Journal of Biomolecular NMR, **21:** 49–61, 2001.
KLUWER/ESCOM

Derivation of ^{13}C chemical shift surfaces for the anomeric carbons of oligosaccharides and glycopeptides using *ab initio* methodology

Chet W. Swalina, Randy J. Zauhar, Michael J. DeGrazia & Guillermo Moyna*
Department of Chemistry & Biochemistry, University of the Sciences in Philadelphia, 600 South 43rd Street, Philadelphia, PA 19104-4495, U.S.A.

Received 31 March 2001; Accepted 17 June 2001

Key words: ab initio, anomeric carbon, carbon chemical shift surface, GIAO, glycosidic bond, oligosaccaride conformation

Abstract

The dependence between the anomeric carbon chemical shift and the glycosidic bond $\langle\phi, \psi\rangle$ dihedral angles in oligosaccharide and glycopeptide model compounds was studied by Gauge-Including Atomic Orbital (GIAO) *ab initio* calculations. Complete chemical shift surfaces versus ϕ and ψ for D-Glcp-D-Glcp disaccharides with (1→1), (1→2), (1→3), and (1→4) linkages in both α- and β-configurations were computed using a 3-21G basis set, and scaled to reference results from calculations at the 6-311G** level of theory. Similar surfaces were obtained for GlcNAcThr and GlcNAcSer model glycopeptides in α- and β-configurations, using in this case different conformations for the peptide moiety. The results obtained for both families of model compounds are discussed. We also present the determination of empirical formulas of the form $^{13}C\delta = f(\phi,\psi)$ obtained by fitting the raw *ab initio* data to trigonometric series expansions suitable for use in molecular mechanics and dynamics simulations. Our investigations are consistent with experimental observations and earlier calculations performed on smaller glycosidic bond models, and show the applicability of chemical shift surfaces in the study of the conformational behavior of oligosaccharides and glycopeptides.

Introduction

Nuclear magnetic resonance (NMR) is perhaps the most powerful technique available for the study of macromolecular dynamics and three dimensional (3D) structure in solution. Although the general methods have been extensively developed and are now well established (Wüthrich, 1986; Evans, 1995), progress continues in the development of new experimental techniques and more accurate empirical relations between NMR observables and structural parameters (Clore and Gronenborn, 1997). Most of the advances have been due to the increase in magnetic fields, up to 18.8 T in present commercial magnets, and the production of hardware necessary to implement elaborate pulse sequences that permit more structure-related information to be extracted from the molecular spin systems. These experiments are best exemplified by the recent reports on the determination of residual dipolar couplings in partially aligned proteins (Tjandra and Bax, 1997), and on the identification of hydrogen bonding networks in proteins through $^{h}J_{NC'}$ scalar coupling measurements (Cornilescu et al., 1999). In the field of molecular modeling, the availability of inexpensive and powerful computer hardware has made possible accurate descriptions of biomolecular systems using empirical molecular mechanics force fields based on high level *ab initio* calculations, such as the Cornell and Merck force fields (Cornell et al., 1995; Halgren, 1996). When coupled to data derived from NMR measurements, molecular modeling simulations can now produce structural models that in many cases rival in accuracy those obtained by X-ray crystallography. Furthermore, models derived from NMR data provide information on the dynamics of the molecular

*To whom correspondence should be addressed. E-mail: g.moyna@usip.edu

system which are not easily obtained by any other type of experimental physicochemical method.

One of the NMR parameters for which a better theoretical understanding is being attained is the chemical shift, perhaps the most characteristic measurement obtained from NMR experiments. Both proton and carbon chemical shifts have been recognized by several authors as important indicators of regular secondary structure, particularly in proteins (Szilágyi, 1995), and have been employed on occasion in structural refinement using either empirical or theoretical relationships (Moyna et al., 1998; Pearson et al., 1995). While the former are based on the parametrization of relationships known from classical physics, such as ring current effects and peptide group anisotropies, against large databases of chemical shifts for a particular type of compound, the latter is based on the fact that accurate computation of chemical shifts can be made based solely on the molecular structure and conformation by means of *ab initio* molecular orbital (MO) calculations (de Dios, 1996). The accuracy of these *ab initio* methods has improved greatly in the past ten years, particularly after the efficient implementation of Ditchfield's Gauge-Including Atomic Orbital (GIAO) approach by Pulay and coworkers (Ditchfield, 1974; Wolinski et al., 1990), as well as the development of other protocols that eliminate the choice of gauge origin problem, such as IGLO, LORG, and CSGT (Kutzelnigg, 1980; Hansen and Bouman, 1984; Keith and Bader, 1993). The availability of these methods and of powerful and inexpensive computer systems not only allows for *ab initio* computation of chemical shifts in relatively large systems, such as peptide models, but also to employ larger basis sets and to include electron correlation schemes, either by use of density functional theory (DFT) or Møller-Plesset perturbation theory, both of which have been shown to improve the accuracy of calculations with respect to experimental data (Chesnut, 1994; de Dios, 1996). Furthermore, extensive studies of the effect that systematic variations of structural parameters have on chemical shifts, such as regular variations of torsion angles, can nowadays be performed by *ab initio* methods. In this regard, the extensive work of Oldfield and de Dios on protein and peptide models is representative (Oldfield, 1995; de Dios, 1996). Their findings indicate that the dependence of the Cα and Cβ chemical shifts with respect to the peptide backbone and side chain conformation can be described by two- or three-dimensional surfaces of ^{13}C chemical shift as a function of the ϕ, ψ, and χ dihedral angles. Since the chemical shift of a nucleus is determined mainly by its local environment, surfaces for individual amino acids can be derived, and several such surfaces have already been reported (Jiao, 1993; de Dios and Oldfield, 1994). Furthermore, the raw *ab initio* ^{13}C shift surfaces can be fitted to periodic functions of ϕ, ψ, and χ, allowing the use of experimental Cα and Cβ ^{13}C chemical shifts directly in protein structure refinement protocols (Pearson et al., 1995, 1997).

Similar to proteins, polysaccharides are composed of sugar monomers linked through glycosidic bonds to form, in some cases, large repetitive structures. These macromolecules are crucial in a wide variety of biological processes, including energy storage in living organisms, cellular and molecular recognition, and control of cellular structure and shape, to mention a few (Dwek, 1996). Central to all these functions is the ability of polysaccharides to adopt a wide range of dynamic conformations in solution (Imberty and Pérez, 2000). Thus, knowledge of the conformational characteristics of these compounds is vital to the pharmaceutical, medical, and drug-design communities. Contrary to proteins, in which the conformations of amino acid side chains are crucial determinants of their structure and function, polysaccharide structure is dictated mainly by the glycosidic bond dihedral angles ϕ and ψ (Figure 1). Experimental reports have shown that the ^{13}C chemical shift of the anomeric carbon in polysaccharides has a periodic dependence with the ϕ and ψ dihedral angles, and indicate that they could be used as probes of oligosaccharide conformation (Saitô, 1986; Jarvis, 1994). Despite this, and of the extensive studies carried out for peptide models mentioned above, only a few brief reports have appeared on systematic theoretical calculations of ^{13}C chemical shifts in polysaccharides and their dependence with the glycosidic bond conformation. These studies have either focused on variations of a single conformational parameter, or have used small model compounds to represent the glycosidic bond (Durran et al., 1995; Wilson et al. 1996, Hricovíni et al., 1997; Zhang et al., 1998). More comprehensive studies in this field are necessary, as they could have a large impact in the understanding of carbohydrate conformation and function. In particular, methods for structure refinement based on ^{13}C NMR data could be indispensable in the study of large polysaccharides, glycoproteins, and glycolipids for which the measurement of NOEs and J-couplings is experimentally unfeasible. In this report we wish to present an exhaustive *ab initio* study on the dependence of the ^{13}C

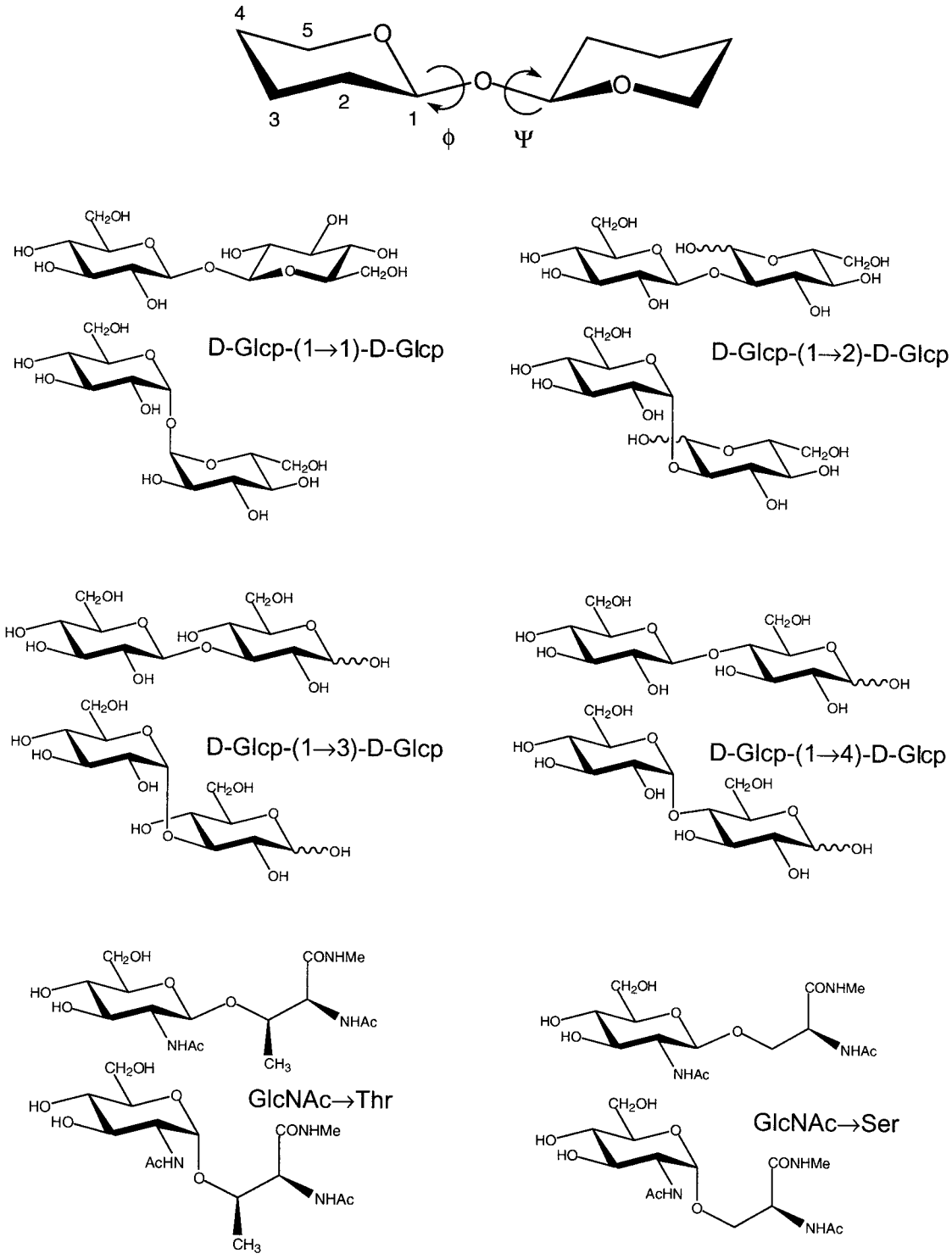

Figure 1. Definition of dihedral angles ϕ and ψ and sugar numbering scheme (top), and model molecules used for the GIAO ^{13}C chemical shift surface calculations. For disaccharides, ϕ is defined as $\langle H1 - C1 - O1 - Cn' \rangle$, and ψ as $\langle C1 - O1 - Cn' - Hn' \rangle$. For glycopeptides, ϕ is defined as $\langle H1 - C1 - O1 - C\beta \rangle$, and ψ is defined as $\langle C1 - O1 - C\beta - H\beta \rangle$ (Thr) or $\langle C1 - O1 - C\beta - H\beta_{proR} \rangle$ (Ser).

chemical shift of the anomeric carbons with the glycosidic bond conformation for a series of representative disaccharide and glycopeptide models (Figure 1). We describe the determination of *ab initio* ^{13}C chemical shift surfaces versus the ϕ and ψ dihedral angles, as well as the derivation of empirical functions of the form $^{13}C\delta = f(\phi,\psi)$ suitable for use in MD simulations and other structural refinement protocols.

Computational methodology

Input structures

D-Glcp-D-Glcp disaccharide models, including (1→1), (1→2), (1→3), and (1→4) glycosidic linkages in both α- and β-configuration, as well as glycopeptide model structures were built with Sybyl 6.5 (Tripos, Inc.). For glycopeptide models, the aglycone consisted of the amino acid capped as the *N*-acetylcarboxyl-*N*-methylamido derivative (AcN-Ser-CONHMe and AcN-Thr-COMHMe). In order to generate the chemical shift surfaces as a function of the glycosidic bond conformation, an 18 × 18 grid in the $-180° \leftrightarrow 180°$ range for ϕ and ψ at 20° intervals was constructed, to give a total of 324 input structures for each model molecule. For each structure in the grid, the ϕ and ψ were kept constant with dihedral constraints on heavy atoms and the monosaccharides held in the chair conformation, while the rest of the molecule was optimized using the AM1 semiempirical Hamiltonian as provided in Spartan 5.0.1 (Wavefunction, Inc.), resulting in an adiabatic energy surface for each model molecule. Sybyl and Spartan calculations were performed on Silicon Graphics 02 R10000 workstations.

Isotropic ^{13}C chemical shift calculations

Hartee-Fock (HF) theory and the Gauge-Including Atomic Orbital (GIAO) method as implemented in Gaussian 98 were employed for all NMR calculations (Wolinski et al., 1990; Frisch et al., 1998). Due to the computational expense of calculations at the 6-311G** theory level, shift surfaces were computed using the 3-21G basis set, and scaled to results from reference 6-311** level calculations. In order to obtain the scaling factor, duplicate GIAO ^{13}C calculations using the 3-21G and 6-311G** basis sets were performed on fully optimized (AM1 semiempirical) models of the eight disaccharide for which surfaces were to be derived. As discussed below, the computed ^{13}C shifts for a total of 96 carbons of the different models obtained using both basis sets were then correlated (Figure 2), and the resulting linear relationship employed to scale 3-21G results. For particular test cases, ^{13}C shift surfaces were computed using either a locally-dense basis set (6-311** for the anomeric carbon and atoms directly attached to it, 3-21G for the remaining atoms), or at the 6-311G** theory level for comparison to results from scaled calculations. In all cases, the isotropic ^{13}C chemical shift was estimated by subtracting the isotropic chemical shielding of the carbon atom to the one obtained at the same theory level for the methyl carbons of the NMR reference tetramethylsilane (TMS). In order to maintain consistency, the geometry of TMS was also optimized using the AM1 semiempirical prior to its use in ^{13}C shift computations. Gaussian 98 calculations were performed either on a HP/Convex Exemplar SPP-200, a HP N-4000 96-processor cluster, or a Beowulf cluster consisting of 16 Pentium-III 450 MHz compute nodes operating under LINUX.

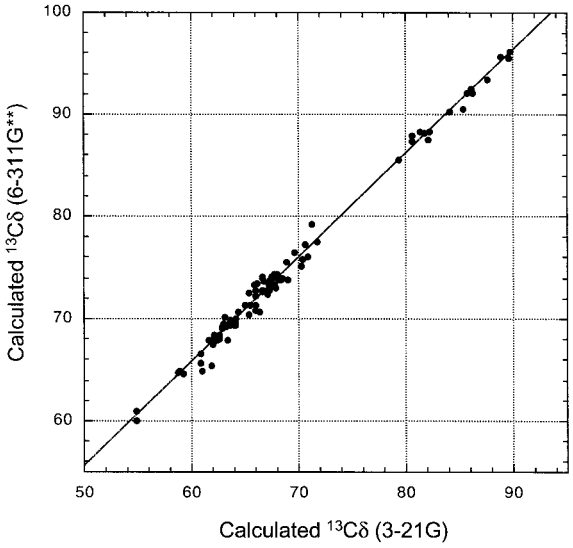

Figure 2. Correlation of GIAO ^{13}C calculations with 6-311G** and 3-21G basis sets. The two are related by the equation $^{13}C\delta_{6-311G^{**}} = 1.019 \times {}^{13}C\delta_{3-21G} + 4.6775$ ($r^2 = 0.992$).

Fitting of raw ab initio *data*

Isotropic ^{13}C shifts for the anomeric carbons obtained as described above were employed in the derivation of empirical equations relating the glycosidic bond $\langle\phi, \psi\rangle$ dihedral angles to the ^{13}C chemical shift. The raw 3-21G *ab initio* chemical shift data scaled to results from representative 6-311** calculations were fitted to trigonometric series expansions of general

form (Equation 1):

$$
\begin{aligned}
{}^{13}C\delta(\phi,\psi) = \; & \sum_i [A_i \sin(i\phi) + B_i \cos(i\phi) \\
& + C_i \sin(i\psi) + D_i \cos(i\psi)] \\
& + \sum_{i,j,\alpha,\beta} [A_{i,j,\alpha,\beta} \sin(i\alpha)\cos(j\beta) \\
& + B_{i,j,\alpha,\beta} \sin(i\alpha)\sin(j\beta) \\
& + C_{i,j,\alpha,\beta} \cos(i\alpha)\cos(j\beta)] + C_0,
\end{aligned}
$$

where α and β can be either φ and ψ. The data were fitted using Mathematica 3.0 (Wolfram Research, Inc.), using series with 91 (i, j = 1 to 3) or 325 (i, j = 1 to 6) terms. The resulting functions, implemented as Perl scripts, are available from the authors.

Results and discussion

Scaling of GIAO ^{13}C shielding calculations

One of the most important factors that weighs against the use of high level *ab initio* chemical shielding calculations, even in modestly large molecules, is their computational expense. For example, a single calculation of isotropic ^{13}C shieldings at the HF/6-311G** level of theory for one of the disaccharide models discussed here took, on average, 12 h of CPU time on a 550 MHz dual Pentium-III workstation. Therefore, the generation of complete ⟨φ, ψ⟩ surfaces for the number of model compounds we intended to analyze would have been impossible at a 20° resolution. Smaller basis, such as the Pople 3-21G set, reduced considerably the CPU time of a single-point NMR calculation to approximately 30 min. Unfortunately, it is well known that the calculated chemical shieldings are dependent on the size of the basis sets used, and they become asymptotic to experimental results as the basis sets become increasingly diffuse and correlation corrections are included (Chesnut, 1994; de Dios, 1996). However, earlier studies on α-(1→4)-glucan models indicate that HF/GIAO calculations at the 3-21G level of theory can faithfully reproduce the trends in ^{13}C shielding variations observed using a 6-31G** basis set (Durran et al., 1995). Furthermore, it has been shown that experimental chemical shifts for molecules containing H, C, N, and O can be estimated with equal accuracy using small (3-21G) or relatively large (6-31G*) basis sets if the results are scaled empirically to experiment using linear relationships (Forsyth and Sebag, 1997). These studies indicate that the errors arising from the use of less-than-optimal basis sets are systematic, and can be corrected by simple scaling procedures. We therefore investigated if this was the case in the calculation of ^{13}C chemical shifts in oligosaccharides. For that purpose, the results from GIAO chemical shift calculations using 3-21G and 6-311G** basis sets on the same set of AM1-optimized model disaccharides were compared (Figure 2). It is evident from the plot that there is an almost perfect linear correlation between results obtained using both basis sets (r^2 correlation coefficient of 0.992), and shieldings comparable to those obtained at the HF/6-311G** level of theory can thus be approximated by scaling calculations preformed with the 3-21G basis set. Therefore, this scaling procedure was employed to derive all ^{13}C chemical shift surfaces.

Generation of input conformations and shift surfaces for model glycosides

As mentioned above, a grid of 324 input structures spaced by 20° intervals in ⟨φ, ψ⟩ space was created by torsional driving of the glycosidic bond φ and ψ dihedral angles. The dihedral constraints used during the driving were applied to heavy atoms (i.e., ⟨O5 − C1 − O1 − Cn′⟩ for φ and ⟨C1 − O1 − Cn′ − C(n + 1)′⟩ for ψ) because deformation of the ⟨H1 − C1 − X⟩ and ⟨X − Cn′ − Hn′⟩ angles can result if restrains on hydrogens are used (Burkert and Allinger, 1982). Initially, we attempted to employ a general molecular mechanics force field (Tripos 6.0) to generate these input structures (Clark et al., 1989). However, it was realized that this molecular mechanics method failed to maintain the geometry around tetrahedral centers for structures of the grid with large non-bonded energy components. These structures present severe steric clashes between the two monosaccharides and have little physical meaning, but they are required in the process of creating continous energy and chemical shift surfaces. Torsional driving with the AM1 semiempirical method did not present these problems (Dewar et al., 1985), and was therefore employed in the generation of the input structures. There are two additional reasons that make the use of this strategy preferable. First, it is known that the anomeric and *exo*-anomeric effects are primary contributors to the observed behavior of the anomeric carbon chemical shift (Jarvis, 1994), and semiempirical methods reproduce these effects faithfully (Tvaroska and Carver, 1991; Woods et al., 1991). Second, the correlation between calculated and experimental chemical shifts has been shown to

improve as the level of theory employed in the optimization of the input structures used for the NMR calculations increases (Forsyth and Sebag, 1997), giving additional support for the choice of semiempirical geometry optimization methods. The resulting 324 conformers obtained for each model disaccharide using this approach were subsequently used as input in the HF/GIAO ^{13}C chemical shift calculations, using the 3-21G basis set followed by scaling to results from higher level of theory as described above. Since the shielding of the anomeric carbon is most sensitive to glycosidic bond conformation, we analyzed the ^{13}C chemical shift surface for this atom. Figure 3 shows the anomeric ^{13}C shift versus $\langle\phi, \psi\rangle$ surface for the disaccharides models D-Glcp-α-(1→1)-D-Glcp and D-Glcp-β-(1→1)-D-Glcp. Similar results were obtained for the remaining disaccharide models.

Although we were confident that the procedure described above was appropriate for the calculation of chemical shift surfaces of all the models, we decided to compare surfaces derived using a small basis set and scaling to those obtained with larger basis sets, using the same grid of AM1-optimized structures as input. In one case, a locally dense basis set was employed (Chesnut and Moore, 1989), using a 6-311G** basis for the anomeric carbon and atoms directly attached to it, and a 3-21G basis for all other atoms of the disaccharide model. The combination of these two sets has been reported to work well together for atoms of the second period (Hehre et al., 1986). For the other comparison, a uniform HF/6-311G** level of theory was employed for all atoms. As mentioned above, the computational cost of calculations using a more diffuse basis set is extremely high, and only representative model disaccharides were studied. Figure 4 shows chemical shift surfaces for the D-Glcp-α-(1→4)-D-Glcp disaccharide model obtained using the three methods described above. Qualitatively, the three surfaces are very similar, indicating that the simple scaling protocol can reproduce the trends in the ^{13}C chemical shift variation obtained from calculations at higher level of theory remarkably well. The same was found quantitatively when the differences between the surfaces were analyzed. The RMS deviation between the scaled surface and those created using the locally dense and the large basis sets were 0.990 and 0.980 ppm, respectively, for all 324 points on the grid. Similar results were obtained for the D-Glcp-β-(1→4)-D-Glcp disaccharide model, as well as for the α-GlcNAcSer/Thr glycopeptide models discussed below (data not shown). From these comparisons, it is clear that scaled ^{13}C chemical shift surfaces obtained with inexpensive basis sets can effectively reproduce results obtained from much more demanding calculations, justifying their use in the present study. It is also worth pointing out that, apart from the qualitative similarities, the RMS deviation between the surfaces obtained with the locally dense 6-311G**/3-21G and the balanced 6-311G** basis sets is only 0.270 ppm, corroborating earlier reports which indicate that there is good agreement among the two approaches (Chesnut and Moore, 1989).

We compared our results to earlier studies in which the dependence of the ^{13}C chemical shielding with the glycosidic bond $\langle\phi, \psi\rangle$ dihedrals was analyzed with HF/GIAO methods (Durran et al., 1995; Wilson et al., 1996; Zhang et al., 1998). In these reports, calculations were performed using either 3-21G or 6-31G* basis sets on models of α-(1→4)-linkages which had only a reduced number of atoms surrounding the anomeric carbon to reproduce the glycosidic bond electronic environment. The ^{13}C shielding and shift surfaces for the anomeric carbon reported in these studies show a variation of nearly 8 ppm, and an extrema at $\langle\phi = 0°, \psi = 0°\rangle$. Our results for disaccharides in the α-configuration correlate very well with these studies, displaying the same range of variation in chemical shift and, in our case, a chemical shift maximum centered around $\langle\phi = 0°, \psi = 0°\rangle$ (see Figure 3).

Our next set of simulations involved the glycopeptides models GlcNAcThr and GlcNAcSer in α- and β-configurations. Using an analogous protocol to the one presented above, we obtained ^{13}C chemical shift versus $\langle\phi, \psi\rangle$ surfaces for both models, and representative results are shown in Figure 5. For these models we also analyzed the effect of the peptide conformation on the anomeric carbon chemical shift. Three different surfaces for each model glycopeptide were computed. In one, the conformation of the peptide was allowed to relax freely during the generation of the conformer grid. In the second, the peptide was held in the C_5 (extended) conformation, and in the remaining case, in the α_r (α-helical) conformation. The three surfaces obtained for each glycopeptide model were virtually identical, corroborating that the chemical shift of the anomeric carbon is dictated mainly by its local environment. There are marked differences, however, between the results for the two models. For α- and β-GlcNAcThr, the ^{13}C shift surfaces resemble those obtained for D-Glcp-D-Glcp models described above, presenting a single distinct chemical shift max-

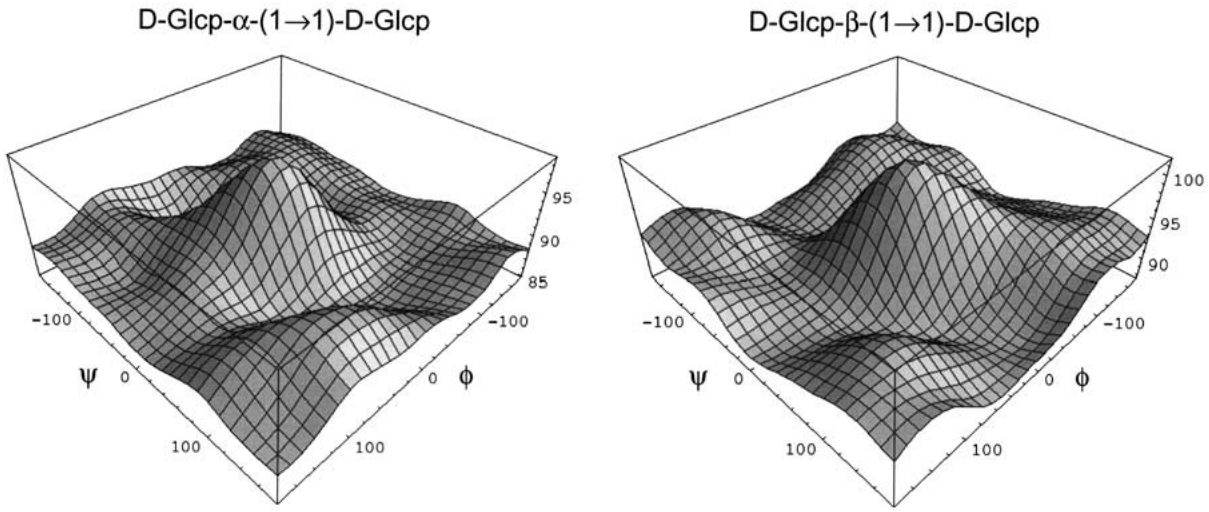

Figure 3. ^{13}C chemical shift surfaces for the anomeric carbon of D-Glcp-α-(1→1)-D-Glcp and D-Glcp-β-(1→1)-D-Glcp disaccharide models.

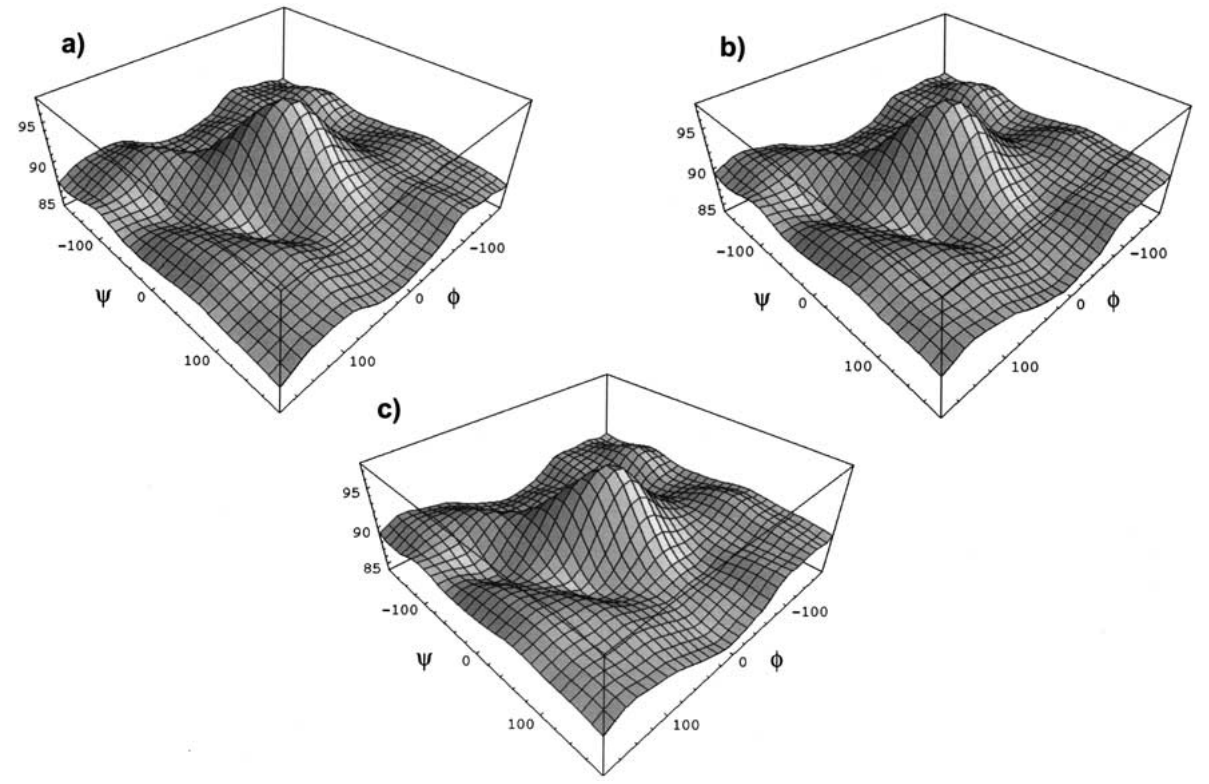

Figure 4. ^{13}C chemical shift surfaces for the anomeric carbon of D-Glcp-α-(1→4)-D-Glcp obtained using scaled 3-21G (a), locally dense 6-311G**/3-21G (b), and uniform 6-311G** (c) basis sets.

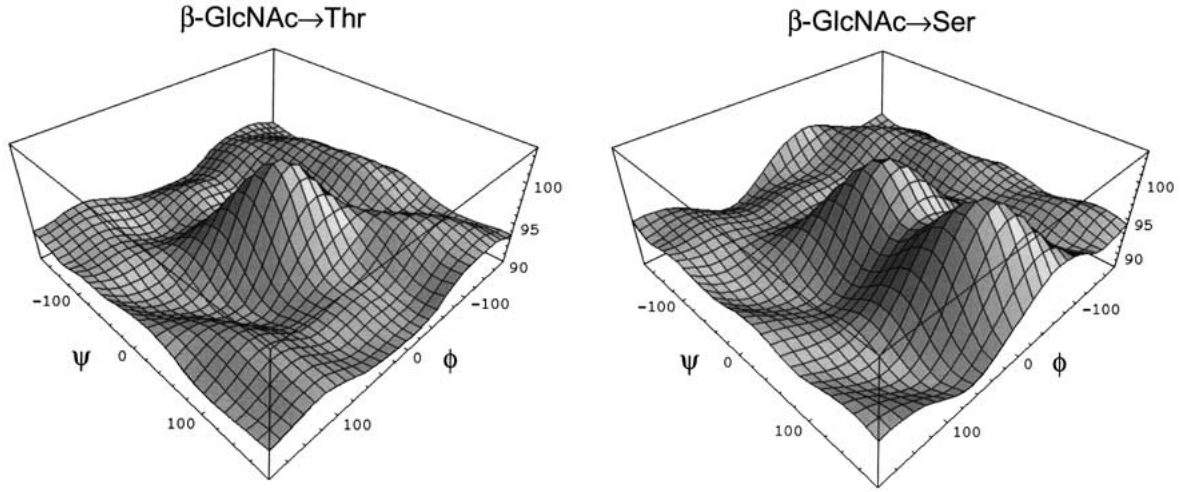

Figure 5. ^{13}C chemical shift surfaces for the anomeric carbon of GlcNAc→Thr and GlcNAc→Ser glycopeptide models in β-configuration.

ima around ⟨ϕ = 0°, ψ = 0°⟩. On the other hand, the surfaces obtained for α- and β-GlcNAc→Ser show two prominent chemical shift maxima. For the α-configuration models, there is a maximum near ⟨ϕ = −10°, ψ = 130°⟩ and a second one around ⟨ϕ = 0°, ψ = 0°⟩, while there are maxima near ⟨ϕ = 0°, ψ = 120°⟩ and ⟨ϕ = 0°, ψ = 0°⟩ for β-configuration models. Since the GlcNAc→Ser and the GlcNAc→Thr glycopeptide models differ only by a methyl group in the amino acid side chain, we decided to investigate if this small structural variation alone could account for the marked differences between the two chemical shift surfaces. Figure 6 shows 3D representations of the β-GlcNAc→Ser and β-GlcNAc→Thr conformers corresponding to grid points with ϕ = 0° and ψ = 120°. The only difference is that in the β-GlcNAc→Thr model the anomeric carbon is positioned directly on top of the Cα-Cγσ-bond of the threonine side chain methyl group, at a distance of approximately 2.5 Å. A similar positioning of the anomeric carbon with respect to the threonine side chain methyl group is seen for the β-GlcNAc→Thr model with ϕ = −10° and ψ = 130°. The differences in the chemical shift surfaces between the Thr and Ser glycopeptide models can therefore be explained, at least qualitatively, by the sterically induced polarization of the electron density around the anomeric carbon caused by the methyl group. While this effect will shield the anomeric carbon in the GlcNAc→Thr conformers discussed above, lowering its chemical shift, it will be absent in the corresponding GlcNAc→Ser conformers. To our knowledge, this is the first time a systematic *ab initio* study of the dependence of the anomeric carbon shift versus the ⟨ϕ,ψ⟩ dihedral angles for glycopeptide models is reported.

Derivation of ^{13}C chemical shift empirical functions

The next step in our study was the derivation of empirical functions of the form $^{13}C\delta = f(\phi,\psi)$ relating the anomeric ^{13}C chemical shift with the ⟨ϕ,ψ⟩ dihedral angles of the glycosidic bond for all disaccharide models. The periodicity of the chemical shift surfaces indicates that functions reproducing their topology could in principle be estimated the raw by fitting the *ab initio* data to series of trigonometric functions. Le and coworkers (1995) have uesed this type of function to represent the ^{13}C chemical shift variation of the Cα and Cβ carbons against the ⟨ϕ,ψ⟩ dihetral angles in peptide models. Therefore, we followed this approach, and fitted the data to a series with sines an cosines terms of the form $\sin(n\psi)$, $\cos(n\psi)$, $\sin(n\psi)$, and $\cos(n\psi)$, including all cross-terms (Equation 1). We analyzed how the variation of the number of terms used in the fit affected the ability of the resulting empirical equations to reproduce the *ab initio* results by comparing the RMS deviations between the raw and the estimated ^{13}C chemical shifts obtained from $^{13}C\delta(\phi,\psi)$ functions with 325 ($n = 6$) and 91 ($n = 3$) terms for all 324 points of the ⟨ϕ,ψ⟩ grid. The combined results for all disaccharide and glycopeptide chemical shift surfaces (nearly 4000 data points) show that the RMS deviation between the raw and calculated shifts rises from 0.31 to 0.56 ppm when going from 325 terms to 91 terms in the fit. Since one of

our goals is to employ these functions in structural refinement protocols and MD simulations (see below), the small increase in accuracy does not justify the use of more than 91 terms for the empirical ^{13}C chemical shift functions.

Comparison to experimental results

One of the most important features of theoretical chemical shift models is their ability to faithfully reproduce, and eventually predict, experimental NMR data. Most reports in which theoretical chemical shift estimation methods are evaluated correlate calculated values to experimental data for sets of relatively small rigid molecules (Chesnut, 1994; Forsyth and Sebag, 1997). Normally, the solid state and solution structures of the test compounds used in these studies are likely to resemble those obtained from geometry optimizations, and molecular flexibility does not have a considerable effect on the results. Peptides and disaccharides, are, on the other hand, highly flexible, and exist as an ensemble of conformers in solution. The measured chemical shift for a particular carbon in these molecules is therefore an average of the shift for that carbon in all conformers of the ensemble over a time period of milliseconds to seconds (de Dios, 1996; Imberty and Pérez, 2000). Thus, attempting to correlate experimental results to NMR calculations using a single geometry-optimized input structure would be erroneous. Since the relationship between the chemical shift versus $\langle\phi,\psi\rangle$ is explicitly described in chemical shift surfaces, they have the potential to address this issue. However, determining their accuracy is still problematic, as one would need to know the experimental chemical shift for isolated conformers with fixed ϕ and ψ dihedral angles. In the case of peptide chemical shift models, experimental NMR data for Cα and Cβ carbons can be obtained from solution studies of globular proteins for which X-ray structures are available. Since their structure in solution and in the crystal are similar, at least for residues not exposed to solvent (Billeter, 1992), experimental isotropic Cα and Cβ chemical shifts for a certain residue can be compared to those calculated from the ^{13}C chemical shift surface for the residue using the dihedral angles from the X-ray model. This approach is at the heart of the Z-surface method (Le et al., 1995), which has been successfully employed in the structural refinement of *Staphylococcus aureus* nuclease using chemical shift restraints (Pearson et al., 1995).

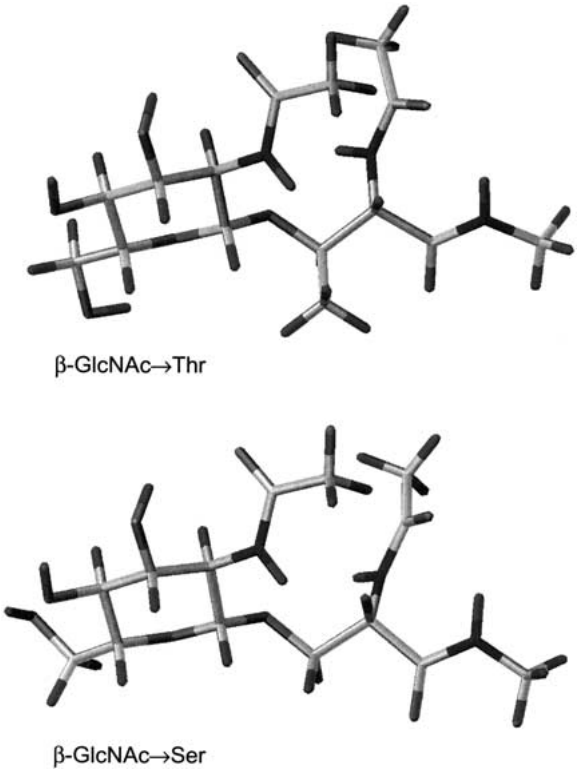

Figure 6. Molecular models of β-GlcNAc→Thr (top) and β-GlcNAc→Ser (bottom) corresponding to conformers of the $\langle\phi,\psi\rangle$ dihedral angle surfaces with $\phi = 0°$ and $\psi = 120°$.

Unfortunately, the number of high-resolution X-ray structures available for polysaccharides is far smaller than for proteins, amounting to less than a hundred for disaccharides, trisaccharides, and tetrasaccharides combined (Cambridge Structural Database, 1999). Furthermore, the large number of hydroxyl groups present in these molecules makes them interact strongly with water and become highly solvated in aqueous solution, and there is no evidence indicating similarities between solution and X-ray structures (Imberty and Pérez, 2000). Other sources of data or comparison methods are therefore needed to assess the accuracy of theoretical chemical shift models for oligosaccharides. One alternative is to calculate the anomeric chemical shift as the Boltzmann average of the chemical shift surface, $^{13}C\delta(\phi,\psi)$, over the corresponding $E(\phi,\psi)$ energy surface of the molecule (Equation 2):

$$\langle {}^{13}C\delta\rangle = \int [e^{-E(\phi,\psi)/k_BT} \times {}^{13}C\delta(\phi,\psi)]\mathrm{d}\phi\,\mathrm{d}\psi / \int [e^{-E(\phi,\psi)/K_BT}]\,\mathrm{d}\phi\,\mathrm{d}\psi.$$

Table 1. Comparison between experimental and calculated anomeric ^{13}C chemical shifts for D-Glcp-D-Glcp disaccharides

Model disaccharide	Anomeric ^{13}C chemical shift			
	Experimental	Calculated[a]	Calculated[b]	Calculated[c]
D-Glcp-α-(1→1)-D-Glcp	94.0	87.3	87.2	94.3
D-Glcp-α-(1→2)-D-Glcp	97.1	92.5	90.2	97.3
D-Glcp-α-(1→3)-D-Glcp	99.8	87.6	93.5	100.6
D-Glcp-α-(1→4)-D-Glcp	100.7	93.4	92.0	99.1
D-Glcp-β-(1→1)-D-Glcp	100.7	95.7	91.5	98.6
D-Glcp-β-(1→2)-D-Glcp	103.2	92.1	98.0	105.1
D-Glcp-β-(1→3)-D-Glcp	103.2	95.5	94.1	101.2
D-Glcp-β-(1→4)-D-Glcp	103.6	90.3	95.2	102.3

[a]Calculated from a single geometry-optimized structure.
[b]Calculated according to the Boltzmann distribution of conformers using Equation 2.
[c]Calculated as in (b), but using ^{13}C chemical shift surfaces referenced to dioxane.

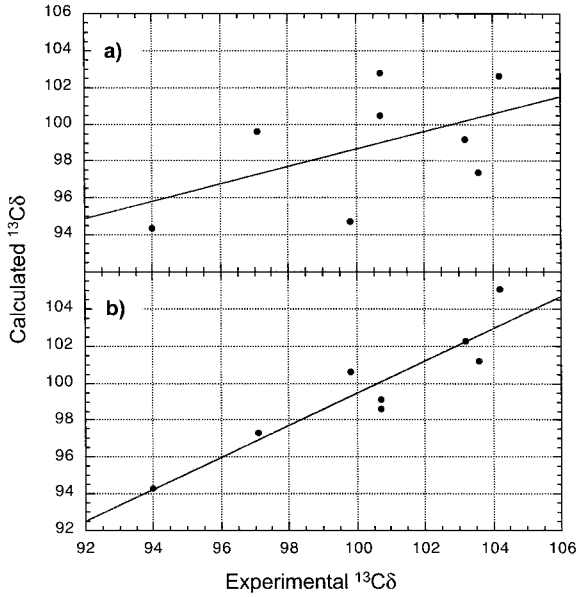

Figure 7. Correlation between experimental and calculated anomeric ^{13}C chemical shifts for D-Glcp-D-Glcp disaccharides. The shifts were calculated using a single geometry-optimized structure as input (a), or as a Boltzmann average using Equation 2 (b).

This approach takes into account the expected distribution of conformers for the molecule, and the calculated average chemical shift can be thus compared to solution NMR data. We therefore decided to apply this procedure to all our model disaccarides. The energy surfaces employed in the computation of the Boltzmann-averaged chemical shifts at the 3-21G level of theory had been calculated during the derivation of the chemical shift surfaces, and compare well to disaccharide energy surfaces reported by others (French et al., 2000). The experimental data sets were taken from Bock et al. (1984). Table 1 and Figure 7 summarize comparisons between experimental and calculated results for the anomeric carbons of all disaccharide models using two methods. In one case, the anomeric ^{13}C chemical shifts were calculated simply by using AM1-optimized structures as input, and in the other following the procedure described above using a temperature of 300 K. It is obvious that the anomeric ^{13}C shifts calculated from a single structure are, as mentioned above, very inaccurate. The RMS deviation against experiment is 9.2 ppm, and the r^2 correlation coefficient between calculated and experimental shifts is only 0.27. Furthermore, the slope of the linear regression is 0.48, which deviates considerably from the ideal value of 1.00. On the other hand, the results from chemical shift estimations in which Boltzmann-weighted averaging was taken into account are very encouraging. While the RMS deviation between calculated and experimental shifts is still large (7.7 ppm), the r^2 correlation coefficient for the linear regression rose to 0.81. The slope of the correlation line is 0.88 in this case, indicating better agreement with experiment. We realized that the large RMS deviation was likely due to systematic errors, whose origins can be associated with the choice of structure optimization method, the lack of correlation corrections in the *ab initio* calculations, or with the lack of solvation in the computation of the energy surfaces for the disaccharides. Perhaps the largest error originates from the reference used to compute the isotropic ^{13}C shifts. While all the experimental

oligosaccharide NMR data employed here were referenced to dioxane, the ^{13}C shift surfaces were derived using TMS as reference. Thus, a systematic correction to the chemical shift surfaces has to be applied. To do this, the basis set scaling equation was first recomputed with respect to dioxane instead of TMS, using the value of 67.6 ppm between the two reference compounds found experimentally (Wishart et al., 1995). A correction factor for all points in the surfaces was then obtained by taking the difference between the calculated and experimental shifts of dioxane with respect to TMS at the 3-21G level of theory, followed by scaling the result with the re-computed scaling equation. We did this and obtained a correction factor of +7.1 ppm which was applied to all the surfaces. If the Boltzmann-averaged ^{13}C chemical shifts are computed using corrected ^{13}C chemical shift surfaces, the RMS deviation against experimental data drops to 1.4 ppm (Table 1).

As mentioned above, additional assessment of the accuracy of the ^{13}C chemical shift surfaces can be done if NMR data from isolated conformers with fixed $\langle\phi,\psi\rangle$ angles is available. These type of data can be obtained from CP-MAS NMR spectra of crystalline disaccharides with reported X-ray structures. We obtained solid-state CP-MAS data for cellobiose (β-D-Glcp-(1→4)-D-Glcp) and trehalose (α-D-Glcp-(1→1)-D-Glcp) crystals (Hricovíni et al., 1997; Zhang et al., 1998), whose $\langle\phi,\psi\rangle$ dihedrals determined from X-ray are $\langle 43.3,-17.9\rangle$ and $\langle -57.4,-58.3\rangle$, respectively (Chu and Jeffrey, 1968; Jeffrey and Nanni, 1985). If the corrected $^{13}C\delta(\phi,\psi)$ surfaces for each of the disaccharides are evaluated at the X-ray $\langle\phi,\psi\rangle$ angles, ^{13}C chemical shifts of 106.5 and 93.6 ppm are obtained for the anomeric carbons of cllobiose and trehalose, which are in very good agreement with the experimental values of 104.9 and 93.0 ppm.

Conclusions

In this report we presented an exhaustive *ab initio* study of the dependence of the anomeric carbon chemical shift with the glycosidic bond dihedral angles for a representative series of disaccharide and glycopeptide models. Our results, which are in agreement with experimental observations and earlier theoretical calculations, indicate that the relationship between the chemical shift of the anomeric carbon and the $\langle\phi,\psi\rangle$ dihedral angles can be conveniently represented by empirical functions $^{13}C\delta(\phi,\psi)$. In principle, these equations could be used directly in pseudo-energy penalty terms of the for $E_\delta = K_\delta[^{13}C\delta_{\exp} - {}^{13}C\delta(\phi,\psi)]^2$ during the refinement of polysaccharide structure from ^{13}C NMR data, following an approach similar to the one employed in structure refinement and determination through the use of J-couplings and NOEs. However, we believe that it would be more appropriate to employ the $^{13}C\delta(\phi,\psi)$ functions as part of Z-surfaces (Le et al., 1995; Pearson et al., 1995), and use these to study the conformer distribution of carbohydrate-containing biomolecules in solution.

Excellent agreement was found between anomeric ^{13}C chemical shifts estimated with the $^{13}C\delta(\phi,\psi)$ functions and experimental values from solid-state NMR data on crystalline samples. The chemical shift surfaces also reproduce with relatively good accuracy the anomeric chemical shifts of oligosaccharides in solution if the Botlzmann distribution of conformers is taken into account. It is worth noting that the energy surfaces used to compute the Boltzmann-averaged ^{13}C shifts were derived *in vacuo*, therefore neglecting the effects that hydrogen bonding has over the energy map and conformational distribution of the disaccharides (Naidoo and Brady, 1999). The incorporation of water in the derivation of the energy surfaces, either explicitly or through the use of continuum solvation models (Foresman et al., 1996; Zauhar and Varnek, 1996), should result in a more realistic distribution of conformers, thus improving the accuracy of chemical shift estimations with respect to solution NMR data. We are currently investigating different routes to include these effects into our simulations. Alternatively, the chemical shifts can be back-calculated from MD simulations in which the solvent has been represented explicitly (Pearson et al., 1993). Since the conformer distribution is sampled explicitly through the MD simulations, only the $^{13}C\delta(\phi,\psi)$ functions are needed in this approach. We have done this using structural data from a 1 ns MD trajectory in water for the disaccharide α-Manp-(1→3)-β-GlcpNAc-OMe (Vishnyakov et al., 1999) and its corresponding ^{13}C chemical shift surface, which was derived following the protocol described in this report. Our preliminary results are encouraging (DeGrazia, personal communication), giving a back-calculated ^{13}C chemical shift range for the α-Manp anomeric carbon whose median is within 0.5 ppm from the reported experimental value.

Finally, the approach used for the calculation of chemical shifts in the present study was limited to HF/GIAO methodology. Several reports indicate that DFT methods are preferable due to their implicit

ability to incorporate electron correlation effects at low computational expense into carbohydrate chemical shielding estimations (Hricovíni et al., 1997; Dejaegere and Case, 1998). We have carried out an exhaustive comparison of shift calculations done with HF theory and different DFT methods which seems to corroborate these findings (Swalina, personal communication). We are currently analyzing the accuracy of ^{13}C chemical shift surfaces derived entirely with DFT for the same series of disaccharide and glycopeptide models.

In summary, we have extended the chemical shift surface method first developed for peptides to oligosaccharides and related moelcules. Although the present study concentrates on the anomeric carbon of disaccharide models, the approach can be extended to the non-anomeric carbons of the glycosidic linkage (i.e., Cn′ or Cβ). The chemical shift of these centers is also affected by the glycosidic bond torsion (Saitô, 1986), and the corresponding chemical shift surfaces would constitute independent but complementary indicators of linkage conformation. As pointed out in a recent review by Imberty and Pérez (2000), the conformational dependence of chemical shifts in oligosaccharides has been poorly understood. Despite additional comparisons to experimental data are required to further validate our results, we are hopeful that the studies presented here will help revert this situation. The findings from our ongoing investigations will be reported in due course.

Acknowledgements

This work was partially supported by the National Center for Supercomputer Applications under grants CHE990009N and CHE99048N. The authors also wish to thank Ivana Mihalek and Chuck Fisher, University of Kentucky Center for Computational Sciences, for their invaluable assistance during the initial setup of the shielding calculations.

References

Billeter, M. (1992) *Quart. Rev. Biophys*., **25**, 325–377.
Bock, K., Pedersen, C. and Pedersen, H. (1984) *Adv. Carbohydr. Chem. Biochem*., **42**, 193–225.
Burkert, U. and Allinger, N.L. (1982) *J. Comput. Chem*., **3**, 40–46.
Chesnut, D.B. (1994) *Annu. Rep. NMR. Spectrosc*., **29**, 71–122.
Chesnut, D.B. and Moore, K.D. (1989) *J. Comput. Chem*., **10**, 648–659.
Chu, S.S.C. and Jeffrey, G.A. (1968) *Acta Crystallogr. Sect. B*, **24**, 830–838.
Clark, M., Cramer, R.D. and Van Opdenbosch, N. (1989) *J. Comput. Chem.*, **10**, 982–1012.
Clore, G.M. and Gronenborn, A.M. (1997) *Proc. Natl. Acad. Sci. USA*, **95**, 5891–5898.
Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) *J. Am. Chem. Soc*., **117**, 5179–5197.
Cornilescu, G., Hu, J-S. and Bax, A. (1999) *J. Am. Chem. Soc*., **121**, 2949–2950.
de Dios, A.C. (1996) *Prog. NMR Spectrosc*., **29**, 229–278.
de Dios, A.C. and Oldfield, E. (1994) *J. Am. Chem. Soc*., **116**, 5307–5314.
Dejaegere A.P. and Case, D.A. (1998) *J. Phys. Chem. A*, **102**, 5280–5289.
Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Sewart, J.J.P. (1985) *J. Am. Chem. Soc*., **107**, 3902–3909.
Ditchfield, R. (1974) *Mol. Phys*., **27**, 789–807.
Durran, D.M., Howlin, B.J., Webb, G.A. and Gidley, M.J. (1995) *Carbohydr. Res*., **271**, C1–C5.
Dwek, R.A. (1996) *Chem. Rev*., **96**, 683–720.
Evans, J.N.S. (1995) *Biomolecular NMR Spectroscopy*, Oxford University Press, New York, NY.
Foresman, J.B., Keith, T.A., Wiberg, K.B., Snoonian, J. and Frisch, M.J. (1996) *J. Phys. Chem. A*, **100**, 16098–16104.
Forsyth, D.A. and Sebag, A.B. (1997) *J. Am. Chem. Soc*., **119**, 9483–9494.
French, A.D., Kelterer, A.-M., Johnson, G.P., Dowd, M.K. and Cramer, C.J. (2000), *J. Mol. Graphics Mod*., **18**, 95–107.
Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Zakrzewski, V.G., Montgomery, J.A., Stratmann, R.E., Burant, J.C., Dapprich, S., Millam, J.M., Daniels, A.D., Kudin, K.N., Strain, M.C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G.A., Ayala, P.Y., Cui, Q., Morokuma, K., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Cioslowski, J., Ortiz, J.V., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R.L., Fox, D.J., Keith, T.A., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P.M.W., Johnson, B.G., Chen, W., Wong, M.W., Andres, J.L., Head-Gordon, M., Replogle, E.S. and Pople, J.A. (1998) *Gaussian 98* (Revision A.7), Gaussian, Inc., Pittsburgh PA.
Halgren, T.A. (1996) *J. Comput. Chem*., **17**, 490–519.
Hansen, A.E. and Bouman, T.D. (1984) *J. Chem. Phys*., **82**, 5035–5047.
Hehre, W.J., Radom, L., Schleyer, P.V.R. and Pople, J.A. (1986) Ab Initio *Molecular Orbital Theory*, John Wiley & Sons, New York, NY.
Hricovíni, M., Malkina, O.L., Bízik, F., Nagy, T. and Malkin, V.G. (1997) *J. Phys. Chem. A*, **101**, 9756–9762.
Imberty, A. and Pérez, S. (2000) *Chem. Rev*., **100**, 4567–4588.
Jarvis, M.C. (1994) *Carbohydr. Res*., **259**, 311–318.
Jeffrey, G.A. and Nanni, R. (1985) *Carbohydr. Res.*, **137**, 21–30.
Jiao, D., Barfield, M. and Hruby, V.J. (1993) *J. Am. Chem. Soc*., **115**, 10883–10887.
Keith, T.A. and Bader, R.F.W. (1993) *Chem. Phys. Lett*., **210**, 223–231.
Kutzelnigg, W. (1980) *Isr. J. Chem*., **19**, 193–200.
Le, H., Pearson, J.G., de Dios, A.C. and Oldfield, E. (1995) *J. Am. Chem. Soc*., **117**, 3800–3807.

Moyna, G., Williams, H.J., Nachamn, R.J. and Scott, A.I. (1999) *J. Peptide Res*., **53**, 294–301.
Naidoo, K.J. and Brady, J.W. (1999) *J. Am. Chem. Soc*., **121**, 2244–2252.
Oldfield, E. (1995) *J. Biomol. NMR*., **5**, 217–225.
Pearson, J.G., Oldfield, E., Lee, F.S. and Warshel, A. (1993) *J. Am. Chem. Soc*., **115**, 6851–6862.
Pearson, J.G., Le, H., Sanders, L.K., Godbout, N., Havlin, R.H. and Oldfield, E. (1997) *J. Am. Chem. Soc*., **119**, 11941–11950.
Pearson, J.G., Wang, J.-F., Markley, J.L., Le, H. and Oldfield, E. (1995) *J. Am. Chem. Soc*., **117**, 8823–8829.
Saitô, H. (1986) *Magn. Reson. Chem*., **24**, 835–852.
Szilágyi, L. (1995) *Prog. NMR Spectrosc*., **27**, 325–443.
Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.
Tvaroska, I. and Carver, J.P. (1991) *J. Chem. Res. (S)*, 6–7.
Vishnyakov, A., Widmalm, G., Kowalewski, J. and Laaksonen, A. (1999) *J. Am. Chem. Soc.*, **121**, 5403–5412.
Wilson, P.J., Howlin, B.J. and Webb, G.A. (1996) *J. Mol. Struct*., **385**, 185–193.
Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135–140.
Wolinski, K., Hinton, J.F. and Pulay, P. (1990) *J. Am. Chem. Soc*., **112**, 8251–8260.
Woods, R.J., Szarek, W.A. and Smith, V.H. (1991) *J. Chem. Soc., Chem. Commun*., 334–337.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York, NY.
Zauhar, R.J. and Varnek, A. (1996) *J. Comput. Chem*., **17**, 864–877.
Zhang, P., Klymachyov, A.N., Brown, S., Ellington, J.G. and Grandinetti, P.J. (1998) *Solid State Nucl. Magn. Reson*., **12**, 221–225.