



Published in final edited form as:

J Comput Chem. 2014 September 30; 35(25): 1854–1864. doi:10.1002/jcc.23697.

Factors Affecting the Computation of the ^{13}C Shielding in Disaccharides

Pablo G. Garay^[a], Osvaldo A. Martin^{[a],[b]}, Harold A. Scheraga^[b], and Jorge A. Vila^{[a],[b]}

Harold A. Scheraga: has5@cornell.edu; Jorge A. Vila: jv84@cornell.edu

^[a]IMASL-CONICET, Universidad Nacional de San Luis, Italia 1556, 5700-San Luis, Argentina

^[b]Baker Laboratory of Chemistry, Cornell University, Ithaca, New York

Abstract

Knowledge of the three-dimensional structures of glycans and glycoproteins is useful for a full understanding of molecular processes in which glycans are involved, such as antigen-recognition and virus infection, to name a few. Among the ubiquitous nuclei in glycan molecules, the ^{13}C nucleus is an attractive candidate for computation of theoretical chemical shifts at the quantum chemical level of theory to validate and determine glycan structures. For this purpose, it is important to determine, first, which carbons can be used as probes to sense conformational changes and, second, all factors that affect the computation of the shielding, at the density functional theory (DFT) level of theory, of those carbons. To answer such questions, we performed a series of analyses on low-energy conformations, obtained by sampling the glycosidic torsional angles (φ , ψ) every 10° , of 12 disaccharides. Our results provide evidence that: (i) the carbons that participate in the glycosidic linkage are the most sensitive probes with which to sense conformational changes of disaccharides; (ii) the rotation of the hydroxyl groups closest to the glycosidic linkage significantly affects the computation of the shieldings of the carbons that participate in the glycosidic linkage; (iii) it is not possible to obtain the shieldings of one disaccharide from the computed values of a different disaccharide or from those disaccharides that differ in the anomeric state; and (iv) a proper basis set distribution, a functional, and a step size, with which to sample the conformational space, are necessary to compute shieldings accurately and rapidly.

Keywords

glycans; quantum-chemical calculation of ^{13}C -shieldings; validation of glycan structures; glycoproteins

Introduction

During the last two decades, the science of glycobiology has gained momentum, and the accumulated evidence has prompted us to no longer treat glycans as molecules of secondary

importance compared to other biomolecules such as proteins and nucleic acids. For example, it is estimated that 50% of all proteins are glycosylated.^[1] However, only 3.5% of the proteins in the Protein Data Bank^[2] (PDB) occur in a glycosylated form.^[3] There are many reasons for such a low percentage of glycosylated proteins in the PDB, among others, the following: glycan chains are very flexible and, therefore, are often removed to facilitate crystal growth; even when not removed, glycans often yield poor quality electron density, preventing accurate resolution of the three-dimensional (3D) structure.^[3] The rather large number of errors in the available PDB entries containing glycans^[3] point to another issue: the need for more and better methods for evaluating the quality of the 3D structures of glycan and glycoprotein molecules. Consequently, the need for accurate and fast validation methods to detect flaws in glycan and glycoprotein structures, at the residue/disaccharide level, is crucial.

Over the last several years, we have focused on developing computational tools, such as the CheShift-2 server,^[4,5] for automated validation of X-ray- and Nuclear Magnetic Resonance (NMR)-determined protein structures, provided that the observed $^{13}\text{C}^\alpha$ and/or $^{13}\text{C}^\beta$ chemical shifts are available. The latter accomplishment encouraged us to start developing a new methodology, based on density functional Theory (DFT)-computed ^{13}C shieldings, to validate, refine, and determine glycan, glycoprotein, and other glycoconjugated molecules. Achievement of this goal would be an important step forward for the structural glycoscience field, because it is well-known that the measurement of nuclear overhauser effect (NOEs) and J-couplings are experimentally either difficult or unfeasible to obtain for such carbohydrates. As mentioned above, the available structural data for glycans are sparse. As a consequence, it is unlikely that we can envisage, in short term, the development of knowledge-based, rather than physics-based, methods for predicting chemical shifts in glycans. This is contrary to common practice in the protein field in which several knowledge-based methods are available to predict chemical shifts in proteins (Han et al.,^[6] and references therein), mainly because of the large number of high-resolution protein structures in the PDB.

To attain the ambitious goal of developing a physics-based method with which to validate, refine, and determine glycan, glycoprotein, and other glycoconjugated structures, it is necessary to start by examining, in detail, all the factors affecting the computation, at the DFT-level of theory, of the ^{13}C shielding, as a function of the conformational changes in disaccharides, for example, by testing the relative ability of the ^{13}C nuclei to sense variations of (φ , (ψ) glycosidic torsional angles.

The ^{13}C nucleus is an attractive candidate for computation of chemical shifts at the quantum chemical level of theory to validate and determine glycan structures. In this regard, there is experimental evidence^[7,8] showing that the ^{13}C chemical shift of the carbon that participates in the glycosidic linkage has a periodic dependence on the φ and ψ dihedral angles. Such evidence led Swalina et al.^[9] to assume that the carbons participating in the glycosidic linkage could be used as probes for oligosaccharide structural determination. However, to the best of our knowledge, there is no rigorous test of such assumption. In addition, a few brief reports appeared about systematic theoretical calculations of ^{13}C chemical shifts in polysaccharides and their dependence on the conformation of the glycosidic bond.^[10–13] In

addition, a physics-based method to determine the 3D structures of oligosaccharides has been proposed.^[9,14] This method is proof of a concept that the chemical shifts of carbons can be used to obtain structural information of glycans. However, some possible limitations are involved in the proposed method of these authors: (i) the carbons that participate in the glycosidic linkage were adopted as the probes with which to sense disaccharide conformations without performing tests to assure that these carbons are, in fact, the best choice; (ii) the effects of the rotamer states of the hydroxyl groups were not considered; (iii) the 20° step used to sample the torsional φ and ψ angles may have been too crude for an accurate prediction of chemical shifts, because the ^{13}C chemical-shift surface is rough; (iv) the basis set 3-21G chosen to treat all atoms for the DFT-calculations, may not be accurate enough; and (v) neither Swalina et al.^[9] nor Sergeev and Moyna^[14] analyzed the transferability of the results between disaccharides. All these limitations, and other factors affecting an accurate computation of the ^{13}C shieldings, are addressed in the following sections.

Materials and Methods

Generation of disaccharide conformations

Even though glycans can be large and flexible molecules, their conformations can be described essentially by the torsional angles φ , ψ , and ω . From this point of view, the smallest representative unit of a glycan is a disaccharide (two carbohydrates linked by a glycosidic bond) and, hence, in the present work, we will focus on studying disaccharides. The torsional angles (φ , ψ) of the glycosidic link of the disaccharides were defined using the NMR convention, namely φ (H1–C1–O–C4') and ψ (C1–O–C4'–H4'), see Figure 1.^[15]

An initial template of each disaccharide was built with the software SWEET2.^[16] Although different puckering of the pyranose ring, such as C (for chair), B (for boat), S (for skew), and H (for half-chair),^[17] may influence the ^{13}C shieldings of the probe nucleus, we decided to fix the puckering to C (chair), and among the possible C conformations to $^4\text{C}_1$, because this is the most frequently observed puckering of the pyranose ring in both X-ray^[18] and NMR-determined structures.^[19] Then, the energy of the corresponding geometry was minimized using the molecular mechanics force-field MM3.^[20] To obtain the resulting energy-minimized conformation, each of the glycosidic torsional angles (φ , ψ) was varied every 10° using the PyMOL package^[21] and conformations with an internal energy higher than –100 kcal/mol were removed. As a result, we generated an ensemble for each disaccharide of about 550 conformations, although the exact number of conformations differs for each disaccharide; thus, for example, 581 conformations were generated for maltose [α -D-Glcp-(1-4)- α -D-Glcp]. It is important to highlight that conformations are generated using a rigid geometry approximation, that is, the bond lengths and bond angles are fixed, and different conformations differ only in their torsional angles. The 581 conformations were generated by assuming a fixed arbitrary value for each of the torsional angles of the hydroxyl groups, unless otherwise noted.

The dependence of the ^{13}C shielding on variations of the torsional angle ω (O1–C6'–C5'–H5'), present in the glycoside (1–6) link, was not treated in this work. However, in future applications, we plan to allow the ω torsional angle to sample three rotameric states, namely

+60°, -60°, and 180° rather than only the two, viz., +60° and -60°, frequently seen in structures deposited in the PDB; the reason to increase the number of rotamers beyond those most commonly seen in the PDB is based on the fact that the PDB contains only a small fraction of a large diversity of glycans present in nature.

Computation of the ^{13}C shieldings

For a given disaccharide conformation, a functional and a basis set distribution (BSD) of the ^{13}C isotropic-shielding values were always computed using the gauge invariant atomic orbital procedure^[22] and DFT methods as implemented in the GAUSSIAN 03 suite of programs.^[23] The gas-phase ^{13}C isotropic-shielding calculations were always computed without explicit consideration of inter- or intramolecular interactions. Implicit in this decision is that the conformations of a given disaccharide could be influenced by intra- or intermolecular interactions but, once these conformations are established, the ^{13}C isotropic-shielding values depend mainly on the torsional angles φ , ψ , and the rotation of the hydroxyl groups.

Computations of the entropy

From the distribution of shieldings, computed for an ensemble of disaccharide conformations, it is possible to calculate its entropy S using the following equation:

$$S = -\sum_i P_{(i)} [\ln P_{(i)}] \quad (1)$$

where $P_{(i)}$ is the probability of observing a given carbon shielding with value i . The probability $P_{(i)}$ was obtained from the computed histogram of the shieldings using bins of 0.5 ppm.

We use the entropy, S , as a measure of the sensitivity of a nucleus to variation in torsional angles, for the following reason. The ideal nucleus with which to build an application that provides structural information from the chemical shifts is one in which there is a one-to-one correspondence between chemical shift and conformation. In this circumstance, the distribution function of chemical shifts would be uniform. Unfortunately, chemical shifts are multivalued functions of the torsional angles, and chemical shifts tend to exhibit more peaked distributions closer to a Gaussian distribution. Consequently, a convenient way to measure the uniformity of a shielding-distribution is the entropy of the distribution. In fact, the entropy of the shielding distribution is a measure of the information provided by a given nucleus on variations of the φ and ψ torsional angles.

Results and Discussion

Test of the step size to sample the conformational space

It is important to determine accurately whether the samplings of the (φ , ψ) torsional angles (Fig. 1) must be specified every 10°, rather than 20° used in Ref. [9], because the sampling step determines: (i) the total number of conformations for which the shielding values must be computed at the DFT-level of theory for significant accuracy, whereas as is known, the computations are very CPU-time demanding, and (ii) the accuracy with which the chemical-

shift value can be obtained by interpolation of the computed shieldings. Consequently, to obtain the error associated with a chosen step size quantitatively, for example, 20° , we start by analyzing a 1D problem, that is, by fixing the torsional angle φ to an arbitrary value. Then, the shielding values are computed by DFT by sampling the torsional angle ψ every 20° , namely for $\psi = 0^\circ, 20^\circ, 40^\circ, 60^\circ, 80^\circ$, and so forth, to obtain the intermediate shielding values by “interpolation,” namely for $\psi = 10^\circ, 30^\circ, 50^\circ, 70^\circ$, and so forth; from here on, the “interpolated” shielding values are denoted as $\delta_{\text{interpolate}}$. The shielding values for the set of torsional angles ψ of $10^\circ, 30^\circ, 50^\circ, 70^\circ$, and so forth, are also computed directly by DFT; from here on, the “true” shielding values are denoted as δ_{true} . Then, the absolute difference between the “interpolated” and the “true” shielding values, that is, $\text{shielding} = |\delta_{\text{interpolation}} - \delta_{\text{true}}|$, are used to assess whether the sampling every 20° rather than 10° is a good approximation. This procedure can be generalized straightforwardly to other sets of (φ, ψ) torsional angles of the glycosidic linkage of the maltose disaccharide.

In Figure 2, we plot the normalized histogram of the shielding values obtained using two interpolation methods, namely a linear one (Fig. 2a) and a cubic one (Fig. 2b). Despite the differences between the interpolation methods used, the results shown in Figure 2 indicate that more than 50% of the shielding differences, between the “interpolated” and the “true” values are larger than 0.5 ppm. If 0.5 ppm is adopted as a cutoff value beyond which the computed shielding difference is considered important, by analogy with results obtained from sequence-dependent effects in proteins,^[24] then the results shown in Figure 2 highlight the importance of the use of 10° rather than 20° sampling, for an accurate shielding sampling of the torsional angles of the glycosidic link. This result is in line with those obtained for proteins.^[4]

It is worth noting that sampling the (φ, ψ) torsional angles at, say, 5° rather than 10° may further improve the accuracy of the shielding calculations. However, adoption of this finer-grid sampling would require a significantly larger number of conformations for which the carbon shieldings need to be computed by DFT and, hence, not feasible with existing computational resources.

Determination of a probe carbon with which to sense glycosidic torsional angle variations

To determine which ^{13}C nuclei are the most sensitive ones with which to sense the variation of the glycosidic torsional angles φ and ψ (Fig. 1), the shielding values for all the ^{13}C nuclei of 12 disaccharides were computed. Explicitly, eight disaccharides with glycosidic linkage (1–4), viz., for $\alpha\text{-D-Glcp-(1-4)-}\alpha\text{-D-Xxxp}$, where Xxx = Glu (Glucose), All (Allose), Alt (Altrose), Gal (Galactose), Gul (Gulose), Ido (Idose), Man (Mannose), Tal (Talose), and four disaccharides with glycosidic linkage (1-Y), viz., for $\alpha\text{-D-Glcp-(1-Y)-}\alpha\text{-D-Glcp}$, where Y = 1, 2, 3, and 6, will be considered. It should be noted that Y = 5 is not included because the C5 nucleus cannot participate in a glycosidic linkage (Fig. 1). From the DFT-computed shieldings with a B3LYP functional and a 6-311+G(2d,p) basis set, for ~550 conformations of each of the 12 disaccharides, the entropy distribution for each disaccharide was obtained using eq. (1). The results of these analyses are summarized in Figure 3. From this figure, it can be seen that, for each of the eight disaccharides with glycosidic linkage (1–4), the nuclei with the highest entropy (darkest blue) are those that participate in the glycosidic linkage,

namely columns $^{13}\text{C}1$ and $^{13}\text{C}4'$ for the results from rows **A** to **H**. From the same Figure 3, it can also be seen that for each of the four disaccharides with glycosidic linkage (1- Y), that is, with $Y=1, 2, 3$, and 6 , the nuclei with the highest entropy (darkest blue) are also those that participate in the glycosidic linkage, namely columns $^{13}\text{C}1$, $^{13}\text{C}1'$, $^{13}\text{C}2'$, $^{13}\text{C}3'$, and $^{13}\text{C}6'$ for the results from rows **I** to **L**. It is worth noting that some panels of Figure 3 display comparable darkness, for example, see columns $^{13}\text{C}4'$ and $^{13}\text{C}6'$ of row **H**. This happens because the adopted coloring representation used to highlight the entropy differences between nuclei makes it difficult to distinguish it clearly, although the carbons that participate in the glycosidic linkage are always the most sensitive nuclei, in terms of entropy, with which to sense the variations of the φ and ψ torsional angles.

Overall, these results, which support the assumption of Swalina et al.,^[9] provide solid evidence that the carbons that participate in the glycosidic linkage can be adopted as probes with which to sense variations of the glycosidic torsional angles.

Analysis of the computed ^{13}C shielding distributions for maltose and the experimental ones from amylose

If the polysaccharide amylose is regarded as a collection of linked maltose molecules [α -D-Glcp-(1-4)- α -D-Glcp], then an interesting observation arises. The ^{13}C NMR spectra observed for amylose in aqueous solution, by Saitô,^[7] show three closely spaced resonance peaks, namely those for the $^{13}\text{C}3$, $^{13}\text{C}2$, and $^{13}\text{C}5$ nuclei, and three peaks significantly displaced among them, namely those for the $^{13}\text{C}1$, $^{13}\text{C}4$, and $^{13}\text{C}6$ nuclei, respectively.^[7] Interestingly, the DFT-computed distribution of the mean shielding value for the ^{13}C nuclei of maltose, shown in Figure 4, follows a comparable trend to that observed by Saitô^[7] for amylose. Thus, from Figure 4, we can identify two clusters, depending on how the shielding distributions are grouped, viz., $^{13}\text{C}3$ (~99 ppm), $^{13}\text{C}2$ (~101 ppm), and $^{13}\text{C}5$ (~101 ppm), as one cluster, with almost superimposed shielding, and $^{13}\text{C}1$ (~78 ppm), $^{13}\text{C}4$ (~107 ppm), and $^{13}\text{C}6$ (~115 ppm), as another cluster, with spread out shielding, with the mean shielding value for each carbon indicated here in parentheses.

Taken as a whole, the good correspondence between the DFT-computed shielding values from maltose and the resonance values observed by Saitô,^[7] for amylose, from NMR spectroscopy support our quantum-chemical-based analysis of the chemical shifts in glycans.

Test of BSDs

The smallest representative unit of a glycan is a disaccharide and hence, in the present work, we will focus on studying disaccharides. To improve the very CPU-time consuming DFT calculations, we have tested 11 BSDs to find a trade-off between accuracy and speed in the quantum chemical computations of the shieldings of disaccharides. For this purpose, the DFT-computations of shielding were performed on all atoms of the 581 conformations of maltose using the B3LYP functional. Initially, every C, O, and H atom in the disaccharide of Figure 5a was treated with a uniform 6-311+G(2d,p) BSD; designated from here on, as BSD [00]. The shielding values computed using the BSD [00] were adopted as an “internal-reference” for comparison with the shielding values computed by the other BSD's.

The dependence of both the accuracy and the speed of the DFT-computed shieldings, relative to the 581 shielding values computed with the BSD [00], was tested with a uniform 3-21G BSD (Fig. 5b) and with nine locally dense BSD's,^[25] namely, six locally dense BSD's with the 311+G(2d,p)/3-21G basis set (Figs. 5c–5h) and 3 locally dense BSD's with the 6-31G/3-21G basis set (Figs. 5i–5k). Although the total number of possible combinations of locally dense BSD's is larger than the nine tested in this work, an accurate and fast BSD, with which to compute the shielding of the $^{13}\text{C1}$ and $^{13}\text{C4'}$ nuclei, was determined straightforwardly (see below).

The nine locally dense BSD are illustrated for the $^{13}\text{C1}$ nucleus in each panel of Figures 5c–5k in which a colored disaccharide representation is inserted, with blue, red, and green representing the 6-311+G(2d,p), 6-31G, and 3-21G basis sets, respectively; this notation refers to the basic basis sets of Pople and coworkers,^[26] as implemented in Gaussian-03.^[23] The corresponding results for the $^{13}\text{C4'}$ nucleus are shown in Supporting Information Figure S1.

Analysis of the results of Figures 5b–5k indicates, first, that a proper locally dense BSD is crucial for reproducing the results obtained with the “internal-reference” BSD [00] accurately and, second, that the computation of the shielding with the BSD [04] represents the best option, in terms of both accuracy and CPU-time, among all 10 BSD's tested. Computation of the shielding with the BSD [04] ($R^2=0.997$, $t=01:19:38$, see Fig. 5e) is ~4 times faster than that with the BSD [02] ($R^2=0.999$, $t=05:30:50$, see Fig. 5c) although ~10 times slower than that with the BSD [01] ($R^2=0.917$, $t=00:07:55$, see Fig. 5b). It is worth noting that a BSD [01], shown in Figure 5b, was used, after extrapolating it to a large basis set, by Swalina et al.^[9] Even though the BSD [01] is notably faster than the BSD [04], the latter provides appreciably better accuracy. Consequently, we will adopt the BSD [04] for our future computations to develop a physics-based method to validate and determine glycan structures.

Test of density functional

The results obtained using the B3LYP functional were adopted here as an “internal-reference” because this functional has proven to be a very good choice to predict the shielding tensor for a great variety of compounds containing ^{13}C , ^{15}N , and ^{17}O nuclei.^[27–29] However, B3LYP is not the fastest functional with which to compute shieldings,^[30] and this could severely limit the applicability of the methodology because of the large number of conformations that must be computed for each possible combination of disaccharides. Consequently, computation of the shielding for each $^{13}\text{C1}$ of the 581 conformations of maltose using six alternative functionals, all of them faster than B3LYP, was performed for maltose, with the OPBE, OPW91, BPW91, OB98, BPBE, and OLYP functionals.^[30] The results of this test are shown in Figure 6. In each panel of this figure, the shielding values for the $^{13}\text{C1}$ nucleus computed with the B3LYP functional with basis set [00] are plotted on the y-axis versus the shielding values on the x-axis, computed using each of the remaining six functionals with the BSD [04]. The coefficient of determination, R^2 , and the averaged CPU time, computed over all the 581 conformations, are inserted in each panel. The corresponding results for the $^{13}\text{C4'}$ nucleus are shown in Supporting Information Figure S2.

Although all the functionals show similar accuracy, we focus our attention on OB98 and OLYP because these two functionals are slightly more accurate ($R^2=0.981$ and $R^2=0.984$, respectively) than the remaining ones to reproduce the results obtained with B3LYP, using BSD [04] and a uniform basis set [00]. Finally, the functional OB98, rather than OLYP, was selected because it will enable us to integrate the shieldings computed for disaccharides with those computed previously in our group for each of the 20 naturally occurring amino acids.^[4,30] It is very important to be consistent in this manner to compute shielding in disaccharides and peptides with the same functional because this will enable us to validate, refine, and determine glycan, protein, and glycoprotein structures.

Overall, the $^{13}\text{C}1$ shielding (δ) computed with B3LYP and the BSD [00] can be obtained rapidly and accurately, with the OB98 functional and the BSD [04], using the following linear regression:

$$\delta_{\text{B3LYP}} = -3.766 + 1.012 \times \delta_{\text{OB98}} \quad (2)$$

where -3.766 and 1.012 represent the linear regression parameters from Figure 6d, for both the y intercept and the slope of the regression, respectively.

Analysis of the transferability of the calculated shieldings

As mentioned in the Test of BSDs section, DFT computations of shieldings are expensive and, hence, it is important to determine whether the shielding computed from a given disaccharide could be used as a model template with which to compute the shielding for several other disaccharides. Because this question is very important for building a large database of disaccharide shieldings, it will be analyzed here in detail.

The transferability of the shielding values between disaccharides was tested by analyzing the coefficient of determination (R^2) computed for both the $^{13}\text{C}1$ and the $^{13}\text{C}4'$ nucleus, respectively, that is, these tests were performed only for glycosidic (1–4) linkages. Thus, the shielding values from ~550 conformations were computed for each of eight disaccharides, namely for: $\alpha\text{-D-Glcp-(1-4)-}\alpha\text{-D-}\mathbf{Xxxp}$, where \mathbf{Xxx} = Glc (Glucose), All (Allose), Alt (Altrose), Gal (Galactose), Gul (Gulose), Ido (Idose), Man (Mannose), and Tal (Talose). The results for the shielding correlations between all eight different disaccharides are plotted in Figure 7. In this figure, those windows showing a coefficient of determination $R^2 > 0.9$ are highlighted by red frames; only those 5 and 7 out of a total 28 windows possess $R^2 > 0.9$ for the $^{13}\text{C}1$ and $^{13}\text{C}4'$ nucleus, respectively. This result enables us to conclude that there is not sufficient transferability of shielding between different disaccharides. In other words, it is not possible to reproduce all the computed shielding values of the $^{13}\text{C}1$ or $^{13}\text{C}4'$ nuclei of one disaccharide from those of other disaccharides, accurately. Therefore, explicit computation of the carbon shielding values for each disaccharide separately, is needed.

Analysis of the α/β anomers

An anomer is one of two stereoisomers of a cyclic saccharide that differs only in its configuration at the hemiacetal carbon ($^{13}\text{C}1$). In hemiacetals, the anomers are designated as α or β according to the configurational relationship between the hydroxyl group ($-\text{OH}$) of the anomeric center ($^{13}\text{C}1$) and the reference atom ($^{13}\text{C}5$) of the anomer. Thus, the α anomer

is defined as the hydroxyl group of the anomeric center ($^{13}\text{C}1$) being on the opposite side of the ring as the reference atom ($^{13}\text{C}5$); while the β anomer is defined as the hydroxyl group of the anomeric center ($^{13}\text{C}1$) being on the same side of the ring as the reference atom ($^{13}\text{C}5$).

To determine the influence of the anomeric state on the $^{13}\text{C}1$ shielding values, we analyzed the coefficient of determination (R^2) between the anomeric $^{13}\text{C}1$ shielding values computed for two arbitrarily selected disaccharides, namely α -D-Glcp-(1-4)- α -D-Glcp (α -maltose, generally called maltose) and β -D-Glcp-(1-4)- β -D-Glcp (cellobiose), using 581 conformations of each disaccharide. The result ($R^2=0.033$), shown in Figure 8a, indicates that an accurate prediction of shieldings will require the explicit computation of the carbon shielding values for each anomeric state separately.

As an additional analysis to test the influence of the distant-anomeric $^{13}\text{C}1'$ nucleus (Fig. 1) on the shielding of the $^{13}\text{C}1$ nucleus, the shieldings for the following, arbitrarily selected, disaccharides were computed: α -D-Glcp-(1-4)- α -D-Glcp (maltose) and α -D-Glcp-(1-4)- β -D-Glcp (β -maltose). The results shown in Figure 8b, in terms of the coefficient of determination R^2 (0.977), indicate that the state (α/β) of the distant-anomeric $^{13}\text{C}1'$ nucleus does not influence the shielding of the $^{13}\text{C}1$ nucleus. In other words, the distant-anomeric $^{13}\text{C}1'$ nucleus is too far away to have any significant influence on the shielding of the carbons participating in the glycosidic link.

Influence of the hydroxyl-group rotamers on the computed shielding of $^{13}\text{C}1$

Hydroxyl groups close to the carbons that participate in the glycosidic linkage should have an effect on the shielding of those carbons, for example, as occurs for serine in proteins, where the rotation of the hydroxyl group, two bonds from the $^{13}\text{C}^\alpha$ nucleus, has a significant effect on the computation of the shielding.^[30]

To determine whether rotation of the hydroxyl groups attached to the various carbon atoms influences the shielding value of the carbons that participate in the glycosidic linkage, the following test was performed for maltose. Once the torsional angles of the glycosidic linkage were fixed, for example, $\varphi = \psi = 0^\circ$, each hydroxyl group in one of the monomers (the one at the nonreducing end) was allowed to adopt three rotameric states, namely $+60^\circ$, -60° , and 180° . The results of this analysis indicate that rotation of the hydroxyl group could have a significant effect on the computation of the shielding of the nearby carbon atom. The influence of the rotating hydroxyl group decreases with the distance measured as the number of bonds around the ring. Thus, the computed average shielding and standard deviation values for the $^{13}\text{C}1$ nucleus are: 2.9 ± 1.2 , 0.4 ± 0.2 , 0.2 ± 0.2 , and 0.3 ± 0.1 ppm, following the rotations of the hydroxyl groups attached to the $^{13}\text{C}2$, $^{13}\text{C}3$, $^{13}\text{C}4$, and $^{13}\text{C}6$ nucleus, respectively. As a result, considerations of the torsional angle variations of the hydroxyl groups attached to the first neighbors of the carbons that participate in the glycosidic linkage are crucial; in fact, their omission will introduce an error of $\sim 3 \pm 1$ ppm in the computations of the shieldings of the carbons that participate in the glycosidic linkage.

The above conclusion is in line with the basis set analysis shown in Figure 4. The BSD [04] treats the hydroxyl group attached to the $^{13}\text{C}2$ nucleus with a 6-311+G(2d,p) basis set (Fig. 4e), while the BSD [05] differs from the BSD [04] only by the basis set used to treat the

same hydroxyl group, namely with a 3-21G basis set (Fig. 4f). Comparison of the coefficients of determination between these two BSD's, that is, $R^2=0.997$ and 0.990 for the BSD [04] and [05], respectively, reveals the influence of the hydroxyl group on the computed shielding value of the $^{13}\text{C1}$ nucleus of maltose.

The fact that the torsional angle variations of the hydroxyl groups, attached to the first neighbors of the carbons that participate in the glycosidic linkage, are crucial for an accurate computation of the shieldings, enables us to estimate the total number of conformations needed to represent the accessible conformational space of a given disaccharide. Thus, the computation of the $^{13}\text{C1}$ and $^{13}\text{C3'}$ shieldings for $\alpha\text{-D-Glcp-(1-3)-}\alpha\text{-D-Glcp}$ will require sampling of φ and ψ every 10° , and of the torsional angles of the hydroxyl groups attached to C2, C2', and C4' every 120° . This implies that a total number of $\sim 35,000$ conformations must be generated. Even though numerous conformations ($\sim 40\%$) will be removed, because they are high-energy conformations, the remaining total number of conformations ($\sim 21,000$) for which the shielding must be computed, at DFT-level of theory, is enormous.

Sensitivity of the shielding of the carbon probe to changes in the φ , ψ torsional angles

As to whether the use of the carbon probes shielding are sensitive enough for validation and refinement of glycan structures deserve to be analyzed. In this regards, on Figures 9a and 9b we show, for the $^{13}\text{C1}$ and $^{13}\text{C4'}$ nuclei, respectively, the equal-shielding surfaces plot on a Ramachandran map using 581 conformations of maltose (generated as described in Materials and Methods section). As shown in Figure 9a, the $^{13}\text{C1}$ shielding are very sensitive to the φ , ψ torsional angle variations, for example, for a fixed $\psi=0^\circ$, a change of 10° step of the φ torsional angles, namely from -50° to -40° , would lead to a shielding change of 1.87 ppm. An additional graphical analysis for four different disaccharides, namely for $\alpha\text{-D-Glcp-(1-4)-}\alpha\text{-D-Xxxp}$, where $\text{Xxx}=\text{Gal}$ (Galactose), Alt (Altrose), Ido (Idose), and Tal (Talose), shown in Supporting Information Figures S3 and S4, lead to similar conclusions.

Overall, variations of the (φ , ψ) torsional angles every 10° would lead to changes, on average, larger than 1.0 ppm in the computed shielding values of the carbon probes.

Conclusions

After a systematic analysis of all the factors affecting the computation of the ^{13}C shielding in disaccharides, our results indicate that: (i) the carbons that participate in the glycosidic linkage are always the best probes with which to sense torsional angle variations of the glycosidic link of disaccharides; (ii) the rotameric states of the hydroxyl groups closer to the glycosidic linkage have a significant effect on the ^{13}C shieldings of the carbons participating in the glycosidic linkage; (iii) the shielding values of the carbons participating in the glycosidic linkage for a given disaccharide cannot be accurately reproduced by other disaccharides; (iv) the anomeric (α/β) state should also be considered explicitly for an accurate computation of the shielding of the carbons that participate in the glycosidic linkage; (v) a fast and accurate prediction of the computed ^{13}C shielding values can be obtained using the OB98 functional with a BSD [04]; and (vi) a step of 10° , rather than 20° ,

should be used to sample the (ϕ , ψ) torsional angles, otherwise the ^{13}C shielding interpolations will not be accurate enough.

Taking into account all the factors affecting the computation of the ^{13}C shielding, listed above, building an accurate method for validation, determination, and refinement of glycan structures will require that we compute the ^{13}C shielding for a large number (up to ~21,000) of low-energy conformations for each disaccharide, and consideration of as many as possible disaccharides. In this regard, our future work will focus on building a look-up table of shieldings of the carbon participating in the glycosidic linkage. The resulting look-up table will contain, among others, the shielding-values for the following disaccharides: maltose, cellobiose, lactose, trehalose, saccharose, isomaltose, and disaccharides forming complex molecules such as galactooligosaccharides, peptidoglycans, proteoglycans, gangliosides, and so forth. Once the look-up tables are built, we will then be in condition to validate, refine, and determine glycan structures, if the chemical shifts of the carbon probe are available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Contract grant sponsor: National Institutes of Health; contract grant number: GM14312 (H.A.S.); Contract grant sponsor: IMASL-CONICET; contract grant number: PIP-112–2011-0100030 (J.A.V.); Contract grant sponsor: UNSL; contract grant number: Project 328402 (J.A.V.)

The research was conducted using the resources of Blacklight, a facility of the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center.

References

1. Apweiler R, Hermjakob H, Sharon N. *Biochim Biophys Acta*. 1999; 1473:4. [PubMed: 10580125]
2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. *Nucleic Acids Res*. 2000; 28:235. [PubMed: 10592235]
3. Lütke T. *Acta Crystallogr D*. 2009; 65:156. [PubMed: 19171971]
4. Vila JA, Arnautova YA, Martin OA, Scheraga HA. *Proc Natl Acad Sci USA*. 2009; 106:16972. [PubMed: 19805131]
5. Martin OA, Vila JA, Scheraga HA. *Bioinformatics*. 2012; 28:1538. [PubMed: 22495749]
6. Han B, Ginzinger SW, Liu Y, Wishart DS. *J Biomol NMR*. 2011; 50:43. [PubMed: 21448735]
7. Saitô H. *Magn Reson Chem*. 1986; 24:835.
8. Jarvis MC. *Carbohydr Res*. 1994; 259:311. [PubMed: 8050104]
9. Swalina CW, Zauhar RJ, DeGrazia MJ, Moyna G. *J Biomol NMR*. 2001; 21:49. [PubMed: 11693568]
10. Durran DM, Howlin BJ, Webb GA, Gidley MJ. *Carbohydr Res*. 1995; 271:C1.
11. Wilson PJ, Howlin BJ, Webb GA. *J Mol Struct*. 1996; 385:185.
12. Hricovíni M, Malkina OL, Bízik F, Nagy LT, Malkin VG. *J Phys Chem A*. 1997; 101:9756.
13. Zhang P, Klymachyov A, Brown S, Ellington J, Grandinetti P. *Solid State Nucl Magn Reson*. 1998; 12:221. [PubMed: 9800267]
14. Sergeyev I, Moyna G. *Carbohydr Res*. 2005; 340:1165. [PubMed: 15797132]
15. McNaught AD. *Pure Appl Chem*. 1996; 69:1919.

16. Bohne A, Lang E, von der Lieth CW. *Bioinformatics*. 1999; 15:767. [PubMed: 10498779]
17. IUPAC-IUB Joint Commission on Biochemical Nomenclature. *Eur J Biochem*. 1980; 111:295. [PubMed: 7460897]
18. Chu SSC, Jeffrey GA. *Acta Crystallogr*. 1968; 24:830.
19. Wyss DF, Wagner G. *Curr Opin Biotechnol*. 1996; 7:409. [PubMed: 8768899]
20. Allinger NL, Yuh YH, Lii JH. *J Am Chem*. 1989; 111:8566.
21. Schrödinger, LLC. In *PyMOL Molecular Graphics System*. San Diego, CA: 2013.
22. Wolinski K, Hinton JF, Pulay P. *J Am Chem Soc*. 1990; 112:8251.
23. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA.; Vreven, JT.; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, JE.; Hratchian, HP. In *Gaussian 03, Revision E.01*. Inc; Wallingford CT: 2004.
24. Vila JA, Serrano P, Wüthrich K, Scheraga HA. *J Biomol NMR*. 2010; 48:23. [PubMed: 20644980]
25. Chesnut DB, Moore KD. *J Comput Chem*. 2009:648.
26. Hehre, WJ.; Radom, L.; Schleyer, P.; Pople, JA. *Ab Initio Molecular Orbital Theory*. Wiley; New York: 1986.
27. Facelli JC. *J Phys Chem B*. 1998; 102:2111.
28. Ferraro MB. *J Mol Struct Theochem*. 2000; 528:199.
29. Bagno A. *Chem Eur J*. 2001; 7:1652. [PubMed: 11349906]
30. Vila JA, Baldoni HA, Scheraga HA. *J Comput Chem*. 2009; 30:884. [PubMed: 18780343]

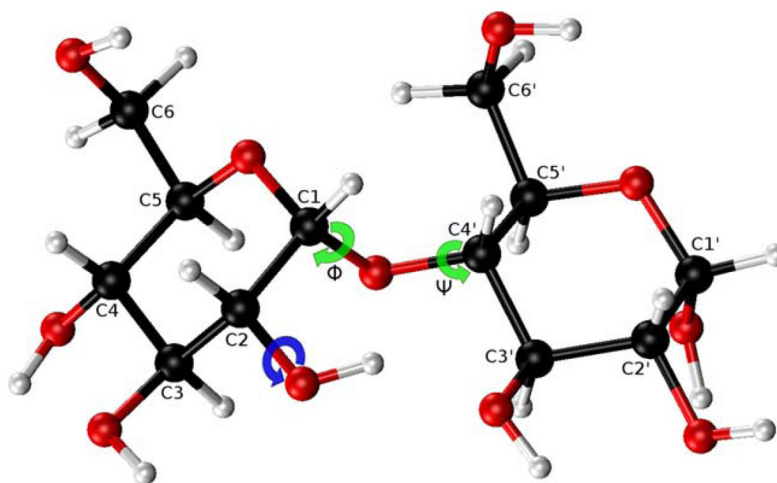


Figure 1.

Ball and stick representation of the maltose disaccharide [α -D-Glcp-(1-4)- α -D-Glcp] in an arbitrary conformation and with the pyranose ring in the 4C_1 conformation of the C (chair) puckering. Each carbon atom of the disaccharide (in black) is labeled following the IUPAC recommendation.^[15] Red and gray colors are used to highlight the oxygen and hydrogen atoms, respectively. The torsional angles (φ , ψ) of the glycosidic link are highlighted in green. A blue arrow illustrates the χ torsional angle (H2-C2-O-H) of the hydroxyl group attached to the C2 nucleus; all other hydroxyl groups in the disaccharide exhibit similar torsional angles.

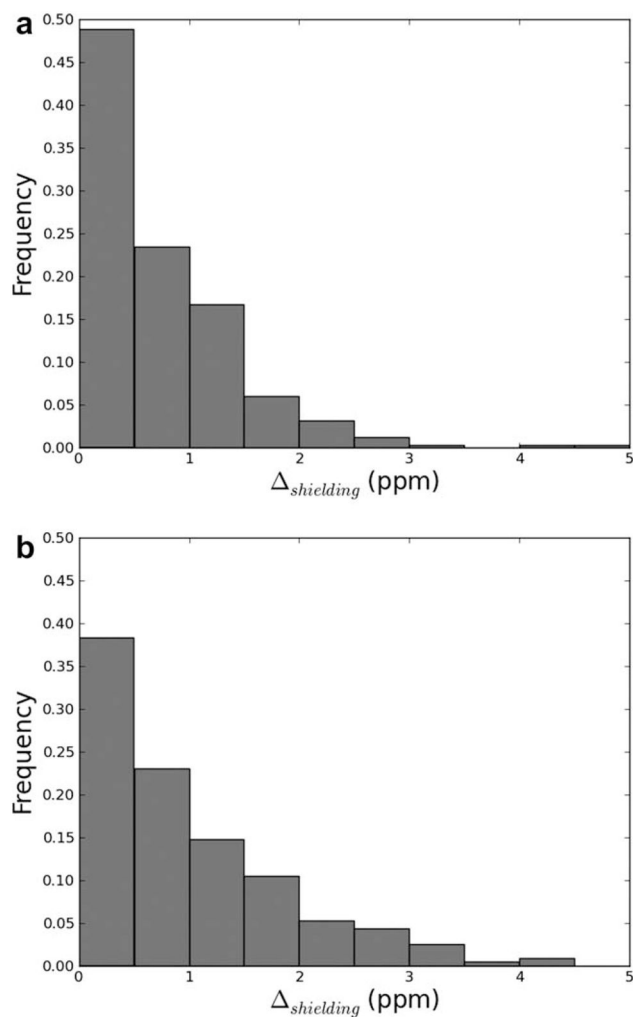


Figure 2.

a) Normalized histogram of the $\Delta_{shielding}$ values ($\Delta_{shielding} = |\delta_{interpolation} - \delta_{true}|$) obtained using a linear interpolation method. The $\Delta_{shielding}$ values are grouped within intervals of 0.5 ppm. b) same as (a) for a cubic interpolation method.

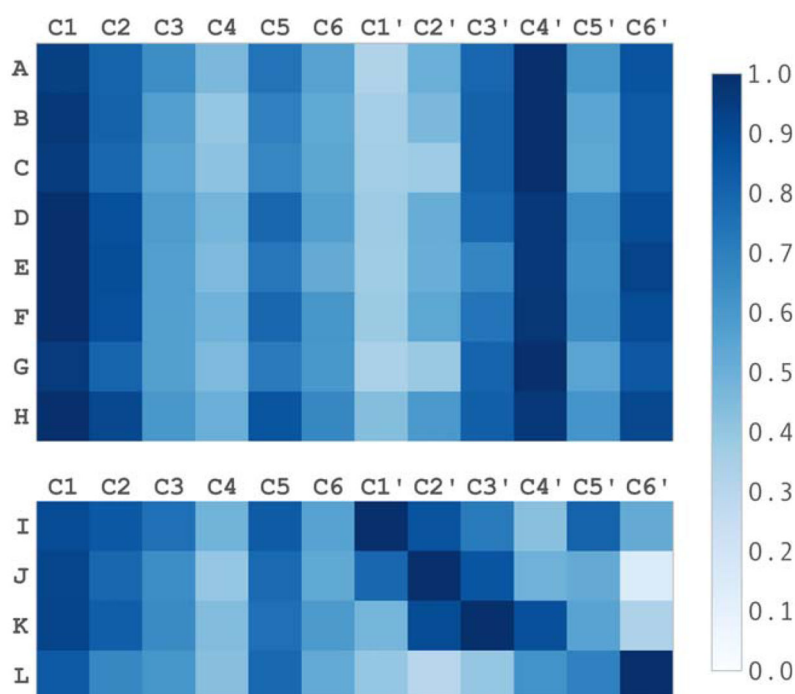


Figure 3.

Normalized entropy, computed using eq. (1), for all the ^{13}C shielding distributions of 12 disaccharides. Each of the 12 rows of the figure corresponds to results obtained by using a particular disaccharide, according to the following taxonomy: **A–H** columns are for eight disaccharides with glycosidic linkage (1–4): $\alpha\text{-D-Glcp-(1-4)-}\alpha\text{-D-}\text{Xxxp}$, where **Xxx** = Glu (Glucose), All (Allose), Alt (Altrose), Gal (Galactose), Gul (Gulose), Ido (Idose), Man (Mannose), and Tal (Talose), and **I–L** columns are for four disaccharides with glycosidic linkage (1-*Y*) where *Y* = 1, 2, 3, 6, $\alpha\text{-D-Glcp-(1-}Y\text{)-}\alpha\text{-D-Glcp}$. The blue-colored scale of the normalized entropy is shown as a separate column.

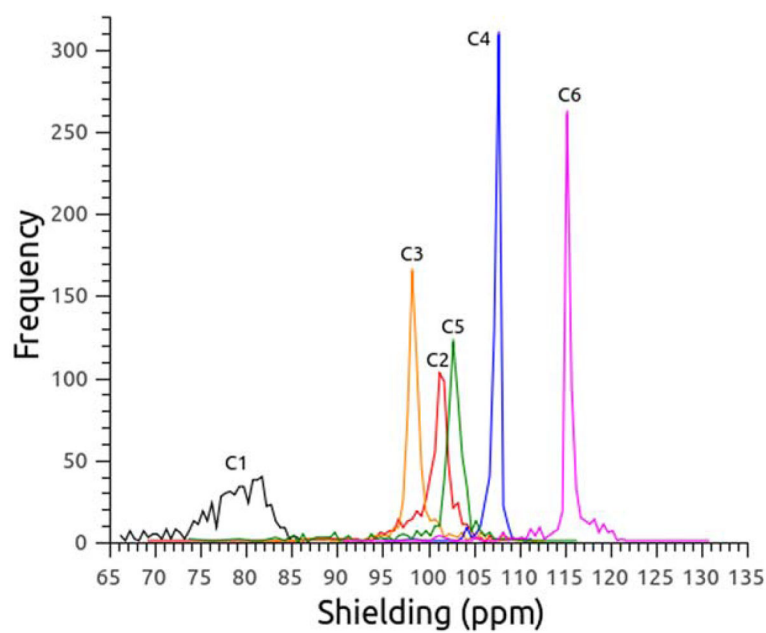
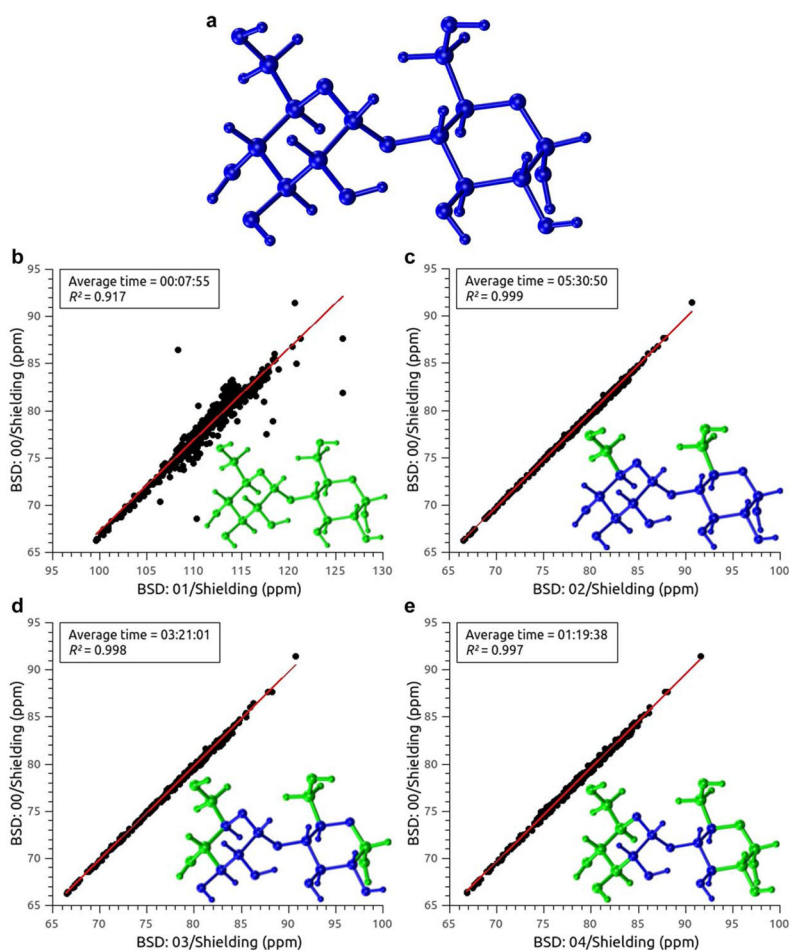


Figure 4. Histograms of the gas-phase, computed ^{13}C shieldings from 581 conformations of maltose [α -D-Glcp-(1 \rightarrow 4)- α -D-Glcp] as described in the Materials and Methods section. The shielding distributions for each of the six carbons of the ring (Fig. 1) are distinguished using different colors. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



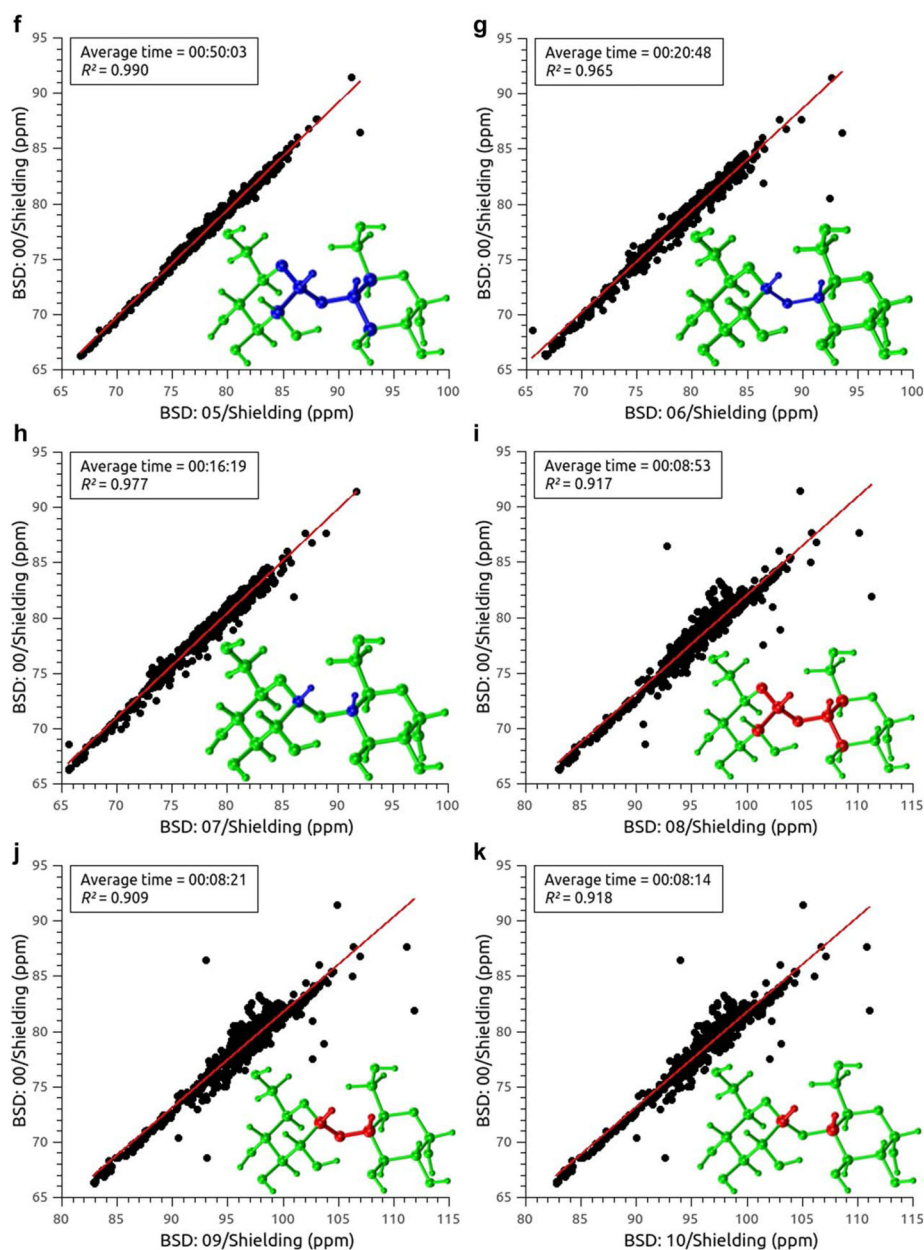
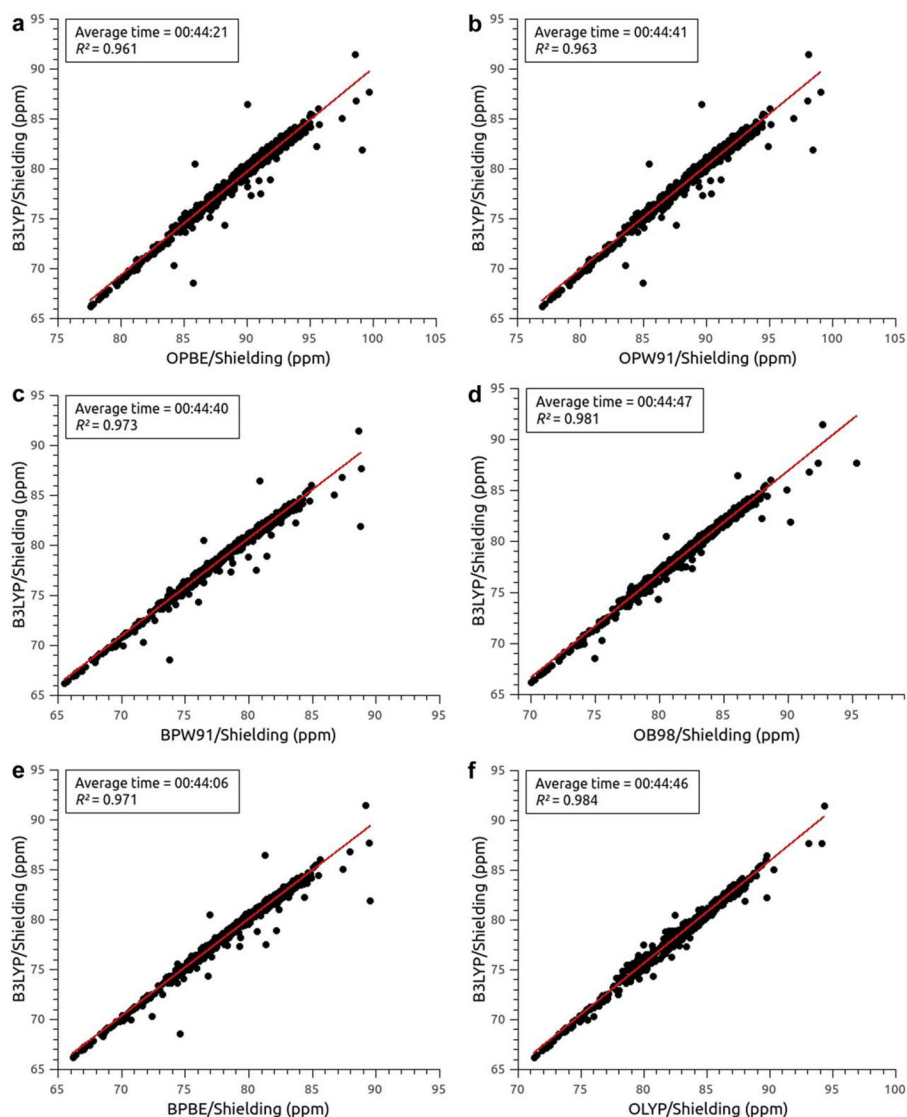


Figure 5.

The maltose structure in panel (a) is colored blue, so as to illustrate a uniform BSD 6–311+G(2d,p) ([00]). Graphical representations of the maltose structure, colored according to the other BSD's used to compute the $^{13}\text{C}1$ shieldings, are inserted in each panel, except panel (a), in blue, red, and green indicating the 6-311+G(2d,p), 6-31G, and 3-21G basis set, respectively; b) Correlation between $^{13}\text{C}1$ shielding computed, for 581 conformations of maltose, using the [00] versus [01] uniform BSD; c), d), e), f), g), h), i), j), and k), same as (b) for the BSD [02], [03], [04], [05], [06], [07], [08], [09], and [10], respectively; for (b)–(k), the averaged CPU-time, in hours:minutes:seconds, and the values of the coefficient of determination, R^2 , are inserted as a panel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

**Figure 6.**

a) Correlation between ^{13}C shielding values computed from 581 maltose conformations with the B3LYP functional and the basis set [00] versus the ^{13}C shielding values computed with the OPBE functional and the BSD [04]; inserted, as a panel; the averaged CPU-time, and the value of the square of the correlation coefficient, R , are shown; b), c), d), e), and f) the same as (a) for the functionals OPW91, BPW91, OB98, BPBE, and OLYP, respectively, with the BSD [04]. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

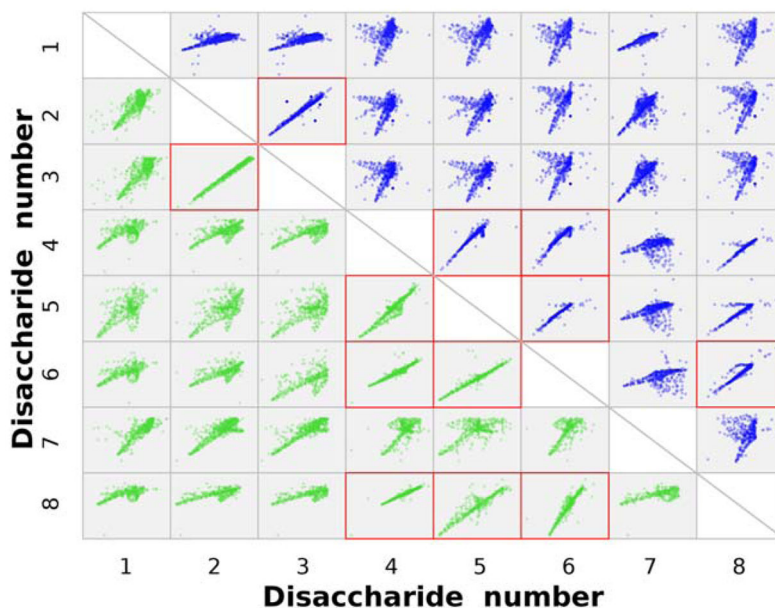


Figure 7.

Scatter matrix displaying the correlation between computed shielding values, of the $^{13}\text{C}1$ and $^{13}\text{C}4'$ nuclei, for eight different (1–4) linked disaccharides, namely for: α -D-Glcp-(1-4)- α -D-**X**xp, where **X**xx =Glucose (1), Allose (2), Altrose (3), Galactose (4), Gulose (5), Idose (6), Mannose (7), and Talose (8); the x and y axes of the figure are labeled with the number identifying each disaccharide. The x and y axes of each square correspond to the computed shielding value for the carbons of the given disaccharide. Each of the 56 square-windows shows the correlations between ^{13}C shieldings, with the 28 square-windows of the upper-triangle for the $^{13}\text{C}1$ nucleus, while the 28 from the lower-triangle are for the $^{13}\text{C}4'$ nucleus. The empty windows on the diagonal represent the correlation of each of the eight disaccharides with themselves, while those pairs of disaccharides for which the coefficient of determination $R^2 > 0.9$ are highlighted by a red-frame. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

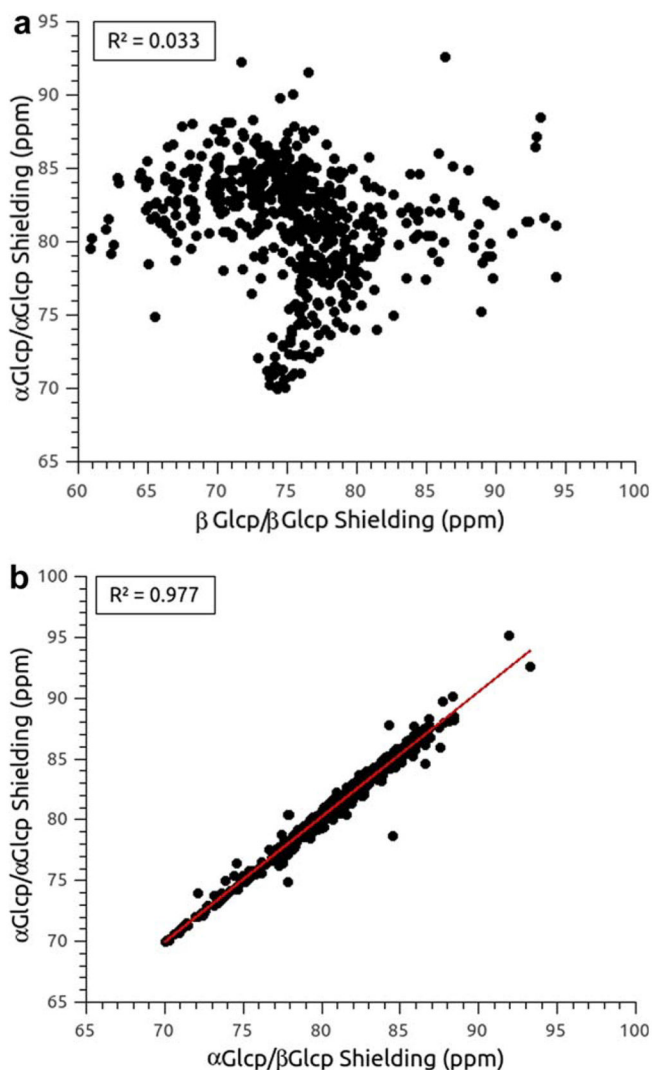


Figure 8.

a) Correlation between $^{13}\text{C}1$ shielding values from the disaccharides $\alpha\text{-D-Glcp-(1-4)-}\alpha\text{-D-Glcp}$ (maltose) and $\beta\text{-D-Glcp-(1-4)-}\beta\text{-D-Glcp}$ (cellobiose), using 581 conformations of each disaccharide; b) same as (a) for the $\alpha\text{-D-Glcp-(1-4)-}\alpha\text{-D-Glcp}$ (maltose) and $\alpha\text{-D-Glcp-(1-4)-}\beta\text{-D-Glcp}$ (β -maltose). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

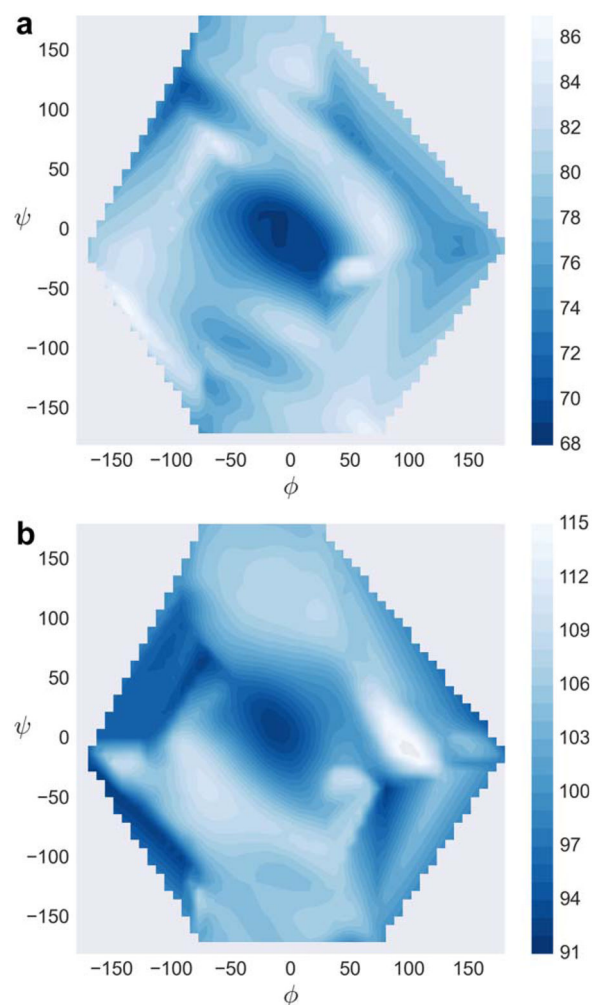


Figure 9.

a) Contour plot for variation of the $^{13}\text{C1}$ shielding of the maltose molecule based on 581 conformations, generated as described in the Materials and Methods section. The equal-shieldings surfaces, every 1.0 ppm, are represented using different shades of blue. Gray color is used to highlight the nonexplored regions of the Ramachandran map. The column on the right-hand represents the range, in ppm, of the shielding variation; and b) same as (a) for the $^{13}\text{C4'}$ shielding. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]