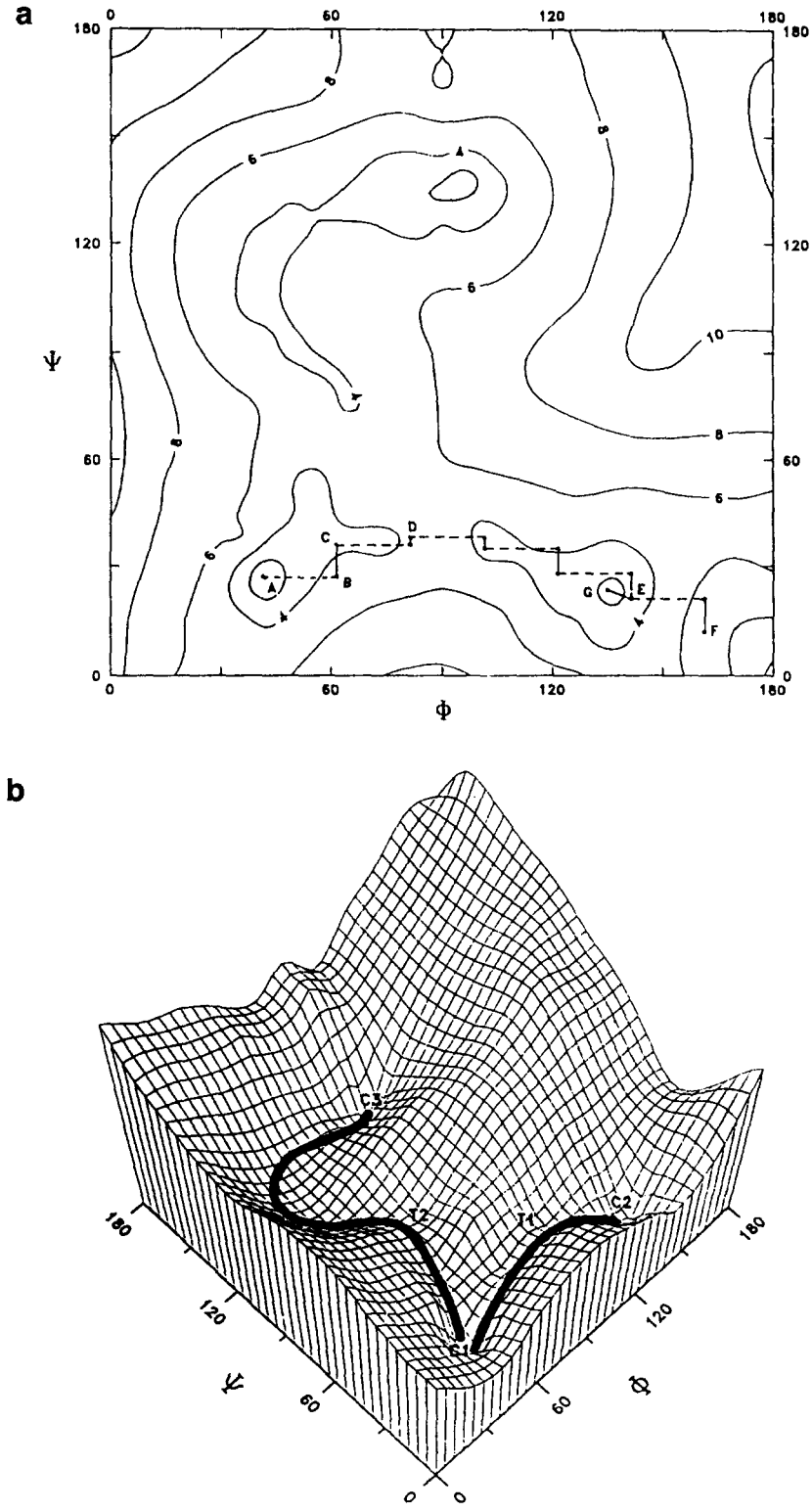


Fig. 3. (a) Contour plot of the CICADA travelling on a model surface. (b) Pathways on the model surface.

instrument operating in the Fourier-transform mode at 296 K. Chemical shifts and vicinal coupling constants, $^3J_{H,H}$, were extracted from a spectrum with a digital resolution of 0.12 Hz/pt. Simulation of the 1D spectrum using the Bruker package PANIC was required to obtain the coupling constants and chemical shifts of strongly coupled protons. ^{13}C T_1 measurements were acquired with the inversion-recovery sequence (180–τ–90–FID), and relaxation times were calculated with the Bruker T_1 routine.

Two sets of phase-sensitive NOESY spectra [31] were acquired with mixing times of 0 and 1 s. A 20 ms variable delay was introduced at the beginning of the mixing time in order to suppress J-peak transfer [32]. The recycle time was set to 5 times the longest T_1 in order to obtain a symmetrical normalized NOESY volume matrix. The diagonal and cross-peak intensities were evaluated from the summed ω_1 subspectra contributing to a specific signal. The standard (average) deviations were calculated for the cross-peak volumes, $a_{k,l}$ (diagonal volumes, $a_{k,k}$). Integration of noise in these spectra gave ± 0.005 units of normalized intensity with respect to the diagonal volume ($\tau_m = 0$ s) for typical peak widths.

Computational methods.—The conformational behavior of ethyl β-lactoside was investigated using the MM3 force field [24,25]. This is a highly detailed force field, which in addition to the classical terms, includes anisotropy of hydrogens, corrections for stereoelectronic effects, cross-term effects like torsion-stretch, torsion-bend and bend-bend interactions. Buckingham type potential for nonbonded interactions and explicit terms for hydrogen bonding. No cut-off of nonbonded interactions was used, and the dielectric constant was set to 4.0 [33].

Grid search.—A complete conformational grid search on ethyl β-lactoside in the MM3 force field was reported previously [23]. In this work an adiabatic map was computed from 36 individual relaxed maps. The relaxed maps were computed using rigid rotation followed by harmonic constraint minimization in 10° increments for Φ and Ψ spanning the whole angular range. Convergence was accepted when the iterative decrease in energy ΔE was less than $N \times 0.00008$ kcal/mol; where N was the number of atoms in the molecule. This approach resulted in 46 656 possible conformations of which 34 443 were accepted for further calculations on the basis of the criterion of relative energy (≤ 20 kcal/mol) and values of χ and χ', redundant couples (χ, χ') being discarded (these torsion angles were not constrained in the relaxed map calculations and transitions occurred).

CICADA search.—The CICADA program provides an intelligent shell for the underlying molecular mechanics or quantum chemistry program. The details of the algorithm are described elsewhere [21]. Fig. 3 describes how the CICADA algorithm travels on a 2D model surface. The search begins from the lowest energy conformer available, say point A or C1 (in practise it can be any conformation). A dihedral angle Φ_i is then selected, based on the lowest expectation value of the rotational energy barrier (random at the beginning). The selected dihedral is rotated by a predefined step, $\Delta\Phi$, and a new point (B) is reached. The molecule is then relaxed, with a sole constraint namely the dihedral angle Φ in the search direction, and a new point is reached (C). This process is repeated until the new point is lower in energy than the previous in which case a saddle point or transition state has been encountered. The previous (highest

energy) point (D) is stored as an approximation to the transition state (T1) and the down-hill travelling is continued until a new point is reached (F) where the energy increases again. The previous (lowest energy) point (E) on the pathway is then fully relaxed and the true minimum (G or C2) is obtained. From the point of view of graph theory, three geometries are important, the two minima (A and G) and the transition-state approximation (D), since these constitute vertices in the graph. All three points are saved together with the pathway (in the form A–D–G). The minimum found can be either a new one or one found previously, of which the lower energy minimum is selected for addition to the graph. Having found a new minimum (G), CICADA will proceed from the lowest energy conformation found so far (A or G) with available new directions. The process is repeated for all the minima within predefined energy limits and for all selected dihedral angles in both possible directions (this results in $2N$ pathways from a given minimum, where N is the number of driven dihedrals).

The CICADA process is controlled by several parameters of which the most important are the selected set of driven dihedrals, the identity criterion D_{min}, the dihedral step increment $\Delta\Phi$, and the stop criterion E_{conf}. It is up to the user to define the set of driven dihedrals in the CICADA search, an option that can be used to make searches on large systems more efficient. In the case of modeling large carbohydrates, it is sensible to exclude dihedrals of hydroxyl groups in the preliminary conformational search, thereby avoiding unnecessary time spent on rotating small side groups in regions of PES of insignificant importance. The identity criterion is used to determine if two conformations are identical; it is formulated as an angle criterion that must be greater than all the differences amongst dihedrals of two conformations. The dihedral step increment is the step size used by the dihedral driver, and this is typically set to 20°, with the result that we are not likely to observe much finer details in resulting transition states. It has to be emphasized that it is not a regular grid increment as all local minima are fully relaxed. The stop criterion is based on relative energy of conformations; if no more conformations are found in an energy window defined with respect to the global minimum then the algorithm will stop the search. The CICADA algorithm will save all conformations and transition states found in the search in standard coordinate files and in special geometry files containing energy and values of selected dihedral angles. Finally, for further advanced analysis CICADA will save a graph containing the accumulated information on interconversion pathways and transition states amongst the conformations.

Data simulation.—All simulated data were calculated using a Boltzmann distribution of complete molecular statics ensembles.

Optical rotation. The optical rotation of the molecules at 589 nm were calculated using a semi-empirical model described by Stevens and Sathyanarayana [34]. The model used is dependent on the topology and geometry and refers to molecules in vacuo, although a correction for solvent water is applied.

Coupling constants. Coupling constants, $^3J_{H,H}$ for vicinal hydrogen atoms of a H–C–C–H segment were calculated using a Karplus type equation [35] with the Haasnoot–Altona parameterization [36]. It accounts for the J dependence on the dihedral angle of the H–C–C–H fragment, on the electronegativity of the participating atoms, and on the orientation of the α and β substituents. The heteronuclear coupling

constant $^3J_{H,C}$ across the glycosidic linkage was calculated by using the equation for the heteronuclear C–O–C–H segment proposed by Tvaroska et al. [37].

Rotamer distributions about primary hydroxyl groups were calculated numerically from the limiting values given by the Haasnoot–Altona equation [36,38] by solving the set of linear equations:

$$\begin{matrix} 0.9 & 10.7 & 5.0 \\ 2.8 & 3.1 & 10.7 \\ 1.0 & 1.0 & 1.0 \end{matrix} \times \begin{matrix} p_{GG} \\ p_{GT} \\ p_{TG} \end{matrix} = \begin{matrix} J_{5.6R} \\ J_{5.6S} \\ 1.0 \end{matrix}$$

where the p's are the fractional populations and the J's are the coupling constants in Hz. For consistency we chose to recalculate all rotamer distributions using this set of equations. The results (Table 1 and Table 2) include negative populations arising from the inherent inaccuracies in the Haasnoot–Altona description and from the simplified model in which only idealized staggered conformations are considered.

Table 1
Observed and calculated [a] ^3J coupling constants of ethyl β-lactoside

^3J	Expt. [e]	MM3/grid	MM3/cic conf	MM3/cic conf + trans
N	—	34.443	1.586	4.457
$^3J_{HH}$				
	β-Gal p			
H-1–H-2	7.8	7.9	7.8	7.8
H-2–H-3	9.8	9.5	9.5	9.5
H-3–H-4	3.4	5.0	4.9	4.9
H-4–H-5	0.5	1.1	1.1	1.1
H-5–H-6S	3.2	2.6	2.7	2.5
H-5–H-6R	8.6	7.8	8.9	9.2
GG:GT:TG [b]	—	21:76:3	8:87:5	6:91:3
GG:GT:TG [c]	20:78:2	33:73:-5	21:83:-4	19:88:-7
	β-Glc p			
H-1′–H-2′	8.0	7.7	7.8	7.8
H-2′–H-3′	9.8	9.2	9.2	9.2
H-3′–H-4′	9.3	9.2	9.1	9.1
H-4′–H-5′	9.3	10.0	9.9	9.9
H-5′–H-6′S	2.1	3.4	3.3	3.1
H-5′–H-6′R	5.0	6.8	6.7	6.9
GG:GT:TG [b]	—	33:57:10	32:58:10	31:60:9
GG:GT:TG [c]	64:46:−10	37:58:5	38:57:4	38:61:1
$^3J_{CH}$				
H-1–C-4′	3.7	3.5	3.4	3.3
H-4′–C-1	4.9 [d]	4.2	4.6	4.6
H-1–C(Et)	4.4	3.2	3.2	3.2

[a] Calculated using the Haasnoot–Altona parametrization of the Karplus equation.
[b] Calculated ensemble distributions (GG if $240 < \chi < 360$, GT if $0 < \chi < 120$ and TG if $120 < \chi < 240$).
[c] Back-calculated population distributions. See Methods section.
[d] From ref. [58].
[e] From ref. [23].

Table 2
Observed and calculated [a] ^3J coupling constants of methyl α-D-galactoside and methyl β-D-galactoside [b]

^3J	Me α-Gal *p*					Me β-Gal *p*				
	Expt.			MM3/ cic conf	MM3/ cic conf + trans	Expt.			MM3/ cic conf	MM3/ cic conf + trans
	c	d	e			c	e	f		
N			—	298	1172			—	179	1379
$^3J_{HH}$										
H-1–H-2			3.0	3.7	3.7		8.0	7.9	7.8	7.8
H-2–H-3			9.8	9.6	9.7		10.0	9.9	9.5	9.5
H-3–H-4			2.3	4.8	4.8		3.8	3.4	4.9	4.9
H-4–H-5			1.0	1.2	1.2		0.8	1.0	1.0	1.0
H-5–H-6S	3.7	6.0	4.6	2.8	2.8	4.83	4.4	4.4	2.9	2.9
H-5–H-6R	8.6	7.8	8.6	7.8	7.8	7.50	7.6	7.9	8.1	7.9
H-6R–H-6S								−11.7		
GG:GT:TG [g]			—	21:73:6	21:73:6			—	17:76:7	18:75:7
GG:GT:TG [h]			16:75:9	31:72:−3	31:72:−3			19:63:18	27:74:−1	29:72:−1
$^3J_{CH}$										
H-1–CMe			3.3	3.0	3.0		3.4		3.1	3.1

[a] Calculated using the Haasnoot–Altona parametrization of the Karplus equation.
[b] Carbon T_1 values for Me β-Gal *p* are as follows: 1.05 (C-1), 1.05 (C-2), 1.01 (C-3), 0.94 (C-4), 1.05 (C-5) and 0.63 (C-6).
[c] Primary hydroxyl pro-*R*/pro-*S* data from ref. [38].
[d] Primary hydroxyl pro-*R*/pro-*S* data from ref. [61].
[e] Ring couplings and heteronuclear coupling from ref. [62].
[f] Simulated for a second-order spin system using the Bruker package PANIC.
[g] Calculated ensemble distributions (GG if $240 < \chi < 360$, GT if $0 < \chi < 120$ and TG if $120 < \chi < 240$).
[h] Back-calculated population distributions. See Methods section.

NOESY volumes. In order to account for internal motions in the treatment of relaxation data [39], it is necessary to know the relative rates of overall tumbling, τ_c, and conformational exchange, $\tau_i(k,l)$. Both the averaging method for the time-dependent geometrical parameters and the appropriate spectral density functions depend on these relative rates. The computational methods appropriate for both the slow internal motion model ($\tau_i(k,l) \gg \tau_c$) and the fast internal motion model ($\tau_i(k,l) < \tau_c$) are as follows.

For the slow internal motion model and isotropic tumbling, the spectral densities are a function of the molecular reorientation time

$$J_n(\omega) = 2\tau_c / \left(1 + (n\tau_c\,\omega)^2\right) \tag{1}$$

In the case of the *fast internal motion model*, the spectral densities were expressed with the model-free formalism *[40]*. When internuclear one-bond proton–carbon distances can be considered to be constant (1.11 Å), heteronuclear relaxation data can be described by:

$$J(\omega) = 2/5\left\{S^2\tau_c / \left(1 + (\tau_c\,\omega)^2\right) + (1 - S^2)\tau / \left(1 + (\tau\omega)^2\right)\right\} \tag{2}$$

where S^2 and τ_e are the generalized order parameter and effective correlation time of internal motion ($\tau^{-1} = \tau_c^{-1} + \tau_e^{-1}$), respectively.

For homonuclear relaxation in the case of the *fast internal motion model*, fluctuations in internuclear distances must be taken into account. The expressions for the spectral densities and the generalized order parameter become

$$J(\omega) = 2/5\Big\{S^2(\mathrm{k,l})\tau_c/\big(1 + (\tau_c\omega)^2\big) + \big(\langle r^{-6}\rangle - S^2(\mathrm{k,l})\big)\tau(\mathrm{k,l})/\big(1 + (n\tau(\mathrm{k,l})\omega)^2\big)\Big\} \quad (3)$$

When the angular, S^2_{ang}, and radial contributions of the dipole interactions are assumed to be independent, S^2 is;

$$S^2(\mathrm{k,l}) = S^2_{ang}(\mathrm{k,l})\langle r_{kl}^{-3}\rangle^2 \quad (4)$$

In these equations the angular order parameter matrix, $S^2_{ang}(\mathrm{k,l})$,, characterizes the amplitudes of the internal motions ($S^2_{ang}(\mathrm{k,l}) = 0$ corresponds to a totally flexible molecule, whereas $S^2_{ang}(\mathrm{k,l}) = 1$ corresponds to a rigid molecule). The theoretical NOESY values were computed using the full relaxation matrix method [41] with in-house software without a correction for leakage due to relaxation mechanisms other than dipole–dipole relaxation. Two approaches were used to optimize the fit between theoretical and experimental data. In the first one (which was a trial-and-error approach), all volumes corresponding to interactions between weakly coupled spins were fit to roughly 110% of the experimental values in order to account for the leakage contribution. In the second procedure, motional models were generated randomly, and optimized values for $S^2_{ang}(\mathrm{k,l})$ and $\tau(\mathrm{k,l})$ were obtained by least-squares fitting of the theoretical NOESY volumes, obtained by back-calculation of the relaxation matrix, to the experimental ones with the fitting algorithm as previously described [29]. In both cases, structural groups (methylene or methyl groups or ring protons) are defined which share a common value of $S^2_{ang}(\mathrm{k,l})$ and $\tau(\mathrm{k,l})$.

3. Results and discussion

NMR data.—Coupling constants for both ethyl β-lactoside and methyl β-D-galactoside, which were obtained by simulating the experimental spectra (0.12 Hz/pt), are collected in Tables 1 and 2, respectively. The low-field methylene resonances of the galactosyl ring were assigned to the pro-*R* protons in order to conform to recent work [38]. Perusal of the literature brings to light a certain variability in the methylene H-5–H-6*R* and H-5–H-6*S* coupling constants of galactosyl derivatives, which may be due to the influence of concentration on the rotamer populations.

The carbon T_1 values for methyl β-D-galactoside (Table 2) were measured in order to evaluate the molecular reorientation time as previously described [23]. It can be seen that carbons bearing axial protons have almost the same T_1 values (1.05 s), whereas the C-4, the only carbon with an equatorial proton, has a shorter longitudinal relaxation time

(0.94 s, 90%). A similar tendency has been reported for galactose (the T_1 of C-4 is roughly 90% of the other ring carbons) [42]. A *rigid-molecule* correlation time of 50 ps is obtained from the average T_1 value of carbons bearing an axial proton when the carbon–proton bond length is taken to be 1.11 Å. In order to obtain this value from the C-4 T_1 longitudinal relaxation time, it is necessary to assume a carbon–proton bond of

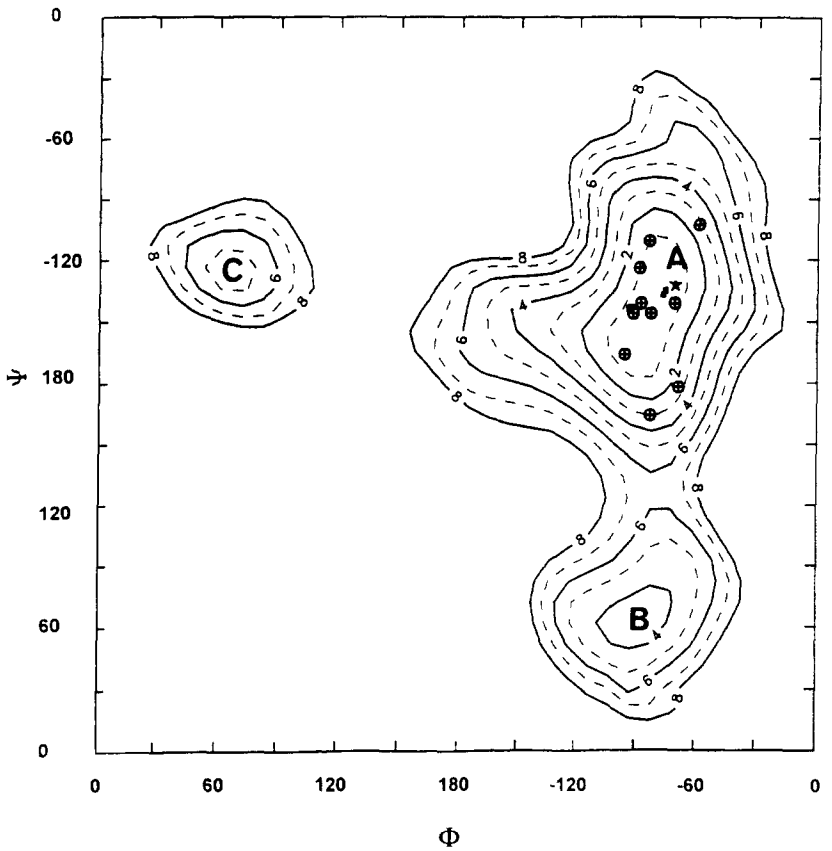
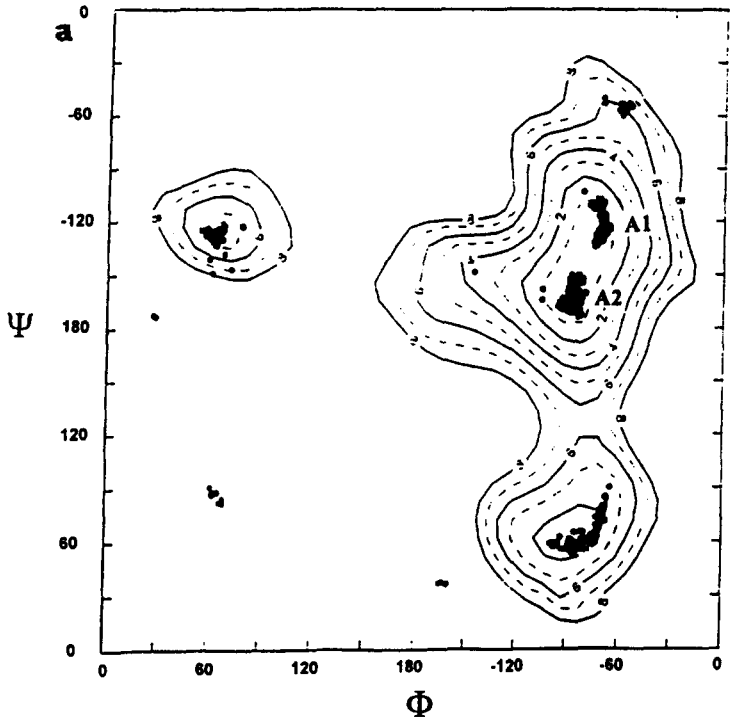
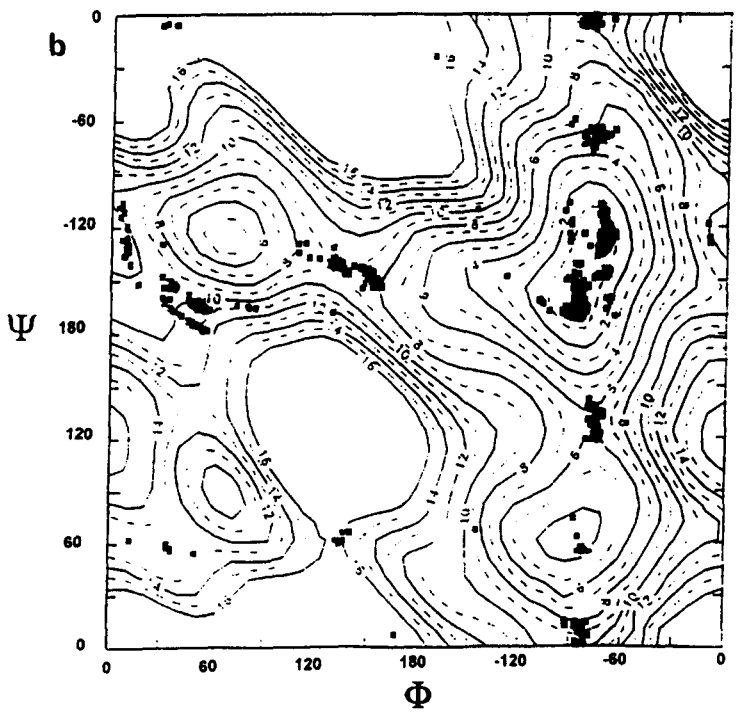


Fig. 4. Adiabatic map of ethyl β-lactoside in the MM3 force field with indication of crystal structures of lactose: (★) the most refined crystal structure of β-lactose [47], (■) crystal structures of α,β-lactose [48–52] and (⊗) crystal structures of lactose found in proteins [53–57]. Isoenergy contours are drawn in 1 kcal/mol intervals out to 8 kcal/mol from the global minimum. The PES is a simple three-well system. The most important well is the almost rectangular **A** well, combining voluminous conflict free conformational space with O-5–O-3′ interresidue hydrogen bonding and optimal exo-anomeric configuration. All crystal structures found so far belong to this well. The highest energy crystal conformation is about 3 kcal/mol above the global minimum. The two other main features are the **B** and the **C** wells. The lowest energy minimum in the **B** well is calculated to be approximately 3.5 kcal/mol above the global minimum, and so far no lactose crystal structure has been found in this well. The **C** well which is disconnected from the main **A** well by a relatively high energy barrier is typically staggered about Φ but in a non-optimal exo-anomeric configuration. Since the **C** well offers no possibility for creation of strong hydrogen bonds, this non-ideal, exo-anomeric position is more than 4 kcal/mol above the global minima.


a
A1
A2
Ψ
Φ
0
-60
-120
180
120
60
0
60
120
180
-120
-60
0


b
Ψ
Φ
0
-60
-120
180
120
60
0
60
120
180
-120
-60
0

1.09 Å. It has been recently shown in an *ab initio* study [43] that the equatorial C–H bond in 2-methoxytetrahydropyran is shorter than the corresponding axial bond.

In the case of the lactosyl derivatives, the C-4 T_1 values are even shorter [23,44], roughly 20% of the other T_1 values at 400 MHz. A detailed study of the multi-field longitudinal relaxation parameters of lactosyl derivatives has shown that molecular reorientation is slightly anisotropic [44], but simulations have indicated that this slightly anisotropic nature of overall tumbling has a negligible effect on homonuclear relaxation data [45,46]. The molecular reorientation time of 0.24 ns, previously established [23] for ethyl β-lactoside based on heteronuclear NOE data, was used in the present work.

The grid search and CICADA search ensembles.—Ethyl β-lactoside (grid search). The adiabatic map of ethyl β-lactoside as a function of Φ and Ψ resulting from grid search [23] is shown in Fig. 4. On the figure is indicated the positions of all crystal conformations of lactose found so far [47–57], including those found in carbohydrate–protein complexes.

Ethyl β-lactoside (CICADA) search of the glycosidic linkage. Fig. 5a,b show respectively the conformations and transition states found by the CICADA conformational search, superimposed on the adiabatic (Φ,Ψ)-map calculated from the systematic grid search. As can be seen from the figures, CICADA concentrates on very minor areas of (Φ,Ψ)-space, namely the populated areas. The conformations occur mainly in the centers of the **A**, **B** and **C** wells with only very minor concentrations in the staggered ($-60°$; $-60°$) position and in the (60°; 90°) position. The former is lowest in energy, being only about 6 kcal/mol from the global minimum, whereas the latter is more than 9 kcal/mol from the global minimum (see Table 3). The remote (energetic) areas or spread islands of CICADA conformers are in fact small pockets trapped in between strained orientations [23]. However, these remote areas are not believed to be of any significance in spectral calculations of β-lactoside; in fact it has not yet been proven that the **B** and **C** wells have any experimental significance. The most important well **A**, which encompasses all the crystalline conformations of lactose, is divided into two regions by the CICADA search. The northern region centered about ($-70°$; $-125°$) will from now on be labelled **A1**, and the southern region centered about ($-85°$; $-160°$) is labelled **A2**. Interestingly, the global minimum found by the grid search belongs to the **A2** region, whereas the global minimum found by the CICADA search belongs to the **A1** region. These two "global minima" can be compared by inspection of Table 4, where it can be seen that the global minimum found by the CICADA search is about 0.26 kcal/mol lower in energy than that found by the grid search and that the most refined crystal structure of β-lactose is located in the **A1** well (see Fig. 4). The shift in the Ψ-angle between the global grid search conformer (A2) and the global CICADA conformer (A1) seems to be induced by a transition of the HO-6 orientation on the *galactose* residue. In the "true" global minimum (CICADA), the primary hydroxyl group HO-6 is pointing towards the O-3′, which therefore acts both as a donor and as an acceptor of hydrogen bonds. The ten lowest lying conformers listed in Table 3 possess

Fig. 5. (a) CICADA conformations (●) and (b) transition states (■) superimposed on the adiabatic map. Isoenergy contours are drawn in 1 kcal/mol intervals out to respective (a) 8 kcal/mol and (b) 16 kcal/mol.

Table 3
List of conformations of ethyl β-lactoside found by the CICADA search. Conformations contributing more than one per cent to the total population are listed in descenting order together with a few local minima from other regions

Φ_i	Ψ_i	Lac [a]	Glc p [b]	Gal p [c]	ΔE [d]	p_i [e]
−69	−127	gGTg	GgG	TTTG	0.0000	6.256
−69	−125	gGTG	Ggg	TTTG	0.3432	3.505
−70	−126	gGTg	TgG	TTTG	0.3720	3.339
−86	−162	gGTG	Gpg	TTTT	0.4536	2.909
−69	−126	gGTg	GgT	TTTG	0.4580	2.888
−85	−161	gGTg	TpG	TTTT	0.5874	2.321
−84	−161	gGTg	GpT	TTTT	0.7023	1.912
−69	−125	gGTG	Tgg	TTTG	0.7518	1.759
−70	−126	gGTg	TgT	TTTG	0.7644	1.722
−70	−126	gGTG	GgT	TTTG	0.7881	1.654
−86	−162	gGTG	Tpg	TTTT	0.9087	1.350
−91	−169	gGTg	TTG	TTTG	0.9470	1.265
−71	−127	gGTt	GgT	TTTG	0.9781	1.200
−85	−159	gGTg	GpG	TTTg	1.0121	1.133
−70	−127	GGTg	GgG	TTTG	1.0224	1.114
−71	−124	gGTg	TTG	TTTG	1.0504	1.063
−85	−164	gGTg	TpT	TTTT	1.0789	1.013
−88	56	gGTg	TTG	TTTG	2.1913	0.155
64	−126	gGTg	TTG	TTTG	3.7461	0.011
−55	−54	gGTG	Tgg	TTTG	5.7801	0.000
62	87	gGTg	GgG	TTTT	9.1155	0.000

[a] Most important dihedral angles defined as follows: χ, Φ', Ψ' and χ'. Letter code is written according to Fig. 2.
[b] Exocyclic hydroxyl dihedrals on the *glucose* moiety. The dihedrals are defined as follows: (1) C-3′–C-2′–O-2′–HO-2′ (2) C-4′–C-3′–O-3′–HO-3′ and (3) C-5′–C-6′–O-6′–HO-6′.
[c] Exocyclic hydroxyl dihedrals on the *galactose* moiety. The dihedrals are defined as follows: (1) C-3–C-2–O-2–HO-2. (2) C-4–C-3–O-3–HO-3. (3) C-5–C-4–O-4–HO-4 and (4) C-5–C-6–O-6–HO-6.
[d] Relative energy in kcal/mol calculated with respect to the global minimum.
[e] Population of conformer in per cent (298 K).

the GG orientation (or G orientation according to Fig. 2 nomenclature) about χ in the **A1** well and the TG (or T) orientation about χ in the **A2** well.

Curiously, the global minimum found in the grid search was not even a true conformer but a part of a catchment region. This can be explained by the combination of the grid increment and the relative easy convergence criterion used in the grid search. If

Table 4
Comparison of the global minima of ethyl β-lactoside found by respectively the grid search and by the CICADA search

	Φ	Ψ	χ	Φ'	Ψ'	χ'	ΔE
Crystal [47]	289	228	50	—	—	72	—
MM3/grid	270	200	69	281	179	66	0.26
MM3/cic	286	233	69	281	179	67	0.00

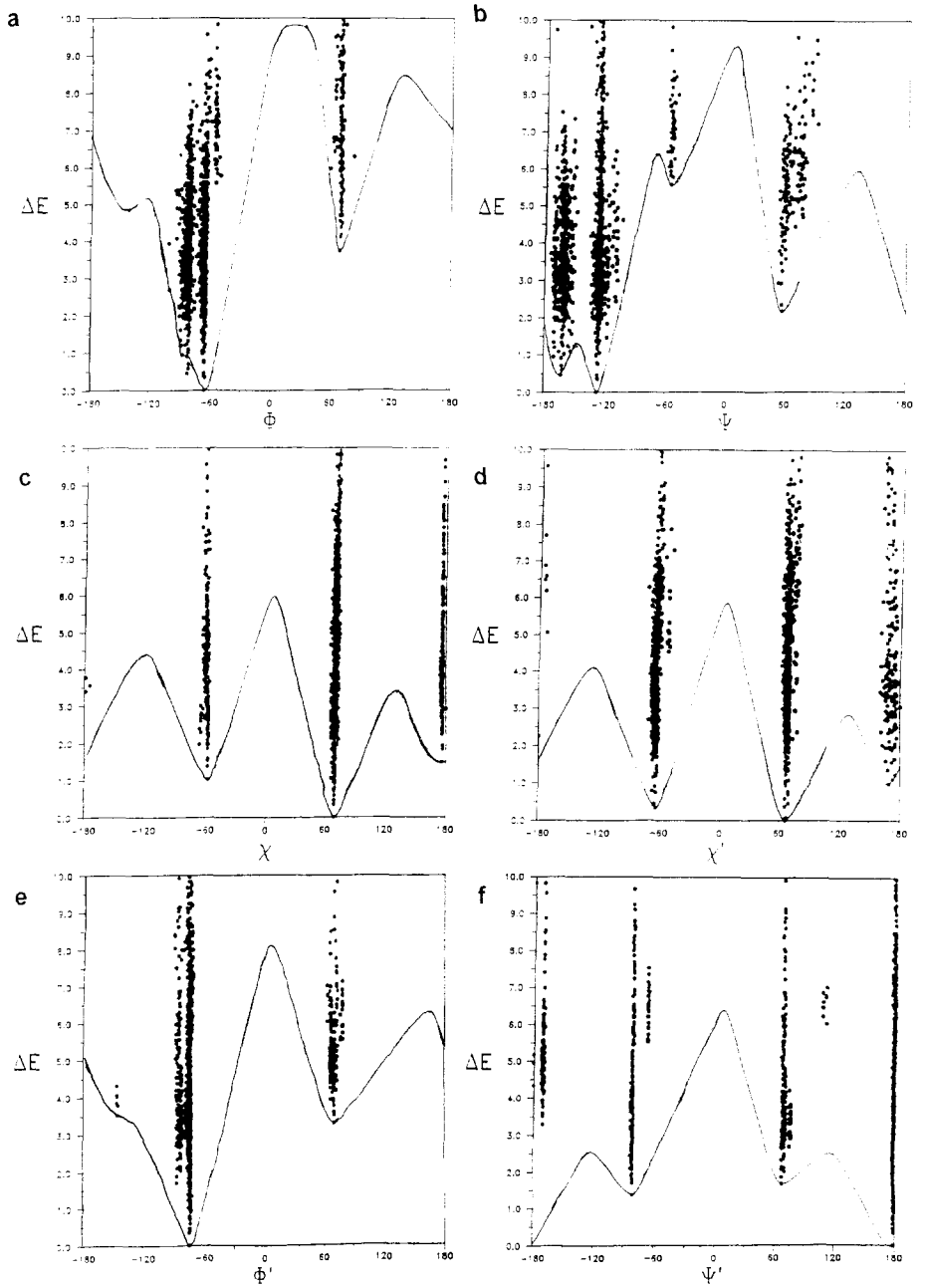


Fig. 6. One-dimensional plots of important dihedral angles versus energy of conformations (●) of ethyl β-lactoside found by the CICADA search. An underlying spline is drawn using information from only the lowest energy conformations and transition states: (a) Φ, (b) Ψ, (c) χ, (d) χ', (e) Φ' and (f) Ψ'.

the true minimum lies between the grid points, then obviously there is no chance for the grid search to find it. However, the conformational (Φ,Ψ)-space is not likely to contain such fine details and the CICADA minimum is only 1 and 3 degrees, respectively, from a pure grid point. More important is the fact that CICADA has $2N$ ways to arrive at a given minimum, and not all are likely to have the same descent conditions. The necessity for the use of a very robust minimization criteria is therefore less stringent for the CICADA search.

In Fig. 5b the underlying adiabatic isoenergy contours from the grid search have been drawn out to include 16 kcal/mol in order to show that the transition states are usually placed in saddle-point regions (representing transitions in Φ,Ψ-space) or, alternatively, in flat regions including minimum regions (representing transitions in the "depth"). As well as determining interconversion pathways amongst conformations, the transition-state information can provide useful information about the N-dimensional space. The very dense population of transition states in energy well A is characteristic for a large and flexible potential energy well. In fact, it is a qualitative measure of the N-dimensional volume or the true partition function. In contrast, the energy well **C** is free of transition states, indicating a very small N-dimensional volume. The complete lack of transition states indicates that this well is populated only by one particular conformation, or that energy barriers amongst conformers populating this well are high. The very low N-dimensional volume of the **C** well may very well be the reason why it is not significant for calculating spectral properties. The barrier between the **A** well and the **B** well is less than the barrier between the **A** and the **C** well, and the presence of some transition states, indicating some N-dimensional volume, makes the **B** well structure the most likely (only) alternative to the **A** well structures.

Fig. 6a and 6b, respectively, represent plots of the relative energy versus the Φ-angle and the Ψ-angle of conformations found by the CICADA search. Similar plots could be made of the transition states. However, because of their relatively unprecise determination in the driven directions (vide supra), we prefer to plot a lowest energy line using only information from the lowest energy transition states and the lowest energy conformations. The resulting line is a one-dimensional "adiabatic type" rotational barrier. In Fig. 6a we see already a clear split in the **A** well conformations corresponding to **A1** and **A2**. The 1D plot obviously contains overlapping "peaks" from different 2D regions; **A1**, **A2**, **B** and (-60; -60) wells all have a narrow range of Φ, and we therefore only "see" the most populated. It is, however, possible to distinguish a third peak corresponding to the (-60; -60) region at about 5.5 kcal/mol. The fourth and very narrow peak at about 60° corresponds to the **C** well. A much better resolution exists in the Ψ-direction (Fig. 6b) where **A1**, **A2**, **B** and (-60; -60) are completely separated in four well-defined peaks and corresponding number transition states. Only the **C** well is hidden (behind the **A2** well). From these one-dimensional energy plots it is apparent that the broad shape and low energy of the combined **A** well will accommodate almost all the populations. It is also apparent that conformational flexibility is much greater in the Ψ-direction than in the Φ direction, in contrast to Hayes et al.'s [58] interpretation of vicinal heteronuclear coupling constant data.

The primary hydroxyl groups. Fig. 6c and 6d represent similar one-dimensional plots of the χ and χ' dihedral of the primary hydroxyl groups. Both dihedral angles have a

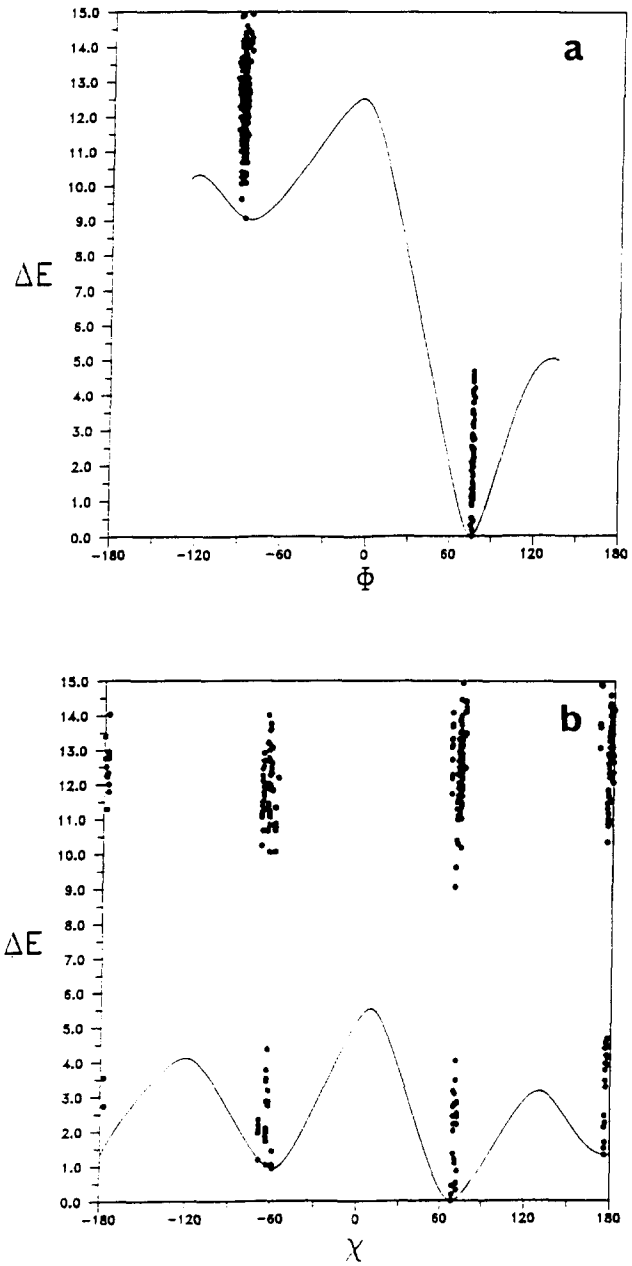


Fig. 7. One-dimensional plots of (a) Φ and (b) χ versus energy for conformations (●) of methyl α-D-galactoside found by the CICADA search. An underlying spline is drawn using information from only the lowest energy conformations and transition states.

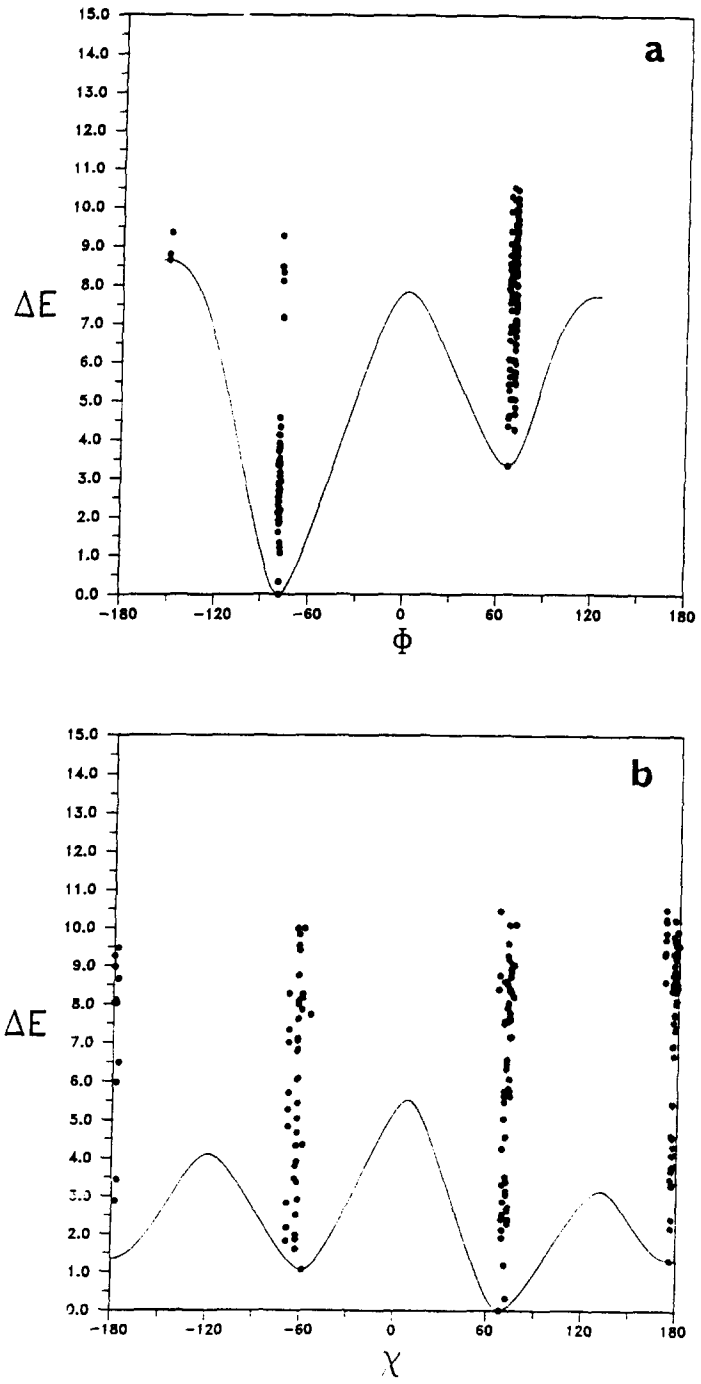


Fig. 8. One-dimensional plots of (a) Φ and (b) χ versus energy for conformations (●) of methyl β-D-galactoside found by the CICADA search. An underlying spline is drawn using information from only the lowest energy conformations and transition states.

relative simple three-fold symmetry with minima in the staggered positions, and both have their global minimum in the GT conformation. While their respective rotational barriers do have approximately the same energies, the primary hydroxyl group on the *galactose* residue has lower energy minima for the GG and TG conformations, resulting in higher populations of the GG and TG conformations.

The methoxy group. One-dimensional plots of the Φ'- and Ψ'-angles of the β-substituted ethoxy group (which can be regarded as a non-perturbed model of the glycosidic linkage) are shown in Fig. 6e and 6f. For the Φ'-angle, as in the case of the Φ-angle, we notice that the staggered T conformation is very unfavorable. No adiabatic energy barrier was found between this conformation and the strongly favoured exoanomeric G conformation, although transition states do exist for individual conformations. Again we see the G peak split into two parts, but in this case it is caused by a slightly different location in Φ of the T and **g** peaks in Ψ (Fig. 6f). A more detailed analysis would demand a 2D analysis of the peaks. The rotational barrier of the g peak is significantly lowered from that of the glycosidic Φ-angle, and the g minimum is also lower in energy. The plot of the Ψ'-angle, which is relaxed from anomer effects and free of steric conflicts, displays a three-fold symmetry with the T minimum being the global one surrounded by two almost isoenergetic *gauche* conformations. All the conformations are centered about almost ideal staggered values, and all the transition states are centered

Table 5
List of conformations of methyl α-D-galactoside found by the CICADA search. Conformations contributing more than one per cent to the total population are listed in descending order

Φ_i	χ_i	Gal p [a]	ΔE [b]	p_i [c]
76	68	GGGG	0.0000	17.883
76	68	gGGG	0.1744	13.323
75	72	GGGT	0.3137	10.532
77	68	TTGG	0.4320	8.626
76	72	gGGT	0.5067	7.604
76	72	TTGT	0.8613	4.179
76	−59	GGGG	0.9293	3.726
77	−64	TgTg	1.0307	3.140
77	−59	gGGG	1.0492	3.044
76	70	GGGg	1.0927	2.828
77	−69	TgTT	1.1774	2.451
76	70	gGGg	1.2045	2.342
76	176	GGGT	1.3188	1.931
77	69	TgEG	1.3608	1.799
77	−59	TTGG	1.4429	1.566
77	176	gGGT	1.5203	1.374
77	176	TTGT	1.6849	1.041

[a] Exocyclic hydroxyl dihedrals on methyl α-D-galactoside. The dihedrals are defined as follows: (1) C-1–C-2–O-2–HO-2, (2) C-2–C-3–O-3–HO-3, (3) C-3–C-4–O-4–HO-4 and (4) C-5–C-6–O-6–HO-6. Letter code is written according to Fig. 2.
[b] Relative energy in kcal/mol calculated with respect to the global minimum.
[c] Population of conformer in per cent (298 K).

Table 6
List of conformations of methyl β-D-galactoside found by the CICADA search. Conformations contributing more than one per cent to the total population are listed in descending order

Φ_i	χ_i	*Galp* [a]	ΔE [b]	p_i [c]
−78	68	gGGG	0.0000	38.665
−78	72	gGGT	0.3190	22.568
−78	−59	gGGT	1.0624	6.436
−78	71	gGGg	1.1970	5.128
−78	176	gGGt	1.3206	4.162
−79	−63	TgTg	1.6074	2.565
−79	−69	TgTT	1.8250	1.777
−78	−63	gGTg	1.8750	1.633
−79	69	TgEG	1.9179	1.519
−79	−63	ggTg	1.9840	1.359
−80	69	TppG	2.1176	1.084
−78	69	gGgG	2.1332	1.056

[a] Exocyclic hydroxyl dihedrals on methyl β-D-galactoside. The dihedrals are defined as follows: (1) C-1–C-2–O-2–HO-2, (2) C-2–C-3–O-3–HO-3, (3) C-3–C-4–O-4–HO-4 and (4) C-5–C-6–O-6–HO-6. Letter code is written according to Fig. 2.
[b] Relative energy in kcal/mol calculated with respect to the global minimum.
[c] Population of conformer in per cent (298 K).

about eclipsed values. The G and the g peaks, however, are split into two peaks, indicating two different 2D positions.

Methyl α-D-galactopyranoside and methyl β-D-galactopyranoside. The methoxy group.—An exhaustive CICADA search on methyl α-D-galactopyranoside and methyl β-D-galactopyranoside was carried out. One-dimensional plots of the two most important dihedral angles, Φ and χ, are shown in Fig. 7 and Fig. 8. It is interesting to note that the classical partition function of the α anomer is about two times larger than that of the β anomer. The number of conformations populated by more than one per cent is also considerably smaller for the β anomer (see Tables 5 and 6), where the two most populated conformers account for more than 60 per cent of the population. The CICADA plot of the α-exoanomeric methoxy group, shown in Fig. 7a, characteristically does not have any conformations in the T orientation, but rather shows two surrounding transition states corresponding to a shoulder region. An almost similar situation exists for the β anomer, shown in Fig. 8a, although it does have a few conformations around $\Phi = -138$, but with no well-defined transition states around them. The energy difference between the g ($\Phi = 60°$) conformation for Me α-D-Gal*p* and the G ($\Phi = -60°$) conformation for Me β-D-Gal*p* is calculated to be 1.09 kcal/mol in favor of the β conformer, a difference almost entirely caused by the bending potential. This result is qualitatively different from the results of a recent ab initio and molecular mechanics study on 2-methoxytetrahydropyrans by Tvaroska and Carver [43]. They found a 0.94 kcal/mol difference in favor of the g α conformer using the 6-311 + + G** basis set, which is in qualitative agreement with that of a difference of 0.47 kcal/mol in MM3 ($\epsilon = 4$). In MM3 the geometry of the two global minima conformers, g (Me α-D-Gal*p*) and G (Me β-D-Gal*p*), are practically identical, except for the orientation of the

methoxy group and the C-1–C-2–O-2–HO-2 dihedral angle, which follows the direction of the O-1 bond, thus maintaining an identical hydrogen-bonding scheme. The preference of the G β conformer over the g α conformer in the methyl galatosides can therefore only be attributed to the more crowded pyranoside ring as compared to the tetrahydropyranoside ring. The overall shape of the underlying spline curves (Figs 7a and 8a), including energy differences between rotamers, agrees well with the results on 2-methoxytetrahydropyrans by Tvaroska and Carver [43]. Only the barrier heights have a tendency to be slightly enlarged in agreement with the more strained pyranoside ring.

The primary hydroxyl groups. The split in energies of the three staggered positions of the primary hydroxyl group of the α anomer (Fig. 7b), representing two different groups of families with almost identical energy differences, is quite unusual. The two families correspond respectively to the G and g families in the Φ-direction. The fact that the two families have the same energy differences and barriers (not shown) points to the largely independent nature of the two major rotatable bonds in methyl α-D-galactopyranoside; however the higher family has different sets of secondary hydroxyl positions for the corresponding minima structures. The difference in adiabatic energy between the GT and the GG conformers is 0.9 kcal/mol with identical secondary hydroxyl positions GGGG (see Table 5). The difference in adiabatic energy between the GT and the TG conformations is 1.3 kcal/mol with different secondary hydroxyl positions (see Table 5). The TG GGGG conformer is found to be as high as 2.1 kcal/mol above the preferred conformer. A similar split in energies for the three staggered positions of the primary hydroxyl group is not observed for the β anomer because of a smaller energy difference between the conformations in the Φ-direction. The adiabatic energy difference between the GT and the GG conformers is 1.1 kcal/mol with almost identical orientations of the secondary hydroxyl groups, but different orientation of the primary hydroxyl group C-5–C-6–O-6–HO-6 (see Table 6). The adiabatic energy difference between the GT and the TG conformations is 1.3 kcal/mol, with identical secondary hydroxyl positions (see Table 6), but with a third orientation of the primary hydroxyl group C-5–C-6–O-6–HO-6. The fact that the adiabatic energy differences and barrier heights between primary hydroxyl group rotamers (Figs 7b and 8b) are practically identical for the α and β anomers indicates that the primary hydroxyl group position is not related to the position of the exo-anomeric methoxy group, although the configurational entropy is larger for the α anomer. The ensemble calculated population distributions are respectively (GG:GT:TG) = 21:73:6 for the α anomer and (GG:GT:TG) = 17:76:7 for the β anomer (see Table 2).

Optical rotation.—Ethyl β-lactoside. In the previous study [23] the $[\alpha]_D^{18}$ of ethyl β-lactoside in water was measured to be very close to zero ([M] = -3 ± 13 deg cm^2 dmol^{-1}) in contrast to that of β-lactose, which has been measured to be around 120 deg cm^2 dmol^{-1} [59]. The influence of the ethoxy group on the optical properties of the molecule is apparently very strong. The geometry-dependent semiempirical approach of Stevens [34] has previously been used to probe the structure of both β-lactose using statistical weights from NMR studies [60] and ethyl β-lactoside using ensemble-averaged molar rotation from a grid search [23].

The molar rotation of the global minimum from the grid search was calculated as 48 deg cm^2 dmol^{-1} corresponding to a **A2** minimum, whereas the ensemble averaged

molar rotation was calculated to be 23 deg cm^2 $dmol^{-1}$. The grid-search investigation was limited because the influential ethoxy group was relaxed only in the G-T conformation and not driven in the grid searches. The error introduced by this limitation was considered to be small since the G-T conformation of the ethoxy group of the global minimum was calculated to be about 1.5 kcal/mol lower than any other combination, and also since the numerical values of the molar rotation were similar, except for that of the G-G conformation, which was calculated to be significantly lower. Since the G-G conformation was second lowest in energy, the calculated Boltzmann-averaged values were then considered as upper limits.

As mentioned previously, the CICADA search algorithm does not suffer from exponential growth in calculation times when a new dihedral angle is included in the search, and therefore the ethoxy group was driven in the search. It was hoped that calculation of optical rotation from the CICADA search would establish the magnitude of the error. The molar rotation of the global minimum found by the CICADA search was calculated to be 120 deg cm^2 $dmol^{-1}$ corresponding to that of the **A1** well and about 80 deg cm^2 $dmol^{-1}$ higher than that of the grid search equivalent in the **A2** well. The ensemble-averaged molar rotation from the CICADA search was calculated to be 40.5 deg cm^2 $dmol^{-1}$ for the set of conformations alone and 49.5 deg cm^2 $dmol^{-1}$ for the set of conformations including the transition states. Both values are higher than the corresponding value obtained from the grid search due to insufficient grid search sampling of the **A1** well. The effect of rotating the ethoxy group cannot, therefore, be established but, as is apparent from Table 3, all conformations populated by more than one per cent are G-T conformations. In fact, the most populated conformer that did not have the ethoxy group in the G-T orientation was populated only 0.353 per cent in a G-g orientation.

Methyl α-D-galactopyranoside and methyl β-D-galactopyranoside.—The ensemble-averaged molar rotation for methyl α-D-galactoside was calculated to be 358 deg cm^2 $dmol^{-1}$, which is in relatively good agreement with the experimental value of 381 deg cm^2 $dmol^{-1}$ [59]. All molar rotations of conformers were found to be positive with a level around 180 deg cm^2 $dmol^{-1}$ for the G minima of the methoxy group and a level around 360 deg cm^2 $dmol^{-1}$ for the g minima.

For the methyl β-D-galactoside the ensemble-averaged molar rotation was calculated to be 38 deg cm^2 $dmol^{-1}$ compared with the experimental value of 0 deg cm^2 $dmol^{-1}$ [59]. The calculated value is dominated by the molar rotation of the two most populated conformers having a calculated molar rotation of respectively 57 and 58 deg cm^2 $dmol^{-1}$, which is representative of the g conformers with the primary hydroxyl group in the GT conformation. Calculated molar rotations around −40 and close to 0 deg cm^2 $dmol^{-1}$ are representative of the G-GG conformers and the G-TG conformers, respectively. On the basis of these results, it can be concluded that the populations of the GG and GT conformers about the primary hydroxyl groups are of approximately the same magnitude in the β anomer, with the GG being the most populated. However, this conclusion is in conflict with the NMR coupling constant data.

Coupling constants.—Ethyl β-lactoside. Calculated coupling constants are listed in Table 1, together with the experimental values. It is readily apparent from the results obtained with the CICADA search that only two significant differences with respect to

the grid search are observed. One of them concerns the rotamer distribution of the primary hydroxyl group of *galactose* where the GT population becomes even more predominant at the expense of the GG population, giving rise to a significantly better reproduction of the H-5–H-6*R* coupling constant. When transition states are included in the calculation, this trend is emphasized even more, but the fit for the coupling constants becomes poorer. The population of the primary hydroxyl group of *glucose* in the CICADA search is calculated to be similar to that of the grid search. Inclusion of transition states again seems to lead to poorer values. The second aspect concerns the calculated H-4′–C-1 coupling constant, which increases by a factor of 0.4 Hz with respect to the grid search, reflecting the shift imposed by the more correctly calculated **A1**:**A2** ratio. The value of 4.6 Hz is in good agreement with the value 4.9 Hz measured by Hayes et al. [58] for methyl β-lactoside. It was expected that the H-1–CEt coupling would increase because of the CICADA driving, but this value remained unchanged at 3.2 Hz compared to the experimental value of 4.4 Hz. Assuming that the parametrization is correct for this type of heteronuclear coupling, it could reflect too strong on exoanomeric parametrization in MM3.

*Methyl α-*D*-galactopyranoside and methyl β-*D*-galactopyranoside.* The calculated coupling constants for methyl α-D-galactopyranoside and methyl β-D-galactopyranoside from the CICADA ensembles are listed in Table 2 along with the experimental values [38,61,62]. While the Haasnoot–Altona equation is able to represent the homonuclear axial–axial ring couplings and the axial–equatorial situation about C-1, the equatorial H-4 coupling with the axial H-3 is poorly reproduced for both the α and β galactopyranosides. For methyl α-D-galactopyranoside we have two different sets of experimental data for the pro-*R*/pro-*S* couplings; Nishida et al. [61] measured the two coupling constants in 1984 to be 7.8 Hz/6.0 Hz, whereas Bock and Duus [38] 10 years later measured the same coupling to be 8.2 Hz/3.7 Hz. Our results from the molecular statics ensembles are in much better accordance with the latter values, resulting in about 70% (see Table 2) of the population being GT conformers (GG:GT:TG = 21:73:6). On basis of their experiment, Nishida et al. [61] suggested a much higher TG population, mainly at the expense of the GT population (GG:GT:TG = 14:47:39).

From the results in Table 1 and 2, we are forced to conclude that the combined in vacuo MM3/Haasnoot–Altona investigation of the H-5–H-6 couplings is not generally able to satisfactorily reproduce both coupling constants (within < 1 Hz). However, the rotamer distributions for the primary hydroxyl group of the β-D-galactose moiety of ethyl β-lactoside are in good agreement with experiment. Although it is possible that the MM3 force field needs minor adjustments for the C-5–C-6 bonds in carbohydrates, it is believed that a significant improvement to the model will need the specific inclusion of solvation effects.

NOESY volumes.—Ethyl β-lactoside. Evidence [23] for internal motion in ethyl β-lactoside has been unambiguously obtained by comparing NOESY volumes calculated for a rigid molecule with the experimental ones. Both strongly enhanced diagonal volumes (H-6′*S* + HA of the ethyl methylene group) and greatly reduced cross-peak (H-3/H-4 of β-Gal*p*) and diagonal volumes (H-1 of β-Gal*p*) are observed for quasi first-order spin systems. Five-fold reduction in the least-squares (χ^2) fit with respect to experimental data is obtained by introducing parameters that take into account internal

Table 7
Experimental (boldface type) (standard and average deviations for a_{ij} and a_{ii} volumes respectively), and theoretical [a] (fast motion models) 400.13 MHz normalized NOESY volumes for spectra (1 s mixing time) of ethyl β-lactoside

		Expt. [e]	MM3/grid	MM3/cic conf
a_{ii}	a_{11}	**0.219 (3)**	0.236	0.250
	a_{22}	**0.674 (9)**	0.774	0.775
	a_{44}	**0.369 (9)**	0.384	0.390
	$a_{1'1'}$	**0.309 (8)**	0.324	0.319
	a_{s1s1} [b]	**0.891 (-)**	0.835	0.857
	a_{s3s3} [c]	**0.560 (4)**	0.536	0.537
a_{ij}	a_{1s1}	**−0.026 (1)**	−0.026	−0.026
	a_{1s2} [d]	**−0.048 (2)**	−0.051	−0.052
	a_{4s1}	**−0.035 (1)**	−0.037	−0.038
	a_{4s2}	**−0.034 (1)**	−0.035	−0.035
	$a_{1's2}$	**−0.047 (2)**	−0.052	−0.053
	$a_{1's3}$	**−0.007 (1)**	−0.009	−0.010

[a] The theoretical values are calculated for five $S^2(k,l)$ values (0.8 for interactions between ring protons and methylene protons, 0.6 for interactions between carbohydrate protons and ethyl group protons, 0.4 for methylene group, 0.2 for ethyl group, and 0.9 for remaining interactions), four τ_i values (0.1 ns for the intracyclic interactions, 0.005 ns for the ethyl group, 0.05 ns for interactions between ring protons and exocyclic ones, 0.1 ns for the methylene group and for the interring interactions), and a τ_c value of 0.24 ns.
[b] s1 contains the spins H-5, H-6r, H-6s, H-6'r and H-B.
[c] s2 contains the spins H-3, H-3', H-4' and H-5'.
[d] s3 contains the spins H-6's and H-A.
[e] From ref. [23].

dynamics. Simulations with model-free spectral densities [40] and appropriate motional parameters lead to good agreement between theoretical and experimental data. This approach has been adopted for all the NOESY volume simulation reported in the present study.

In the case of ethyl β-lactoside, a unique solution did not emerge through systematic evaluation of motional models. The only internal dynamics which were relatively well-defined were those corresponding to the intracyclic interactions ($S^2_{ang}(k,l) = 0.7$ and $\tau(k,l) > 0.1$ ns) and to ones involving the ethyl group ($S^2_{ang}(k,l) = 0.3$ and $\tau(k,l) < 0.04$ ns). Manual fitting to the most unambiguous data (first-order spin systems) gave a very good fit for all three theoretical ensembles with the following motional model $S^2_{ang}(k,l)$ values of 0.8, 0.6, 0.4, 0.2, 0.9 and 0.9, and $\tau(k,l)$ ones of 0.05, 0.05, 0.1, 0.05, 0.1 and 0.5 ns for interactions (a) between ring protons and methylene ones, (b) between ring protons and the ethyl group, (c) within the methylene groups, (d) within the ethyl group, (e) within the pyranosyl rings, and (f) between the rings. For the intracyclic correlation time, we arrived at two models which can equally well represent the experimental data, one in which the dominant intracyclic relaxation is fast 0.001 ns, which corresponds to the value which we have previously obtained [23], and one obtained by a systematic approach in which the dominant intracyclic relaxation is a slow 0.1 ns.

The corresponding simulated data for the three ensembles are collected in Table 7. The theoretical NOESY volumes are almost identical, indicating similar average struc-

Table 8
Experimental (standard deviation) [a] and theoretical [bc] 400.13 MHz normalized NOESY volumes for spectra (1 s mixing time) of methyl β-D-galactoside

	H-1	H-2	H-3	H-4	H-5	H-6*S*, 6*R*	Me
H-1	**0.577 (15)**	**−0.012 (2)**	**−0.033 (2)**		**−0.055 (3)**		**−0.016 (2)**
	0.591	−0.014	−0.031		−0.051		−0.018
H-2		**0.849 (27)**		**−0.008 (3)**	**−0.008 (4)**		**−0.012 (6)**
		0.876	−0.015	−0.004	−0.002		—
H-3			**0.601 (7)**	**−0.049 (2)**	**−0.027 (8)**		
			0.583	−0.050	−0.030		
H-4				**0.604 (8)**	**−0.040 (3)**	**−0.016 (3)**	
				0.615	−0.043	−0.012	
H-5					**0.478 (6)**	**0.008 (2)**	
					0.510	−0.016	
H-6*R*, 6*S*						**0.590** [d] **(5)**	
						0.481	
Me							**0.74** [d] **(1)**
							0.440

[a] Experimental values in boldface and standard deviations in parentheses.
[b] Theoretical values for the cicada conformations were calculated with model-free [40] spectral densities. The motional model that was obtained by least-squares fitting of theoretical data to experimental ones is as follows: S^2_{ang}(k,l) values of 0.7, 0.3, 0.1, 0.3, 0.6 and 0.3 for intracyclic, ring-exocyclic, ring-methyl, exocyclic, exocyclic-methyl and methyl interactions, respectively. The corresponding correlation times for internal motions were 0.2, 0.02, 1, 0.03 and 0.03 ns, while the overall tumbling time was 0.07 ns.
[c] The theoretical NOESY volumes for cicada conformations or conformations and transitions states were identical.
[d] These experimental intensities were arbitrarily reduced for the fitting procedure in order to compensate for strong coupling effects.

tures for the three search procedures. The only notable exceptions are the anomeric proton diagonal volumes, a_{11} and to a lesser extend $a_{1'1'}$. The variation in a_{11} corresponds to a slightly different orientation about the glycosidic linkage. Regarding the notable variation in $a_{s1,s1}$, it should be kept in mind that this volume corresponds to five overlapping signals.

Methyl β-D-galactoside. A complete set of experimental NOESY volumes (boldface) for methyl β-D-galactoside are collected in Table 8, along with data simulated for the CICADA ensemble of conformers. Several of the cross-peak volumes were higher than the corresponding rigid-molecule theoretical values (unpublished results) corroborating the presence of internal dynamics for the pyranosyl ring. Several thousand motional models were evaluated, but a unique representation of the internal dynamics was obtained by least-squares fitting of the experimental data. In these calculations the experimental intensities of both the methyl and methylene protons were arbitrarily [29] reduced in order to compensate for strong coupling [63], and the overall tumbling was set at 70 ps. In the case of the best motional model, S^2_{ang}(k,l) values of 0.7, 0.3, 0.1, 0.3, 0.6 and 0.3 were optimal for the (a) intracyclic, (b) ring/exocyclic, (c) ring/methyl, (d) exocyclic, (e) exocyclic/methyl and (f) methyl interactions, respectively, while the corresponding internal motion correlation times were 0.2, 0.02, 1, 0.03 and 0.03 ns.

Due to the arbitrary scaling of the methyl and methylene intensities, the rates and amplitudes of motions involving these groups are poorly defined. By coincidence, the correlation time of 0.03 ns for internal rotation about the hydroxymethyl group is almost identical to the value of 0.035 ns estimated for the galactosyl hydroxymethyl group in lactoside based on multifield carbon relaxation data [44]. The S^2_{ang}(k,l) value of 0.7 for the galactosyl intraring interactions, which is well defined (8 cross-peak volumes), is in good agreement with the theoretical value of 0.8 established from molecular dynamics simulations of ethyl β lactoside [23]. In this latter study very similar values had been observed for all the theoretical S^2_{ang}(k,l) values of the intracyclic interactions, which validates the use of a group angular order parameter and correlation times in the least-squares fitting of NOESY volumes. The S^2_{ang}(k,l) values are analogous to those reported [23,27] for interactions that are modulated by both fixed (0.6 to 1.0) and flexible (0.2 to 0.4) internuclear distances. Strictly-speaking the maximum theoretical value of τ(k,l) should not exceed $0.5\tau_c$ but larger values only lead to small ($< 5\%$) variations in the theoretical volumes.

In conclusion, the homonuclear relaxation data calculated with the CICADA ensemble can be fitted to reproduce the corresponding experimental data. In the case of ethyl β-lactoside, the fit is comparable to those that obtained with the grid-search approach.

4. Conclusions

It has been demonstrated that the CICADA algorithm for conformational searching provides spectral information of a quality equal to that of a very elaborate grid search. Not only are all low-energy regions of the potential energy surface found, but information about transition states and interconversion pathways are provided from which quasi-dynamical properties of the molecule can be extracted. The conformations found are free of harmonic constraints, which makes the CICADA algorithm ideal for calculating spectral data. In the grid-search approach all grid points are constrained in one or more directions, which influences the second-order derivatives and, in turn, the normal modes of the particular conformation. Thus a true spectral calculation will not be possible from the grid-search ensemble without subsequently relaxing all grid points. The search method will then lose its only advantage, namely its systematic exploration of PES, since it will be necessary to apply special algorithms in order to avoid redundancy.

Even on very complex potential energy surfaces, the CICADA algorithm finds all potential energy wells, but it spends by far the most of its time in highly populated areas of the PES and, therefore, provides a complete description of these essential regions. The algorithm's "from the bottom" philosophy makes it very efficient in spectral sampling, since highly populated areas on complex molecular PES do not grow exponentially with the dimensionality of the system. Thus this approach holds promise for application to larger and more flexible biomolecular systems.

For the relatively small molecule, ethyl β-lactoside, investigated in present study, the CPU consumption of the CICADA search was approximately 6 days on a IBM RISC6000/520 calculating about 20 000 points on PES, of which about 25% were

saved as vertices in the graph. The grid search, to which it was compared, included calculation of 46 656 grid points and consumed about 14 CPU days. Given spectral sampling of equal quality and better approximations to both global and local minima, the gain in efficiency is enormous. It is obvious that the surplus of the graph calculations is insignificant with respect to the total CPU requirements, which is easily compensated by the better convergence properties of the algorithm (only one dihedral is moved at a time; the rest of the molecule is in a minimum configuration).

An important spinoff from the CICADA algorithm is the provision of interconversion pathways by virtue of the transition states. These can be used to study folding phenomena in polymers and as has been done in the present work and they can be used to study barriers amongst rotamers. It was also expected that the transition states would contribute significantly to the spectral information since these points in hyperspace represent stationary points often in low-energy areas of the PES and contribute significantly to the classical partition function. A significant contribution from the transition states was not, however, observed in the present work, perhaps because of their imprecise determination (see algorithm description) and perhaps because of their intermediate position. It is a curious fact that the transition states are sufficiently precisely determined that one of their eigenvalues is negative so that an exact determination is possible using a second-order Newton–Raphson minimizer. However, such an exact determination of the transition states would decrease the efficiency of the CICADA algorithm in finding low energy domains.

Acknowledgements

This work was supported by a grant to S.B.E. in the framework of the program Conception MacroMolécules Assistée par Ordinateur (Organibio) and a fellowship to I.B. from the French Ministère de la Recherche et de l'Enseignement Supérieur. The provision of financial support by P. et M. Curie University, CNRS (URA 1679), and INRA are acknowledged. The authors would like to acknowledge Dr William Mackie (Leeds University, UK) and Dr Anne Imberty (CNRS Nantes, France) for helpful comments on the manuscript.

References

[1] I. Mills, in R. Fausto (Ed.), Recent Experimental and Computational Advances in Molecular Spectroscopy *NATO ASI Series C*, 406 (1992), 79–98. The citation is from the talk given at the NATO meeting held in Ponta Delgada, Sao Miguel (Acores), Portugal.
[2] H.A. Scheraga, *UCLA Symp.*, (1989) 213–221.
[3] V. Villani and A.M. Tamburro, *J. Mol. Struct. (Theochem)*, 308 (1994) 141–157.
[4] S. Kirkpatrick, Jr, C.D Gellat, and M.P. Vecchi, *Science*, 220 (1983) 671–680.
[5] M. Saunders, K.N. Houk, Y.-D. Wu, W.C. Still, M. Lipton, G. Chang, and W.C. Guida, *J. Am. Chem. Soc.*, 112 (1990) 1419–1427.
[6] H. Goto, E. Osawa, and M. Yamato, *Tetrahedron*, 49 (1993) 387–396.
[7] R.E. Bruccoli and M. Karplus, *Biopolymers*, 26 (1987) 137–168.
[8] M. Lipton and W.C. Still, *J. Comput. Chem.*, 9 (1988) 343–355.

[9] F. Willamagna and M.A. Whitehead, *J. Chem. Soc., Faraday Trans.*, 90 (1994) 47–54.
[10] M.H. Lambert and H.A. Sheraga, *J. Comput. Chem.*, 10 (1989) 798–816.
[11] R.E. Bruccoli and M. Karplus, *Biopolymers*, 29 (1990) 1847–1862.
[12] M. Saunders, *J. Am. Chem. Soc.*, 109 (1987) 3150–3152.
[13] D.M. Ferguson and D.J. Raber, *J. Am. Chem. Soc.*, 111 (1989) 4371–4378.
[14] Z. Li and H.A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, 84 (1987) 6611–6615.
[15] G. Chang, W.C. Guida, and W.C. Still, *J. Am. Chem. Soc.*, 111 (1989) 4379–4386.
[16] E.O. Purisma and H.A. Scheraga, *Proc. Natl. Acad. Sci.*, 83 (1986) 2782–2786.
[17] W. Linert, P. Margl, and I. Lukovitz, *Computers Chem.*, 16 (1992) 61–69.
[18] S.R. Wilson, W. Cui, J.W. Moscuwitz, and K.E. Schmidt, *J. Comput. Chem.*, 12 (1991) 342–349.
[19] J.H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, 1975.
[20] M.J.J. Blommers, C.B. Lucasius, G. Kateman, and R. Kaptein, *Biopolymers*, 32 (1992) 45–52.
[21] J. Koca, *J. Mol. Struct. (Theochem)*, 308 (1994) 13–24.
[22] A. Imberty, S. Pérez, and J. Koca, *J. Comput. Chem.*, (1994) in press.
[23] S.B. Engelsen, S. Pérez, I. Braccini, and C. Hervé du Penhoat, *J. Comput. Chem.*, (1995) in press.
[24] N.L. Allinger, Y.H. Yuh, and J.-H. Lii, *J. Am. Chem. Soc.*, 111 (1989) 8551–8567.
[25] N.L. Allinger, M. Rahman, and J.-H. Lii, *J. Am. Chem. Soc.*, 112 (1990) 8293–8307.
[26] D.M. Grant, C.L. Mayne, F. Liu, and T.-X. Xiang, *Chem. Rev.*, 91 (1991) 1591–1624.
[27] M. Hricovini, R.N. Shah, and J.P. Carver, *Biochemistry*, 31 (1992) 10018–10023.
[28] C. Meyer, S. Pérez, C. Hervé du Penhoat, and V. Michon, *J. Am. Chem. Soc.*, 115 (1992), 10300–10310.
[29] N. Bouchemal-Chibani, I. Braccini, C. Derouet, C. Hervé du Penhoat, and V. Michon, *Int. J. Biol. Macromol.*, (1995) in press.
[30] IUPAC-IUB, Commission on Biochemical Nomenclature, *Arch. Biochem. Biophys.*, 145 (1971) 405–421.
[31] S. Macura, Y. Huang, D. Suter, and R.R. Ernst, *J. Magn. Reson.*, 43 (1981) 259–281.
[32] D. Neuhaus and M. Williamson, in *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH Publishers, 1989, p. 292.
[33] A.D. French, R.J. Rowland, and N.L. Allinger, *ACS Symp. Ser.*, 430 (1990) 120–140.
[34] E.S. Stevens and B.K. Sathyanarayana, *Carbohydr. Res.*, 166 (1987) 181–193.
[35] M. Karplus, *J. Chem. Phys.*, 30 (1959) 11–15.
[36] C.A.G. Haasnoot, F.A.A.M. de Leeuw, and C. Altona, *Tetrahedron*, 36 (1980) 2783–2792.
[37] I. Tvaroska, M. Hricovini, and E. Petrakova, *Carbohydr. Res.*, 189 (1989) 359–362.
[38] K. Bock and J.Ø. Duus, *J. Carbohydr. Chem.*, 13 (4) (1994) 513–543.
[39] J.P. Carver, *Curr. Opin. Struct. Biol.*, 1 (1991) 716–720.
[40] G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 104 (1982) 4546–4559.
[41] E.T. Olejniczak, R.T. Gampe, Jr, and S.W. Fesik, *J. Magn. Reson.*, 67 (1986) 28–41.
[42] J.M. Berry, L.D. Hall, and K.F. Wong, *Carbohydr. Res.*, C16–C20 (1977).
[43] I. Tvaroska and J. Carver, *J. Phys. Chem.*, 94 (1994) 9477–9485.
[44] A. Ejchart and J. Dabrowski, *Magn. Reson. Chem.*, 30, (1992) 115–124.
[45] A. Ejchart, J. Dabrowski, and C.-W. von der Lieth, *Magn. Reson. Chem.*, 30, (1992) 105–114.
[46] A.J. Duben and W.C. Hutton, *J. Am. Chem. Soc.*, 112 (1990) 5917–5924.
[47] K. Hirotsu and A. Shimada, *Bull. Chem. Soc. Jpn.*, 47(8) (1974) 1872–1879.
[48] C.E. Bugg, *J. Am. Chem. Soc.*, 95 (1973) 908–913.
[49] W.J. Cook and C.E. Bugg, *Acta Crystallogr., Sect. B*, 29 (1973) 907–909.
[50] C.A. Beevers and H.N. Hansen, *Acta Crystallogr., Sect. B*, 27 (1971) 1323–1325.
[51] J.H. Noordik, P.T. Beurskens, P. Bennema, R.A. Visser, and R.O. Gould, *Z. Kristallogr.*, 168 (1984) 59–65.
[52] D.C. Fries, S.T. Rao, and M. Sundaralingam, *Acta Crystallogr.*, 27B (1971) 994–1005.
[53] B. Shaanan, H. Lis, and N. Sharon, *Science*, 254 (1991) 862–866.
[54] Y.D. Lobsanov, M.A. Gitt, H. Leffler, S.H. Barondes, and J.M. Rini, *J. Biol. Chem.*, 268 (1993) 27034–27038.
[55] E. Rutenber and J.D. Robertus, *Proteins: Structure, Function and Genetics*, 10 (1991) 260–269.
[56] T.K. Sixma, S.E. Pronk, K.H. Kalk, E.S. Wartna, B.A.M. van Zanten, B. Witholt, and W.G.J. Hol, *Nature*, 351 (1991), 371–377.

[57] T.K. Sixma, K.H. Kalk, B.A.M. van Zanten, Z. Dauter, J. Klingma, B. Witholt, and W.G.J. Hol. *J. Mol. Biol.*, 230 (1993) 890–918.
[58] M.L. Hayes, A.S. Serianni, and R. Barker, *Carbohydr. Res.*, 100 (1982) 87–101.
[59] F.J. Bates, *Polariometry, Saccharimetry and the Sugars*; NBS Circular C440, U.S. Government Printing Office, Washington, DC, 1942. Data on methyl α,β-D-galactoside p. 718 and data on β-lactose p. 738.
[60] C.A. Duda and E.S. Stevens, *Carbohydr. Res.*, 230 (1990) 347–351.
[61] Y. Nishida, H. Ohrui, and H. Meguro, *Tetrahedron Lett.*, 25(15) (1984) 1575–1578.
[62] K. Bock and H. Thøgersen, *Ann. Report NMR Spectrosc.*, 13 (1982) 2–49.
[63] J. Keeler, D. Neuhaus, and M.P. Williamson, *J. Magn. Reson.*, 73 (1987) 45–68.