Ab Initio Conformational Space Study of Model Compounds of O-Glycosides of Serine Diamide

Gábor I. Csonka,*[a] Gábor A. Schubert,[a] András Perczel,[b] Carlos P. Sosa,[c] and Imre G. Csizmadia[d]

Abstract: Relative stabilities of rotamers of the *N*-acetyl-*O*-(2-acetamido-2-deoxy-α-D-galactopyranosyl)-L-seryl-*N*′-methyl amide (**1**) and eleven analogous molecules containing β-galactose, α- and β-mannose, α- and β-glucose, and L-threonine were calculated to learn whether they could explain the natural preference for **1** in linkages between the carbohydrate and protein in glycoproteins. The lowest energy rotamers of four O-glycoside models of serine diamide were identified with a Monte Carlo search coupled with molecular mechanics (MM2*). These rotamers were further optimized with an ab initio level of theory (HF/6-31G(d)). Subsequently, B3LYP/6-31+G(d) single point energies were calculated for the most stable HF structures. The most favorable interactions are present in **1** and its glucose analogue. The monosaccharide for the carbohydrate antenna is anchored to the serine residue with an AcNH$\cdots$O=C-NHMe hydrogen bond in the most stable rotamers. The mannose analogue and the β-anomers are considerably less stable according to the MM2* and especially to the ab inito energy values. The three analogues have HF/6-31G(d) energies which are 4–6 kcal mol^{-1} higher; the single point B3LYP/6-31+G(d)//HF/6-31G(d) calculations yield preferences of 3–5 kcal mol^{-1} for **1**. The most stable L-threonine analogues show a behaviour very similarly to the corresponding serine analogues. The ZPE and thermal correction components of the calculated ΔH_{298} and ΔG_{298} values are relatively small (<0.4 kcal mol^{-1}). However, the $T\Delta S_{298}$ term can be as large as 2.6 kcal mol^{-1}. The entropy terms stabilize the α-anomers relative to β-anomers, and ManNAc relative to GalNAc. The largest stabilization effect is observed for one of the rotamers of the α-anomer of ManNAc.

Keywords: computer chemistry · density functional calculations · glycopeptides · Gibbs free energies · heats of formation

Introduction

The natural proteins frequently contain saccharide side chains of variable length. Such proteins are termed glycoproteins. The saccharide residues show a considerable influence on the conformational and physicochemical properties of the proteins; they are important for biological recognition processes, cellular adhesion (binding), and provide coating (e.g. protection from proteases or antibodies). These structures are not coded in the DNA, in fact their synthesis depends on glycosyl transferases, glycosidases, organelles, and on the availability of monosaccharides. As a consequence glycosylation is very sensitive to subtle changes in the environment. Altered glycosylation can serve as an event trigger (e.g. cell death, attack). A great deal of bioinformatics is related to the antenna-like carbohydrate moieties of glycoproteins, because of the huge diversity of carbohydrate structure that is possible. The structures and functions of the sugar chains in glycoproteins are reviewed in refs. [1, 2]. The saccharide residues are covalently linked to the protein backbone either N- (via asparagine) or O-glycosidically (via serine, threonine, tyrosine, or hydroxylysine). A characteristic feature of core carbohydrates is the linking of the protein backbone with the saccharide side chain. The N-glycosidic linkages involve exclusively β-2-acetamido-2-deoxy-D-glucosyl residues as the first residue of the carbohydrate antenna. In the mucin-type

[a] Prof. G. I. Csonka, G. A. Schubert
Department of Inorganic Chemistry
Budapest University of Technology and Economics
1521 Budapest (Hungary)
E-mail: csonka@web.inc.bme.hu

[b] Prof. A. Perczel
Department of Organic Chemistry
Eötvös Loránd University, 1518 Budapest (Hungary)

[c] Dr. C. P. Sosa
IBM and University of Minnesota Supercomputing Institute
Minneapolis, MN 55415 (USA)

[d] Prof. I. G. Csizmadia
Department of Chemistry, University of Toronto
Toronto, Ontario, M5S 3H6 (Canada)

Supporting information for this article is available from the author: the optimized structures of the conformational spaces are published in the World Wide Web: http://web.inc.bme.hu/mols/galnacser/

0947-6539/02/0820-4718 $ 20.00+.50/0

core carbohydrates the first saccharide is always 2-acetamido-2-deoxy-D-galactose (T_N antigen), which is α-glycosidically linked to the hydroxy group of threonine or serine building the inner part of several characteristic saccharide core chains. So far, eight different O-glycosidic core structures have been identified:[1]

1) βGal-1,3-αGalNAc-1-Ser/Thr (T antigen)
2) βGal-1,3-[βGlcNAc-1,6-]αGalNAc-1-Ser/Thr
3) βGlcNAc-1,3-αGalNAc-1-Ser/Thr
4) βGlcNAc-1,3-[βGlcNAc-1,6-]αGalNAc-1-Ser/Thr
5) αGalNAc-1,3-αGalNAc-1-Ser/Thr
6) βGlcNAc-1,6-αGalNAc-1-Ser/Thr
7) αGalNAc-1,6-αGalNAc-1-Ser/Thr
8) αGal-1,3-αGalNAc-1-Ser/Thr

To these core structures other saccharide units of varying structure and length can be attached consisting of GalNAc, GlcNAc, Gal, Fuc, and NeuNAc. An important feature of the exterior facade of the antennae is their microheterogeneity, in contrast to the interior or core portion. The heteropolysaccharide antennae of blood types A and B are typical examples where the structural difference is located only at the most external carbohydrate residue: The constitutional difference (α-D-GalNAc versus α-D-Gal) serves a signaling purpose. It is also known that he occurrence of aberrant carbohydrate side chains in natural glycoproteins often parallels pathological phenomena. An important example for one of tumor-associated carbohydrate antigens is the sialyl-T_N antigen (αNeu-2,6-αGalNAc-1-*O*-Ser/Thr). Synthetic fragments are being tested as vaccines against cancer.[3]

Although the above "constitutional conservatism" of the glycosidic linkage represents about 99% of all cases (cf. database of O-linked glycosylation sites of glycoproteins),[4] there are a few exceptions for O-glycosides: hydroxylysine or hydroxyproline connected to D-galactose, L-fucose α-linked to serine or threonine (e.g. in human factor IX, in an insect neuropeptide, and in epidermal growth factor (EGF) domains of various coagulation and fibrinolytic proteins),[1] or cysteine linked to D-glucose through an S-glycosidic bond.[5] O-Glycopeptides carrying βGlcNAc (*O*-2-acetamido-glucopyranose) as unmodified monosaccharide side chain linked to serine or threonine are found predominantly in nucleoplasmic and cytoplasmic cell compartments.[6] Important biological aspects of *O*-GlcNAc glycosylation (e.g., Alzheimer's disease) are discussed in the literature.[1] Biological investigations revealed that a glycopeptide from type II collagen with a centrally located β-Gal-Hyl (hydroxylysine) structure was recognized by the majority of the autoimmune T cells obtained in a mouse model for rheumatoid arthritis.[7] αMan-Ser/Thr linkage can be found in the phytoalexin elicitoractive glycoprotein. The same linkage was found in the saccharide moiety αNeuNAc-2,3-βGal-1,4-βGlcNAc-1,2-αMan-Ser/Thr that is the major constituent of the O-linked carbohydrates of α/β-dystroglycan complex and contributes to laminin binding.[1] βGlc-Ser linkage was found in blood clotting factor IX (a plasma glycoprotein) which is involved in the blood coagulation cascade. βGlcNAc-Thr linkage can be found in O-linked sialyl-Lewis-X (sLex) tetrasaccharide (as part of the mucin domain of mucosal addressing cell adhesion molecule-1 (MAd-CAM-1), which is a ligand of P- and L-selectin).[1]

The exceptions rarely occur in glycoproteins (e.g. tyrosine residues carrying saccharide side chains are seldom found in nature) but are more typical for proteoglycans, providing nice examples of microheterogeneity. The variability observed in the external facade is not observed at the root of the antenna where the constitution and anomeric configuration of the monosaccharide residue is rigorously preserved.

One may wonder why the core part of the antenna and especially the first carbohydrate residue is preserved so strictly during posttranslational modification. Why is there an overwhelming preference for αGalNAc-1-*O*-Ser/Thr bond, and why is the β linkage typical for N-glycosides? Recent ^{1}H NMR spectroscopic results suggest that α-O-linked GalNAc causes a dramatic alteration in the structure of the peptide backbone.[8] However, the β-linked isomer showed rather small changes in amide chemical shifts relative to the α-isomer. The structural change of the α-linked glycopeptides manifests in the significant increase in the in the lifetime of exchangeable peptide backbone amide protons relative to free peptide. The exchange lifetime of the NH of a GalNAc residue is very sensitive to anomeric stereochemistry (more than 12 h for α and minutes for β).[8] The following conclusion were drawn by Danishefsky et al.: "It is likely that the acetyl group on the GalNAc residue is also necessary to support structural coherence. Furthermore, installation of the initial α-*O*-GalNAc residue in a cluster domain creates a stable

Abstract in Hungarian: *Az N-acetil-O-(2-acetamido-2-deoxi-α-D-galactopiranozil)-L-szeril-N′-metilamid (**1**) és e molekula 11 analógjának (β-galaktóz, valamint α- és β-mannóz, glükóz és L-treonin) relatív energiáit határoztuk meg, hogy ezzel magyarázatot kapjunk arra, hogy a természetben miért **1** fordul elő leggyakrabban a glikoproteinekben, mint a szénhidrátot és a fehérjét összekapcsoló egység. Az szerin diamid O-glikozid modelljeinek legstabilabb rotamerjeit Monte-Carlo kereséssel kombinált molekulamechanikai számítással (MM2*) határoztuk meg. Ezt követően a rotamerek geometriáját ab initio szinten (HF/6-31G(d)) további optimáltuk. A legstabilabb HF szerkezetek energiáit B3LYP/6-31+G(d) szinten is meghatároztuk. A legkedvezőbb kölcsönhatás **1**-ben figyelhető meg. E molekula legstabilabb rotamerjeiben a monoszacharidhoz AcNH ··· O=C-NHMe típusú hidrogénhidas kötéssel kapcsolódik a szénhidrát antenna. A három másik analóg jóval kevésbé stabilabb az MM2* számítások szerint, és ez a különbség még nagyobb az ab initio energiák esetében. A három másik analógnál az energiák 4–6 kcal mol^{-1}-lal magasabbak a HF/6-31G(d) számítások szerint. A B3LYP/6-31+G(d)//HF/6-31G(d) számítások alapján pedig 3–5 kcal mol^{-1}-lal stabilabb az **1**-es molekula a többi analógnál. A ΔH_{298} és ΔG_{298} értékekhez tartozó zéruspont és termikus energia korrekciók viszonylag kicsik (< 0.4 kcal mol^{-1}), viszont a $T\Delta S_{298}$ kifejezés értéke akár 2.6 kcal mol^{-1} is lehet. Az entrópia tag az α-anomert stabilizálja a β-hoz képest, valamint a ManNAc részt tartalmazó molekulákat a GalNAc részt tartalmazó molekulákhoz képest. A legnagyobb stabilizációs hatás az egyik ManNAc részt tartalmazó α-anomerben figyelhető meg.*

  0947-6539/02/0820-4719 $ 20.00+.50/0

scaffold that can accept, without intrinsic change, increased glycosylation allowing the display of antennary glycans. A variety of carbohydrate structures can be accommodated in this way so that the same protein backbone can display a variety of glycans, the nature of which reflect the physiological state of the cell."[8] It was also observed for asparagin-linked glycosylation that the N-acetyl groups of the carbohydrates have critical role in promoting the more compact β-turn conformation through steric interaction with the peptide, and small changes in carbohydrate composition can have large effect on glycopeptide conformation.[9]

In the present paper we focus on the distinct configurations in the vicinity of the glycosidic linkage. In this respect the two possible orientations (α or β) of the anomeric center, C1, and the equatorial or axial orientation of the acetamido group of C2 are critical. This leads to four possible conformations (α_{ax}, α_{eq}, β_{ax}, β_{eq}) which fundamentally influence the orientation of the carbohydrate antenna. Thus 2-acetamido-2-deoxy-D-mannopyranose (ManNAc with an axial NAc group) represents an alternative to GalNAc. It is interesting that ManNAc, though frequently found in other locations of the antenna, never occurs in the O-glycosidic linkage to Ser/Thr. It is expected that the axial or equatorial orientation of the other hydroxyl groups, that is the O4H on the C4 atom is considerably less important and that GalNAc and GlcNAc behave similarly in this respect. However, for comparison purposes, the results for GlcNAc derivatives are also presented in this paper.

Hydrogen bonds play a central and undoubtedly a major role in the structure and energetics of biopolymers.[10] They are involved in the formation of DNA pairing and in the stabilization of protein secondary structural elements (e.g. helices, parallel and antiparallel sheets, and turns). In glycoproteins, both the carbohydrate antenna and the protein parts of the macromolecule incorporate a large number of hydrogen bonds in a well-established manner. However, it is not yet clear what anchors the antenna to the surface of the protein. Both hydrophobic contacts (e.g. between fucose and apolar side chains of selected amino acid residues) and hydrophilic interactions such as hydrogen bonds could be crucial.

We will demonstrate in this paper with the help of ab initio conformational analysis that the essential hydrogen bond(s) can only form if the above-mentioned constitution and configuration are conserved. This theoretical work, along with experimental studies, could provide some explanation as to why mutation and alteration occur rarely in the linker region of glycoproteins.

Scope

Four model O-glycosides are necessary and sufficient to understand the energetic a conformational difference between the four possible epimers at the linking region (C1 and C2 epimers, cf. Figure 1). In this respect α,β-D-galactose and α,β-D-mannose derivatives would be sufficient. However, for comparison reasons we also include α,β-D-glucose derivatives. All compounds contain the N- and C-terminated L-serine or threonine residue and the 2-acetamido-carbohydrate residue. Similar model systems were experimentally investigated earlier on in our laboratory.[11, 12] We selected these compounds because of the relatively small conformational space (10 variable torsional angles), and because the ab initio conformational space of N- and C-protected serine was recently published. The numbering, the abbreviations of our model molecules are shown in Figure 1.

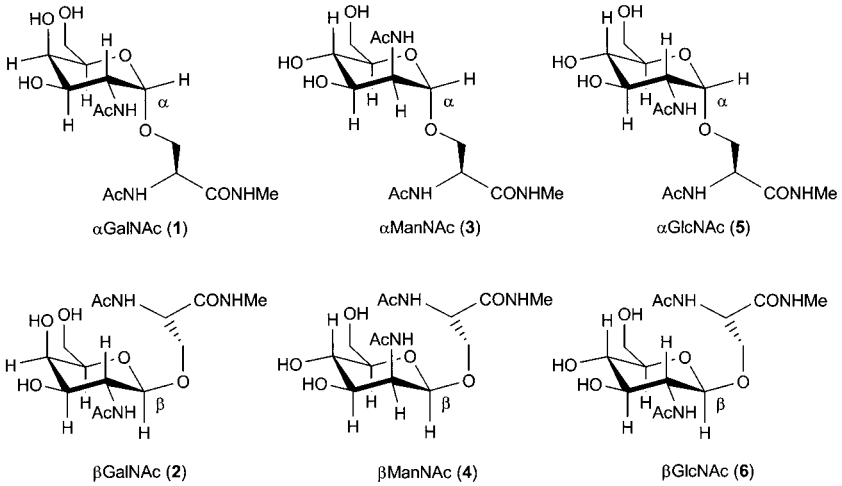


Figure 1. Six different O-glycoside model systems with systematically varied configurations of C1, C2, and C4 of the pyranose ring: *N*-acetyl-*O*-(2-acetamido-2-deoxy-α-D-galactopyranosyl)-L-seryl-*N*′-methy amide (**1**), *N*-acetyl-*O*-(2-acetamido-2-deoxy-β-D-galactopyranosyl)-L-seryl-*N*′-methyl amide (**2**), *N*-acetyl-*O*-(2-acetamido-2-deoxy-α-D-mannopyranosyl)-L-seryl-*N*′-methyl amide (**3**), *N*-acetyl-*O*-(2-acetamido-2-deoxy-β-D-mannopyranosyl)-L-seryl-*N*′-methyl amide (**4**), *N*-acetyl-*O*-(2-acetamido-2-deoxy-α-D-glucopyranosyl)-L-seryl-*N*′-methyl amide (**5**), *N*-acetyl-*O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-L-seryl-*N*′-methyl amide (**6**).

The numbering of the atoms and notation for the most important torsional angles of αGalNAcSer (**1**) are given in Figure 2. The same notation is used for the other molecules. These structures are C1, C2 and C4 epimers. The other structures can be derived from αGalNAcSer by simply changing the corresponding axial (ax) or equatorial (eq) positions either on C1 (ax in the α, and eq in the β-anomer), on C2 (eq in galactose, or glucose and ax in mannose), and on C4 (ax in galactose, and eq in mannose, or glucose) of the carbohydrate residue.

We also study six L-threonine analogues of these molecules: αGalNAcThr (**7**), βGalNAcThr (**8**), αManNAcThr (**9**), βManNAcThr (**10**), αGlcNAcThr (**11**) and βGlcNAcThr (**12**).

Methods

MM Methods: The search for stable conformers in the conformational space of the selected molecules **1**–**4** was carried out using the Macro-

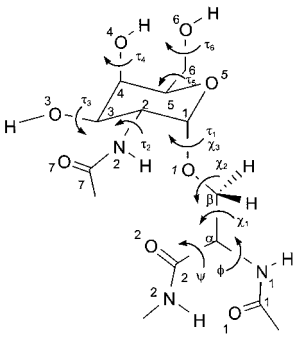
0947-6539/02/0820-4720 $ 20.00+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 20

Model 5.0 program package.[13] The MM2* force field in MacroModel, a slightly varied version of authentic MM2,[14] was used. The most important difference between MM2* and MM2 is in the electrostatic equation. The MM2* force field employs the point-charge Coulomb potential to describe the electronic electrostatic interactions, whereas authentic MM2 uses dipole–dipole interactions. The calculations were performed with variable electrostatic interaction. However, the usual Coulomb equation was also tested (more recent versions of MacroModel use the latter approximation in the MM2* force field). A recent comparison of various molecular mechanics methods showed that the accuracies of relative conformational energies are apparently equal for MM2*, MM2(91) and MM3(92).[15] We have also tested the more recent MMFF94[16] force field.

The conformational searches were carried out with the systematic unbounded multiple minimum search technique (SUMM)[17] which is available in MacroModel. The searches were based on the most stable $^4C_1$ form of the pyranose ring and was limited to only the various rotamers of the rotatable exocyclic groups. The 2000 structures generated by the SUMM procedure were minimized with MM2* to yield unique conformers within an energy window of 5.0 kcal mol^{-1} above the global minimum. An additional conformational search was started from the resulting global minimum geometry, limited to 1500 structures. The elements of the two resulting conformational spaces were compared. It was found that the first SUMM search was sufficient and no new low energy rotamers were found during the latter search. Geometry optimizations were carried out with truncated Newton conjugate gradient (TCNG) technique, with the maximum number of iterations set to 200 by using a convergence criterion of 0.01 for the gradient norm.

Ab initio methods: The lowest energy structures obtained by the MM2*-SUMM searches were fully optimized at the HF/6-31G(d) level of theory. All the optimizations were performed in redundant internal coordinates using the Berny algorithm[18] built into the Gaussian98[19] program. Vibrational frequencies were calculated for the HF/6-31G(d) optimized geometries to confirm that each stationary point is a true minimum on the potential energy surface. The calculated zero-point vibration energy (ZPE) corrections were scaled by 0.8929. ZPE+thermal corrections+RT were used to calculate enthalpies (H at 298.15 K) and Gibbs free energies ($G = H - TS$ at 298.15 K). Single point energy calculations which include correlation energy were performed at B3LYP/6-31+G(d)//HF/6-31G(d) level of theory, using tight convergence criteria.

Multidimensional conformation analysis, notations: Figure 2 shows the torsional angles of the conformational space of the model compounds. The conformational space of an O-glycoside model of serine diamide molecule can be divided into three distinct parts related to the serine backbone, the serine side chain, and the carbohydrate residue. Such division helps the classification and the discussion of the various rotamers. The ψ and Φ torsional angles characterize the serine backbone conformations (cf. Figure 2). Figure 3 shows the notations used for the possibly stable rotamers of the conformational space of ψ and Φ. The stable rotamers are denoted by Greek letters (α, β, γ, δ, ε) and four of these five rotamers have two alternatives (L or D in Figure 3). For example the γ_L rotamer in Figure 3 can be characterized with $\psi \approx 60^\circ$ and $\Phi \approx -60^\circ$, the γ_D rotamer with $\psi \approx -60^\circ$ and $\Phi \approx 60^\circ$. The serine side chain has two torsional angles denoted by χ_1 and χ_2 as depicted in Figure 2. χ_3 denotes the torsional angle



ϕ: $C1_{Ser}$-$N1_{Ser}$-C^{α}-$C2_{Ser}$
ψ: $N1_{Ser}$-C^{α}-$C2_{Ser}$-$N2_{Ser}$
χ_1: $N1_{Ser}$-C^{α}-C^{β}-$O1_{Carb}$
χ_2: C^{α}-C^{β}-$O1_{Carb}$-$C1_{Carb}$
χ_3: C^{β}-$O1_{Carb}$-$C1_{Carb}$-$C2_{Carb}$
τ_1: C^{β}-$O1_{Carb}$-$C1_{Carb}$-$O5_{Carb}$
τ_2: $C1_{Carb}$-$C2_{Carb}$-$N2_{Carb}$-$C7_{Carb}$
τ_3: $C4_{Carb}$-$C3_{Carb}$-$O3_{Carb}$-H
τ_4: $C5_{Carb}$-$C4_{Carb}$-$O4_{Carb}$-H
τ_5: $O5_{Carb}$-$C5_{Carb}$-$C6_{Carb}$-$O6_{Carb}$
τ_6: $C5_{Carb}$-$C6_{Carb}$-$O6_{Carb}$-H

Figure 2. Definitions of the torsional angles and of the numbering of the atoms of αGalNAcSer (**1**). The same definitions were used for **2**–**12**. Carb represents the carbohydrate residue.

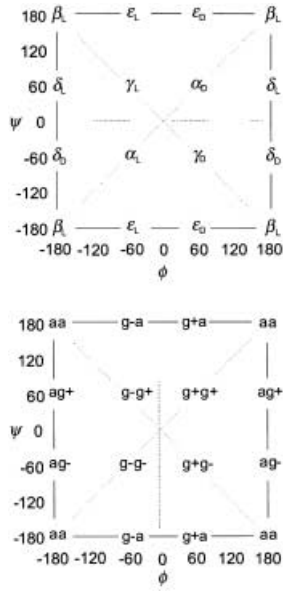


Figure 3. Labeling Schemes using the IUPAC-IUB recommendation for torsional angle definitions for amino acid residue rotamers located on the Ramachandran potential energy surface.

of the O-glycosidic bond (cf. Figure 2). These five torsional angles provide $3^5 = 243$ possible stable rotamers. The carbohydrate residue can be characterized by six torsional angles denoted by $\tau_1 - \tau_6$ in Figure 2. The τ_1 and χ_3 angles are dependent on each other; the τ_2 does not provide the usual three stable rotamers around the bond due to the bulky planar NAc group in position 2. This latter angle usually has only two stable positions (with several exceptions). The total number of the possible stable rotamers one of the selected molecules are $243 \times 2 \times 34 = 39366$ or more. The complete ab initio investigation of such a large conformational space for the given size of a molecule is currently impossible. However, our earlier studies have shown that a Monte Carlo conformational space search coupled with molecular mechanics (MM2*) provide useful, low-energy starting structures for further ab initio investigations. The torsional angle values around 60°, –60°, and 180° are denoted by g, g–, and t, respectively. (We use the letter t for the *anti* position in order to follow the notation generally used in the earlier papers). Torsional angles can be considerably distorted from these supposedly ideal values due to inter- or intraresidue interactions (hydrogen bonds). The value of τ_5 is denoted by capital letter in order distinguish the C-C-C-O type torsional angle from the other type of torsional angles in the carbohydrates.

Structural Background

Ab initio carbohydrate structures: The ab initio HF/6-31G(d) and B3LYP/6-31G(d) structures of the α- and β-anomers of glucose, galactose and mannose are known.[20, 21] and the HF/6-31G(d) structures are summarized in our web site in a three-dimensional structural database.[22] The MM2*, MMFF, HF, and GGA-DFT results are in agreement with each other in that the orientation of the four OH groups on the C1–C4 atoms of the pyranose ring show strong coupling. The most stable conformations tend to maximize the number of possible OH···O interactions and thus provide an intramolecular chain of OH groups and counterclockwise or clockwise patterns appear. (The definition of clockwise direction depends on the carbohydrate ring being in the standard orientation: O5 on the top and O1 on the right.) The formation of these bridges distorts the ideal three-fold

potential energy surface for individual secondary and primary OH groups. These effects greatly reduce the number of possible rotamers.

The position of the primary alcohol group, the internal rotation around the C5–C6 and the C6–O6 bonds (τ_5 and τ_6 in Figure 2) has virtually no influence on the energy difference between the two monosaccharide anomers at HF/6-31G(d) level of theory. NMR results in water for the CH_2OH group of α-D-glucose suggest that two conformations differing in τ_5 denoted by G– ($\tau_5 \approx -60^\circ$, *anti* to H5) and G ($\tau_5 \approx +60^\circ$, *anti* to C4) are populated in about 55:45 ratio at room temperature, while the population of the T conformation (with $\tau_5 \approx 180^\circ$, *anti* to O5) were considered negligible (less than 2%).[23] More recent NMR investigations in water for methyl α- and β-D-glucopyranosides provide similar results with a considerable (about 10%) uncertainty.[24] For galactopyranose derivatives, the corresponding populations are 10–25%, 55–78%, and 2–30%.[25] At HF/6-31G(d) level of theory the energy differences between the T, G–, and G rotamers is about 0.2 kcal mol^{-1} for both anomers of the D-glucose.[26, 27] For β-D-glucose the G– rotamer is the most stable. For α-D-glucose the T rotamer is the most stable. The comparison of the Gibbs free energies, instead of the electronic energies, shows that the G– rotamer of the α-D-glucose is slightly more stable (by 0.02 kcal mol^{-1}) than the T rotamer. Comparison of calculated B3LYP/6-311++G(d,p)// B3LYP/6-31G(d,p) Gibbs free energy values for various α- and β-D-gluco- and galactopyranosides in the gas phase and in water yields that the population of G–, G, and T rotamers depends on the hydrogen bonding and solvent effect.[25] The calculated rotamer populations of the CH_2OH group agreed well with experimental NMR data in water. In the crystal structure of the α-D-glucose,[28] the G– orientation was found for the CH_2OH group. The HF/6-31G(d) energy of the crystal structure is higher by 8 kcal mol^{-1} than the energy of the global minimum, because in this geometry the number of intramolecular hydrogen bridges decrease and intermolecular OH group interactions occur.

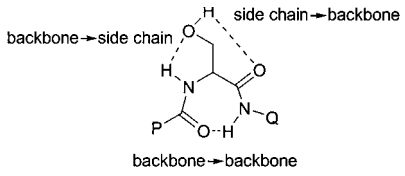
It was observed that the HF/6-31G(d) or cc-pVDZ relative total energies of various rotamers of D-glucose show considerable agreement with the rather expensive composite QM relative energies (in the composite method MP2/cc-pVTZ energies are corrected with CCSD correlation energy correction and cc-pTVQZ basis set correction, an approximation to CCSD/cc-pVQZ results; this is about the most expensive calculations for monosaccharides currently affordable).[26] Although differences can be observed between the HF and correlated geometries, the HF/6-31G(d) method predicts qualitatively correct rotamers (with chain of hydrogen bonds) and excellent energetic order. Another observation is that the relative energies of the family of rotamers change only slightly (about ± 0.2 kcal mol^{-1}) when the geometries are optimized at correlated level of theory (B3LYP or MP2). The influence of the basis sets and methods on the relative energies is considerably larger (about ± 3 kcal mol^{-1}, calculated with the same series of geometries). According to the experience in the literature that best performer among the least expensive methods for monosaccharides are HF/6-31G(d), HF/cc-pVDZ[26] and B3LYP/6-31+G(d).[20, 29–31] Barrows et al.[26] summarized the performance of the best MM methods for energies of D-glucose rotamers and found that the HF/6+31G(d) method is clearly superior compared with any MM parameterization.[26] A lot of work has been made recently to parametrize the molecular mechanics (MM) methods for saccharides in the gas phase using HF/6+31G(d) results.[32–35]

The excellent results obtained by HF/6-31G(d) method for monosaccharides does not necessarily guarantee the quality of the HF results for NAc and serine/threonine substituted monosaccharide rotamers. It is certainly preferable to support the results with calculations that include electron correlation. In this respect B3LYP calculations are among the most affordable and reliable. Such tests have not yet been performed for O-glycosides of amino acids and this is the first ab initio and DFT study for such compounds according to best of our knowledge.

Ab initio serine diamide conformations: The idealized backbone torsional angle values for various secondary structural elements of proteins are summarized in Table 1. N- and C-protected serine has been studied computationally during the past two decades. The global minimum is the γ_L[gg] rotamer (cf. γ_L in Table 1) that is stabilized by three hydrogen bonds shown in Scheme 1.

Table 1. The idealized backbone torsional angle values of selected secondary structural elements of proteins.

	Φ	ψ
α-helix (right handed)	-54°	-45°
α-helix (left handed)	$+54^\circ$	$+45^\circ$
3_{10} helix (right handed)	-60°	-30°
antiparallel β-chain	-139°	$+135^\circ$
parallel β-chain	-119°	$+113^\circ$
collagene helix	-51°	$+153^\circ$
type I turn		
(2nd amino-acid residue)	-60°	-30°
(3rd amino-acid residue)	-90°	0°
type II turn		
(2nd amino-acid residue)	-60°	$+120^\circ$
(3rd amino-acid residue)	-80°	0°
γ-turn (γ_D)	$+60^\circ$	-60°
inverse γ-turn (γ_L)	-60°	$+60^\circ$



Scheme 1. The hydrogen bonds in N- and C-protected serine.

One of the hydrogen bonds is backbone → backbone interaction (N2-H ··· O=C1) resulting in the γ_L ring formation. The other two are backbone → side-chain (N1-H ··· O-H) and side-chain → backbone (O-H ··· O=C) interactions. A total of 44 stable rotamers were located at the HF/3-21G level of theory out of the theoretically possible 81 rotamers (see below). The side-chain → backbone (O-H ··· O=C2) interaction of the serine OH is clearly missing from the conformational space of O-glycosides. Consequently it is expected that

0947-6539/02/0820-4722 $ 20.00+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 20

the less stable rotamers of the conformational space of N- and C-protected serine will occur among the most stable rotamers of glycosylated serine.

Results and Discussion

Energetic order

The MM2*-SUMM conformational search gave 19 rotamers for **1** within 3.63 kcal mol^{-1} of the global minimum (cf. Table 2). The global minimum was found 19 times during the search. The conformation of the serine backbone shows little variation. Only three kinds of rotamers occur: distorted γ_L, γ_D, and a rather distorted α_D (rotamer 14, cf. Table 2). The most stable eight rotamers have the same classifications for the peptide backbone and side chain orientations: γ_L[gtt], covering an energy range of 2.64 kcal mol^{-1}. The essential difference between these rotamers is in the orientation of the primary alcohol group (-CH_2OH) characterized by the τ_5, and τ_6 torsional angles: G – g –, Gg –, Gt, and Tg rotamers occur. Three rotamers (1, 3, and 5, Table 2) fall in the same category with the only difference among them being in τ_2 (the orientation of the NAc group). The value of τ_2 is usually between 160 and 165°. However, in a few rotamers its value is around – 140° (e.g. rotamers 3, 5, and 8). Thus, the NAc group is rotated in the counter-clockwise direction (see Figure 2.) by 90°. As result the C=O oxygen turns below the plane defined by C1, C2 and N2. Comparison of the torsional angles for rotamers 3 and 5 (Table 2, MM2* values) shows that these two rotamers are very similar, and the distinction made by the MM2* method is probably artificial.

We performed a similar conformational space search using a more recent LMOD conformational space search method[36] combined with MMFF94 force field.[16] Comparison of the MM2* and MMFF4 conformational space with the HF/6-31G(d) results showed that the MMFF94 results agree better with the HF/6-31G(d) results for the two most stable rotamers but spurious discrepancies occur in the higher energy regions. Thus the results obtained with the MMFF94 field were not used further and are not discussed here. They are available on the World Wide Web.[37]

Several rotamers of **1** found by the MM2* force field are not stable at the HF/6-31G(d) level of theory (cf. rotamers 3, 5, 7, 8, 11, 14, 17, 18 in Table 2). All MM2* rotamers with $\tau_2 \approx -140°$ disappear from the HF/6-31G(d) conformation space (rotamers 3, 5, 8, and 18). Instead, the ab initio value for τ_2 is consistently about 155°. Rotamer 7 of the MM2* conformation space is transformed into rotamer 4, which is rather similar to 7, in the ab initio conformation space. During this 7 → 4 transformation the value of the χ_2 angle is changed from its – 153° value to 162°. The latter value for χ_2 angle is characteristic for the most stable ab initio rotamers. The most stable ab initio rotamers can be characterized as γ_L[gtt | gttt*Xy*], where *X* and *y* symbolize the various orientations of the primary alcohol group characterized by the τ_5, and τ_6 torsional angles (e.g. G – g, Gg –, Gt or Tg, cf. Table 2). The ab initio results provide a different global minimum for (**1**), namely rotamer 6 in Table 2 from the same family of rotamers. However, rotamer 1 and especially rotamer 2 are only slightly less stable than rotamer 6 (the ab initio total energy difference among these three rotamers is less than 1 kcal mol^{-1}, cf. Table 2). Rotamer 4 is considerably less stable due to the Gt orientation of its primary alcohol CH_2OH group. This is because there is no hydrogen bond for the CH_2OH group in the Gt orientation. The HF/6-31G(d) calculations usually prefer the Tg position for this group in monosaccharides as discussed earlier.

The common feature of the most stable γ_L[gtt | gtttXy] rotamer family of (**1**) is the preference for the (Carb)-AcNH ··· O=C-NHMe(Ser) hydrogen bond (denoted as $(N2_{Carb})H \cdots O2_{Ser}$). This hydrogen bond anchors the galactose residue in an "orthogonal" position (see below) with respect to the serine residue (cf. gtt side chain orientation with two *anti* torsional angles). In the core structures of the O-glycosides O3 and/or O6 of GalNAc is involved in glycosidic linkages (see Introduction).[1] Therefore the O3H and/or O6H type hydrogen donor rotamers (e.g. G – g, Gg –, or Tg) are missing from the conformational space of such glycoproteins.

According to Table 3 the global minimum for **2** (found six times by the MM2*-SUMM search) is qualitatively the same structure by either the MM2* or HF/6-31G(d) methods. All rotamers except 16 from the MM2* calculations are stable. Rotamer 16 is transformed to rotamer 1 by the HF/6-31G(d) geometry optimization (cf. Table 3). The essential difference between these two rotamers is the value of χ_2. In rotamer 1 the value of χ_2 is 122°, a value that causes an eclipsed value (2°) for the H-C$^{\beta}$-O1-C1 torsional angle. In rotamer 16, the value of χ_2 is 178°,a nearly perfect *anti* position with no eclipsed atom pairs. (This difference does not occur in our notation because the three letter notation classifies any angle falling in the 120–240° range as *anti* and denoted by t). It seems somewhat unusual that a rotamer containing eclipsed atoms is the most stable. A survey of the Cambridge Structural Data Base[38] revealed several structures (CSD refcodes: DMGALP, RONHEH and ZOSSEF) that contain eclipsed H-C-O-C torsional angles. The latter ZOSSEF structure contains perfectly eclipsed methyl carbon atoms due to symmetry reasons. Our recent investigation of 2,3-di-*O*-methyl-D-galactopyranosiduronic methyl ester derivatives (to be published elsewhere) show that such eclipsed rotamers are quite frequent among O-methyl groups.[39] Similarly, due to the preferable inter-residue interactions (O5 ··· HN1), the non-eclipsed rotamer 16 seamlessly transformed to rotamer 1 (eclipsed) during the HF/6-31G(d) geometry optimization. Besides the most stable rotamer 1, rotamers 7, 9, 15, and 19 also contain eclipsed atoms (cf. ab initio values for χ_2 in Table 3). The relative MM2* and HF/6-31G(d) energies of the various rotamers differ considerably. This is especially true for rotamers 14 and 17 (cf. HF/6-31G(d) values in Table 3), which are quite stable according to the ab initio results. We note that the non-exoanomeric rotamers with τ_1 about 60° instead of – 60° (cf. rotamers 2 and 4 in Table 3) are predicted to occur with similar energies (within 1 kcal mol^{-1}) to those of exoanomeric ones. This point will be discussed in more detail below.

For model systems containing the mannose residue the situation is somewhat different. While the results for **3** in

  0947-6539/02/0820-4723 $ 20.00+.50/0

Table 4 show the usual difference between the MM2* and HF/6-31G(d) relative energies (slightly different global minimum and different energetic order for various rotamers), the largest differences between MM2* and HF/6-31G(d) relative energies were observed in the conformation space of **4** (up to 7 kcal mol^{-1}, cf. Table 5). The general feature is that the higher energy rotamers of the MM2* conformation space are considerably more stable according to the HF/6-31G(d) method. For example the most stable rotamer of HF/6-31G(d) conformational space is the rotamer 17 of the MM2* conformational space (cf. Table 5).

The most stable rotamers of **5** and **6**, α- and β-GlcNAc-Ser, which are the analogues of **1** and **2**, were constructed from the most stable HF/6-31G(d) rotamers of **1** and **2**. The axial C4 OH was changed to equatorial, and $\tau_4-\tau_6$ torsion angles were taken from the corresponding free D-glucose monomer. This procedure is expected to lead directly to the most stable rotamers of **5** and **6** without an extensive conformational space search. The three most stable rotamers of **5** corresponds to γ_L[gtt | gtttXy] with Xy = Tg, Gg −, and G − g. According to the HF/6-3G(d) results the relative stabilities of Tg, Gg −, and G − g rotamers of **5** are −0.33, 0.06, and 0.51 kcal mol^{-1}, respectively, compared with the most stable rotamer 6 of **1** in Table 2. Thus **5** is slightly more stable than **1**. The reason for this is the following: The oxygen atom of NAc group at C2 stabilizes the counter clockwise direction of OH groups (by a C=O ··· H-O3 interaction). The most stable rotamer of α-D-glucose show such pattern. However, the most stable rotamer of α-D-galactose shows clockwise direction that is destabilized in **1** (cf. rotamer 9 of **1** in Table 2, it is less stable by 3.71 kcal mol^{-1}). Notably according to the HF/6-31G(d) results the rotamers of **6** are rather stable. The relative energy of the most stable γ_L[gtg − | gtttTg] rotamer of **6** is 3.51 kcal mol^{-1}; this is the most stable among the rotamers of various β-anomers (cf. βGalNacSer, 3.92 kcal mol^{-1} in Table 3, βManNAcSer 4.93 kcal mol^{-1} in Table 5). The rela-

Table 2. Relative energies (ΔE in kcal mol^{-1}) and torsional angles of the low energy rotamers of αGalNAcSer (**1**).[a]

		Serine backbone			Serine side chain				Carbohydrate						
	ΔE	Φ	ψ	Code	χ_1	χ_2	χ_3	Code	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	Code
MM2*															
1	0.00	−77	56	γ_L	63	163	−156	gtt	83	165	−162	−170	−54	−39	gtttG − g −
2	0.53	−77	56	γ_L	64	165	−155	gtt	84	164	−161	−173	59	−57	gtttGg −
3	1.57	−77	58	γ_L	62	136	−161	gtt	76	−146	166	−169	−55	−39	gtttG − g −
4	1.84	−77	56	γ_L	64	165	−156	gtt	83	164	−161	−173	68	−180	gtttGt
5	2.02	−77	53	γ_L	57	159	−169	gtt	69	−133	163	−169	−50	−40	gtttG − g −
6	2.59	−77	56	γ_L	63	162	−156	gtt	83	165	−161	−171	−164	75	gtttTg
7	2.64	−73	44	γ_L	78	−153	−160	gtt	79	160	−161	−172	69	−178	gtttGt
8	2.64	−77	52	γ_L	58	161	−169	gtt	69	−133	163	−171	59	−57	gtttGg −
9	2.99	−75	52	γ_L	−55	158	−170	g-tt	70	13	−36	61	−52	53	ggg − gG − g
10	3.06	70	−51	γ_D	180	89	−164	tgt	76	162	−160	−170	−49	−40	gtttG − g −
11	3.11	−77	56	γ_L	63	164	−155	gtt	83	164	−160	−174	174	175	gtttTt
12	3.24	57	−13	γ_D	62	−151	−177	gtt	61	159	−161	−170	−50	−37	gtttG − g −
13	3.44	69	−62	γ_D	−174	−169	−170	ttt	69	160	−158	−173	61	−57	gtttGg −
14	3.44	49	32	α_D	61	146	−174	gtt	64	160	−160	−170	−50	−37	gtttGg −
15	3.49	−77	56	γ_L	63	163	−153	gtt	86	159	−108	66	−53	50	gtg − gG − g
16	3.58	−77	56	γ_L	63	162	−157	gtt	82	165	−160	−173	68	77	gtttGg
17	3.58	−60	69	γ_L	−174	−177	−173	ttt	67	160	−159	−170	−50	−39	gtttG − g −
18	3.62	−77	58	γ_L	62	139	−162	gtt	75	−149	167	−171	67	−180	gtttGt
19	3.63	−77	63	γ_L	−160	−167	−164	ttt	75	159	−158	−173	62	40	gtttGg
HF/6-31G(d)															
1	0.99	−86	69	γ_L	53	160	−155	gtt	80	158	−142	−168	−54	−47	gtttG − g −
2	0.03	−87	68	γ_L	54	162	−156	gtt	81	157	−141	−169	61	−63	gtttGg −
3 → 1[b]	0.99														
4	2.28	−87	68	γ_L	54	162	−157	gtt	79	158	−143	−168	72	−164	gtttGt
5 → 1[b]	0.99														
6[c]	0.00	−86	68	γ_L	54	161	−155	gtt	81	158	−141	−167	−171	79	gtttTg
7 → 4[b]	2.28														
8 → 2[b]	0.03														
9	3.71	−83	73	γ_L	−63	150	−173	g-tt	66	36	−42	37	−58	62	ggg − gG − g
10	4.98	75	−51	γ_D	−176	95	−165	tgt	72	156	−139	−168	−53	−47	gtttG − g −
11 → 6[b]	0.00														
12	7.67	49	49	α_D	57	131	−170	gtt	67	155	−139	−167	−53	−45	gtttG − g −
13	8.29	75	−74	γ_L	−170	−168	−170	ttt	66	155	−139	−170	63	−67	gtttGg −
14 → 12[b]	7.67														
15	3.77	−86	68	γ_D	53	163	−152	gtt	84	149	−85	45	−59	58	gtg − gG − g
16	3.78	−86	70	γ_D	53	160	−157	gtt	79	158	−142	−168	68	74	gtttGg
17 → 10[b]	4.98														
18 → 4[b]	2.28														

[a] The MM2* energy of the most stable rotamer is −62.31 kcal mol^{-1}. For the definition of the torsional angles refer to Figure 2. [b] The notation 3 → 1 means that rotamer 3 provided by the MM2* method is not stable according to the HF/6-31G(d) method and during the geometry optimization it was transformed to rotamer 1. Analogous notation is used for the other unstable MM2* rotamers. [c] The most stable rotamer according to the ab initio HF/6-31G(d) method, total energy = −1307.00775 hartree.

0947-6539/02/0820-4724 $ 20.00+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 20

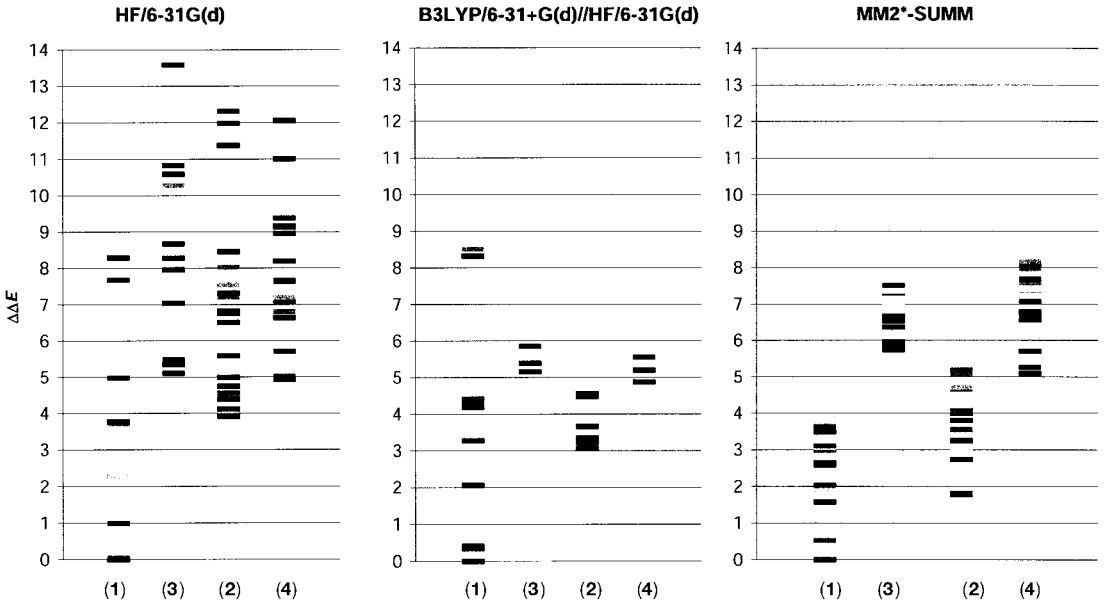


Figure 4. Comparison of HF/6-31G(d), B3LYP/6-31+G(d)//HF/6-31G(d) and MM2* relative energies of **1**, **2**, **3**, and **4** in kcal mol^{-1}. The energy of the most stable rotamer of **1** is selected as the reference.

tive stability of the primary alcoholic rotamers is 3.96 and 4.42 kcal mol^{-1} for Gg− and G−g rotamers, respectively. These rotamers were derived from the second most stable rotamer of **2**. Another rotamer of **6** were derived from the most stable rotamer of **2**. The relative HF/6-31G(d) energy of this γ_L[g−tt|g−-tttG−g−] rotamer is 5.61 kcal mol^{-1}, thus it is considerably destabilized.

Figure 4 shows the essential difference between MM2* and HF/6-31G(d) relative energies. The ab initio energy differences are considerably larger (8–9 kcal mol^{-1} for the first 10–20 rotamers) than that of MM2* (usually 2–3 kcal mol^{-1} for about 20 rotamers). Thus, the MM2* conformational space is energetically more compressed relative to the HF/6-31G(d) conformational space. Our earlier results show that the MM2* method provides different energetic order for the rotamers in the higher energy region.[30, 31] This is not surprising since carbohydrates are rather difficult tests for MM methods because they have densely packed, highly polar functional groups, and the conformational energies depend on stereoelectronic effects. These difficulties certainly influence the relative energies of the molecules investigated in the present paper. On the other hand, earlier studies have found that the HF/6-31G(d) and cc-pVDZ results provide quite good relative energies for monosaccharides that are close to results of the highest-level calculations currently affordable. The B3LYP/6-31+G(d)//HF/6-31G(d) conformational space of the four model compounds resembles the HF/6-31G(d) conformational space (cf. Figure 4). The most noticeable difference is that rotamer 2 is the lowest energy rotamer of **1** (instead of rotamer 6) and the lowest energy rotamer of **2** is 1 kcal mol^{-1} more stable according to the B3LYP/6-31+G(d)//HF/6-31G(d) results (cf. Figure 4 and Table 7). Vibrational frequency analyses for HF/6-31G(d) optimized geometries confirmed that each stationary point is a true minimum on the potential energy surface. Inspection of the components of the calculated ΔH_{298} and ΔG_{298} data (cf. Table 7) shows that the ZPE and thermal corrections to the energy are rather small (the energy order remains the same). However, the $T\Delta S_{298}$ term can be as large as 2.6 kcal mol^{-1} ($T\Delta S_{298} = \Delta H_{298} - \Delta G_{298}$). In general, for the molecules studied here, the entropy differences stabilize the α-anomers relative to β-anomers, and ManNAc relative to GalNAc (cf. Table 7). The largest stabilization effect can be observed for rotamer 6 of αManNAcSer (cf. Table 7). The two anomers of GlcNAcSer are the most stable. This might explain the occurrence of the βGlcNAc monosaccharide side chains on serine or threonine. However, the missing αGlcNAc-type O-glycosides cannot be explained by the stability alone. The relative energies exclude αManNAc as the root of the saccharide antenna, but certainly other factors make the αGlcNAc less preferred. α- or βGlcNAc was never found as O-glycosidic linker in saccharide core chains. βGlcNAc occurs only as a monosaccharide. This could be attributed to structural and/or reactivity properties of αGalNAc that makes the chain continuation "easier". (We recall that βGlcNAc is the exclusive root for N-glycosidic linkage. This will be investigated in a subsequent paper.) Among the β-anomers the βGlcNAcSer is the most stable (cf. Table 7).

A series of HF/6-31G(d) and B3LYP/6-31+G(d)//HF/6-31G(d) calculations were performed on the threonine analogues **7**–**12** of the selected molecules. These molecules were generated from the analogous serine containing molecules. The relative energies are shown in Table 8. Comparison of these relative energies with those in Tables 2–7 shows that the extra methyl group in threonine does not influence considerably (no more than few tenth of kcal mol^{-1}) the relative energies of the most stable rotamers in most of the cases. The only exception is **12**, βGlcNAcThr, for which a considerable destabilization was observed at HF level of theory (cf. Table 8). However, this effect is diminished by the inclusion of the electron correlation effects (nearly 2 kcal mol^{-1} stabilization). Careful analysis of the molecular

  0947-6539/02/0820-4725 $ 20.00+.50/0

Table 3. Relative Energies (ΔE in kcal mol^{-1}) and torsional angles of the low energy rotamers of βGalNAcSer (**2**).[a]

		Serine backbone			Serine side chain				Carbohydrate						
	ΔE	φ	ψ	Code	χ_1	χ_2	χ_3	Code	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	Code
MM2*															
1	1.78	−77	53	γ_L	−55	122	−177	g−tt	−62	−133	163	−168	−49	−51	g−tttG−g−
2	1.82	−76	54	γ_L	59	167	−69	gtg−	51	−128	165	−168	−51	−45	gtttG−g−
3	2.73	−76	48	γ_L	38	72	164	ggt	−82	−132	163	−168	−51	−48	g−tttG−g−
4	2.97	−76	54	γ_L	58	166	−68	gtg−	52	−128	165	−170	61	−55	gtttGg−
5	3.24	61	−62	γ_D	75	−167	177	gtt	−67	−132	165	−168	−52	−41	g−tttG−g−
6	3.25	−77	57	γ_L	−175	−159	159	ttt	−84	−140	167	−168	−52	−40	g−tttG−g−
7	3.55	73	−50	γ_D	−52	147	174	g−tt	−72	−132	163	−170	62	−65	g−tttGg−
8	3.80	−76	48	γ_L	45	142	170	gtt	−75	−131	162	−170	58	−64	g−tttGg−
9	3.99	−78	54	γ_L	−63	120	175	g−tt	−69	−131	163	−169	56	173	g−tttGt
10	4.10	−77	57	γ_L	−175	−159	158	ttt	−85	−139	166	−169	60	−57	g−tttGg−
11	4.12	62	−60	γ_D	77	−168	175	gtt	−69	−132	165	−169	61	−57	g−tttGg−
12	4.21	−76	53	γ_L	59	167	−69	gtg−	51	−128	165	−170	70	179	gtttGt
13	4.27	71	−48	γ_D	−172	−134	168	ttt	−75	−134	165	−168	−52	−40	g−tttG−g−
14	4.56	−76	53	γ_L	58	167	−68	gtg−	51	−128	164	−169	−163	74	gtttTg
15	4.58	−85	−10	α_L	56	128	177	gtt	−68	−133	162	−170	61	−69	g−tttGg−
16	4.58	−76	53	γ_L	−56	178	173	g−tt	−71	−132	163	−168	−52	−43	g−tttG−g−
17	4.65	−77	52	γ_L	45	147	−176	gtt	−62	84	−41	47	−49	61	g−gg−gG−g
18	4.71	−75	48	γ_L	58	−153	178	gtt	−66	−4	−35	59	−54	53	g−g−g−gG−g
19	5.06	69	−47	γ_D	85	123	177	gtt	−68	−130	164	−169	58	−73	g−tttGg−
20	5.08	70	−51	γ_D	−174	−123	167	ttt	−76	−138	165	−169	60	−56	g−tttGg−
21	5.18	66	20	α_D	−46	142	175	g−tt	−70	−133	163	−170	62	−67	g−tttGg−
HF/6-31G(d)															
1[b]	3.92	−87	71	γ_L	−59	125	176	g−tt	−64	−148	172	−169	−52	−52	g−tttG−g−
2	4.39	−86	61	γ_L	53	173	−69	gtg−	54	−131	167	−170	−54	−49	gtttG−g−
3	4.74	−85	55	γ_L	40	67	160	ggt	−81	−145	171	−169	−50	−52	g−tttG−g−
4	4.11	−86	61	γ_L	52	173	−68	gtg−	55	−130	166	−171	61	−59	gtttGg−
5	11.98	72	−40	γ_D	85	−157	178	gtt	−60	171	−157	−166	−53	−46	g−tttG−g−
6	5.58	−87	72	γ_L	−177	−156	163	ttt	−75	172	−159	−166	−53	−45	g−tttG−g−
7	7.29	77	−39	γ_D	−59	137	176	g−tt	−63	−162	179	−169	61	−74	g−tttGg−
8	6.83	−86	56	γ_L	29	91	167	gtt	−74	−147	170	−171	77	−43	g−tttGg−
9	7.21	−88	70	γ_L	−64	121	167	g−tt	−73	−148	173	−168	60	176	g−tttGt
10	4.99	−87	73	γ_L	−176	−156	161	ttt	−77	171	−157	−167	59	−62	g−tttGg−
11	11.38	67	−35	γ_D	79	−173	174	gtt	−65	−142	169	−169	61	−65	g−tttGg−
12	6.51	−86	60	γ_L	53	174	−69	gtg−	54	−132	167	−170	71	−170	gtttGt
13	8.46	76	−52	γ_D	−167	−146	172	ttt	−66	−155	177	−167	−53	−46	g−tttG−g−
14	3.94	−86	61	γ_L	53	173	−69	gtg−	54	−130	166	−168	−171	78	gtttTg
15	6.75	−119	14	γ_L	42	110	172	gtt	−67	−151	172	−171	70	−59	g−tttGg−
16 → 1[c]	3.92														
17	4.55	−86	53	γ_L	42	142	−171	gtt	−52	94	−46	20	−56	67	g−gg−gG−g
18	7.55	−85	44	γ_L	58	−146	−179	gtt	−62	−30	−38	35	−59	61	g−g−g−gG−g
19	12.32	72	−43	γ_D	84	118	178	gtt	−63	−138	168	−170	57	−76	g−tttGg−
20	8.02	76	−53	γ_D	−167	−146	171	ttt	−68	−149	173	−168	61	−61	g−tttGg−
21	7.32	68	31	α_D	−53	136	176	g−tt	−63	176	−165	−168	61	−74	g−tttGg−

[a] The energy of the most stable rotamer 1 in Table 2 is used as reference: −62.31 kcal mol^{-1}. The MM2* energy of the most stable rotamer 1 in this Table is −60.53 kcal mol^{-1}. For the definition of the torsional angles refer to Figure 2. [b] The most stable rotamer according to the ab initio HF/6-31G(d) method, total energy = −1307.00150 hartree. The energy of the most stable rotamer 6 in Table 2 is used as reference: −1307.00775 hartree. [c] The notation 16 → 1 means that rotamer 16 provided by the MM2* method is not stable according to the HF/6-31G(d) method and during the geometry optimization it was transformed to rotamer 1.

geometry provides that the threonine analogues conserve the values of the essential torsional angles. Detailed geometry analysis follows next.

Geometry analysis

Serine backbone conformations: Serine related model peptides have been extensively studied with spectroscopy in the past.[40] Furthermore, many ab initio results for serine peptide models were reported during the last fifteen years.[41, 42]

The distribution of the Φ and ψ torsional angles of the serine residue is presented in Figure 5 as Ramachandran plots for MM2* and HF/6-31G(d) methods. For the definition of these torsional angles refer to Figure 2. Most of the rotamers fall into two categories, γ_L or γ_D (cf. notation for γ turn and inverse γ turn in Figure 3 and codes in Tables 2–6). The distribution of the various backbone rotamers can be followed in Figure 5. The φ and ψ torsional angles have their typical values in the most stable rotamers around −77 and 55°, respectively. As noted earlier, γ_L corresponds to the most stable rotamer of the free serine monomer.[42] α-Helices (left and right turns) can be observed in a few of the rotamers. For **1** and **2** the single α_D rotamer in the MM2* conformational space disappears from the HF/6-31G(d) conformational space (cf. rotamer 14 in Table 2). However, the γ_L rotamer 12 is transformed by HF/6-31G(d) method into a new α_D rotamer

0947-6539/02/0820-4726 $ 20.00+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 20

Table 4. Relative energies (ΔE in kcal mol^{-1}) and torsional angles of the low energy rotamers of αManNAcSer (**3**).[a]

		Serine backbone			Serine side chain				Carbohydrate						
	ΔE	Φ	ψ	Code	χ_1	χ_2	χ_3	Code	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	Code
MM2*															
1	5.75	−77	51	γ_L	54	175	−173	gtt	67	−159	−147	173	58	−56	gtttGg−
2	5.82	−77	64	γ_L	−159	−165	−163	ttt	77	−157	−150	170	60	42	gtttGg
3	5.95	−77	54	γ_L	54	176	−179	gtt	67	−159	−147	170	170	53	gtttTg
4	6.30	69	−63	γ_D	−174	−170	−171	ttt	70	−158	−151	169	60	−55	gtttGg−
5	6.37	−74	44	γ_L	75	−154	159	gtt	82	−157	−152	169	59	−68	gtttGg−
6	6.54	−77	51	γ_L	54	175	−173	gtt	68	−159	−145	175	−52	58	gtttG−g
7	6.55	−74	42	γ_L	79	−148	−160	gtt	80	−157	−151	168	67	177	gtttGt
8	6.57	−104	2	γ_L	82	82	−104	ggg−	86	−155	−148	169	63	179	gtttGt
9	6.61	69	−62	γ_D	−173	−172	−172	ttt	69	−158	−149	168	170	51	gtttTg
10	6.65	71	−49	γ_D	154	−169	−179	ttt	63	−159	−150	172	62	52	gtttGg
11	6.67	−77	51	γ_L	54	176	−174	gtt	67	−159	−147	172	65	175	gtttGt
12	6.94	−77	51	γ_L	54	176	−174	gtt	67	−159	−141	176	−59	−176	gtttG−t
13	7.08	−76	55	γ_L	−53	−177	−172	g−tt	69	−158	−149	172	58	−54	gtttGg−
14	7.21	−81	55	γ_L	−51	−101	−165	g−g−t	74	−158	−146	172	64	69	gtttGg
15	7.31	−76	55	γ_L	−53	−178	−172	g−tt	69	−158	−148	169	170	51	gtttTg
16	7.34	−71	−19	α_L	82	−147	−164	gtt	77	−157	−150	170	68	−179	gtttGt
17	7.50	71	−51	γ_D	−50	−69	−148	g−g−t	93	−158	−148	170	64	179	gtttGt
HF/6-31G(d)															
1	5.34	−86	59	γ_L	51	127	−170	gtt	69	−154	−144	173	60	−58	gtttGg−
2	7.03	−85	91	γ_L	−170	−168	−172	ttt	66	−155	−157	171	64	53	gtttGg
3[b]	5.11	−86	60	γ_L	51	127	−169	gtt	69	−154	−142	173	167	52	gtttTg
4	10.58	75	−76	γ_D	−169	−170	−169	ttt	69	−154	−159	170	62	−60	gtttGg−
5	8.67	−86	45	γ_L	59	−157	−144	gtt	94	−154	−157	170	62	−62	gtttGg−
6	5.48	−86	59	γ_L	50	126	−169	gtt	69	−154	−146	177	−55	63	gtttG−g
7	10.83	−85	42	γ_L	61	−160	−147	gtt	90	−155	−154	169	73	−175	gtttGt
8	8.28	−108	−4	γ_L	67	−93	−142	ggg−	96	−154	−154	170	66	178	gtttGt
9	10.27	75	−73	γ_D	−170	−173	−171	ttt	67	−155	−157	171	165	53	gtttTg
10	13.58	73	−73	γ_D	165	−176	176	ttt	54	−155	−158	171	62	63	gtttGg
11	7.96	−86	60	γ_L	51	128	−172	gtt	67	−154	−143	171	70	−179	gtttGt
12	7.36	−86	59	γ_L	50	128	−169	gtt	69	−155	−137	176	−62	−174	gtttG−t
13c	7.34	−88	74	γ_L	−57	−180	−172	g−tt	66	−154	−156	173	60	−60	gtttGg−
14	6.48	−90	75	γ_L	−60	−95	−163	g−g−t	74	−155	−150	173	65	71	gtttGg
15[c]	7.26	−88	73	γ_L	−57	−178	−170	g−tt	68	−155	−152	173	166	52	gtttTg
16[c]	10.67	−86	−9	α_L	71	−149	−153	gtt	85	−155	−153	169	72	−173	gtttGt
17[c]	8.51	−55	−42	γ_D	−59	−68	−147	g−g−t	91	−154	−151	170	70	−173	gtttGt

[a] The energy of the most stable rotamer 1 in Table 2 is used as reference: −62.31 kcal mol^{-1}. The MM2* energy of the most stable rotamer 1 in this Table is −56.56 kcal mol^{-1}. For the definition of the torsional angles refer to Figure 2. [b] The most stable rotamer according to the ab initio HF/6-31G(d) method, total energy = −1306.99961 hartree. The energy of the most stable rotamer 6 in Table 2 is used as reference: −1307.00775 hartree. [c] Loose geometry optimization.

(this rotamer is in the high-energy region, 7.67 kcal mol^{-1} above the global minimum, cf. Table 2). Only one α_D rotamer can be found in the conformational space of **2**, again in the relatively high energy region (cf. Table 3 and Figure 5). The MM2* and HF/6-31G(d) results for the peptide backbone conformations agree quite well, with only a few exceptions.

Serine side-chain conformations: The values of χ_1, χ_2, and χ_3 torsional angles of the C^{α}-C^{β}-O-$C1_{anomeric}$ bonds control the relative orientation of the serine and the monosaccharide residues thus giving major direction of the carbohydrate antenna. The distribution of the various side chain rotamers can be followed in the Figure 6. The value of χ_1 for **1** is about 60° (*gauche* relative to $N1_{Ser}$ and *anti* relative to $C^{\alpha}H$) in the most stable rotamers. This angle deviates only slightly from the supposedly ideal value of 60° in these rotamers (cf. Table 2). The g− (*anti* relative to $C2_{Ser}$) and t (*anti* relative to $N1_{Ser}$) rotamers of $\chi1$ are considerably less frequent and less stable, and most are missing from the HF/6-31G(d) conformational space. For compounds **2**–**4**, χ_1 occurs considerably more frequently in the g− or t orientation (cf. the most stable rotamer of **2** in Table 3). In these compounds the χ_1 torsional angle shows considerable deviation (15–20°) from the ideal values, indicating strained structures.

The value of χ_2 for **1** is about 180° (*anti* relative to C^{α}) in the most stable rotamers. It can be observed that this angle deviates considerably (usually by ±20°) from the so called ideal value of 180° in these rotamers (cf. Table 2). This torsional angle is a primary influence on the orientation of the carbohydrate antenna, providing an *anti* arrangement for the bulky GalNAc residue. Those rotamers of the MM2* conformational space that have large deviations from the ideal value (40°, e.g. rotamers 3, 14, and 18 in Table 2) are missing from the HF/6-31G(d) conformational space. As noted above for rotamers 1, 7, 9, 15, 19, and 21 of the HF/6-31G(d) conformational space of (**2**), the value of χ_2 is about 120° (cf. Table 3), yielding an eclipsed position for the H-C^{β}-O1-C1 torsional angle. The interresidue, $N1_{Ser}H\cdots O5$ hydrogen bond forces the residues into this strained position. Also, the most stable rotamers of **2** have a bent shape in which the galactose residue turns toward the serine residue. The other most stable rotamer of **2** according to HF/6-31G(d) results

Table 5. Relative energies (ΔE in kcal mol^{-1}) and torsional angles of the low energy rotamers of βManNAcSer (**4**).[a]

	ΔE	Φ	ψ	Code	χ_1	χ_2	χ_3	Code	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	Code
MM2*															
1	5.09	63	−28	γ_D	73	127	−180	gtt	−66	141	−73	168	57	−69	g − tg − tGg −
2	5.25	57	−12	γ_D	62	123	−179	gtt	−64	−163	−144	172	60	−68	g − tttGg −
3	5.71	−76	49	γ_L	48	145	172	gtt	−75	−162	−141	175	52	−63	g − tttGg −
4	6.08	−77	52	γ_L	55	148	179	gtt	−68	−161	−133	176	−48	62	g − tttG − g
5	6.56	−77	51	γ_L	50	148	−176	gtt	−62	−82	57	−82	−51	61	g − g − gg − G − g
6	6.58	−78	55	γ_L	−64	121	176	g − tt	−70	−162	−140	174	52	170	g − tttGt
7	6.66	73	−51	γ_D	−51	151	173	g − tt	−74	−161	−143	174	58	−64	g − tttGg −
8	6.78	61	−18	γ_D	75	131	−177	gtt	−62	144	−73	168	−47	68	g − tg − tG − g
9	7.06	−78	53	γ_L	−54	122	−175	g − tt	−60	−161	−139	174	−52	−173	g − tttG − t
10	7.28	−77	51	γ_L	49	148	−176	gtt	−62	−83	170	178	−48	61	g − g − ttG − g
11	7.37	−77	52	γ_L	46	178	169	gtt	−75	−85	54	−57	−174	−172	g − g − gg − Tt
12	7.38	55	−6	γ_D	68	132	−178	gtt	−62	−163	−151	172	−50	69	g − tttG − g
13	7.41	−76	48	γ_L	45	145	175	gtt	−71	−78	58	−77	51	−62	g − g − gg − Gg −
14	7.51	−77	57	γ_L	63	168	−163	gtt	−49	−163	−128	173	57	−56	g − tttGg −
15	7.60	−78	55	γ_L	−63	121	177	g − tt	−69	−163	−139	175	53	67	g − tttGg
16	7.62	−76	48	γ_L	45	146	176	gtt	−71	−80	170	177	50	−62	g − g − ttGg −
17	7.65	−77	57	γ_L	63	169	−161	gtt	−46	−163	−133	171	170	53	g − tttTg
18	7.68	69	−51	γ_D	163	94	−169	tgt	−55	−167	−113	176	−56	51	g − tg − tG − g
19	7.68	−87	1	γ_L	64	138	−176	gtt	−63	−161	−132	176	−48	69	g − tttG − g
20	7.96	−77	53	γ_L	58	−178	−177	gtt	−62	−161	−141	171	170	53	g − tttTg
21	8.06	−91	−23	α_L	51	−173	168	gtt	−75	−79	55	−57	−174	−172	g − g − gg − Tt
22	8.09	−77	52	γ_L	60	148	−175	gtt	−62	−88	−83	175	−48	63	g − g − g − tG − g
23	8.11	64	−29	γ_D	83	130	−178	gtt	−64	−160	−136	175	−50	71	g − tttG − g
24	8.17	75	−52	γ_D	−43	159	180	g − tt	−67	−161	−137	175	−46	66	g − tttG − g
HF/6-31G(d)															
1 → 2[b]	9.16														
2	9.16	59	−7	γ_D	56	117	176	gtt	−63	−158	−137	171	66	−63	g − tttGg −
3	7.65	−86	44	γ_L	46	146	170	gtt	−71	−157	−142	175	56	−63	g − tttGg −
4	4.97	−86	72	γ_L	52	152	−162	gtt	−42	−161	−118	174	−57	56	g − tttG − g
5	9.13	−86	56	γ_L	46	−178	162	gtt	−76	−91	55	−84	−61	63	g − g − gg − G − g
6	7.08	−87	73	γ_L	−65	122	167	g − tt	−73	−156	−132	171	57	163	g − tttGt
7	9.38	78	−57	γ_D	−62	145	169	g − tt	−71	−155	−141	172	64	−66	g − tttGg −
8 → 12[b]	12.06														
9	6.81	−86	74	γ_L	−59	126	176	g − tt	−64	−155	−132	172	−58	−163	g − tttG − t
10	8.19	−86	70	γ_L	45	177	174	gtt	−67	−88	−176	180	−55	63	g − g − ttG − g
11	9.16	−86	54	γ_L	46	−172	159	gtt	−80	−91	53	−51	−177	−179	g − g − gg − Tt
12	12.06	61	−16	γ_D	68	127	−177	gtt	−55	−156	−138	171	−48	76	g − tttG − g
13	11.01	−85	43	γ_L	43	143	171	gtt	−69	−88	55	−76	54	−63	g − g − gg − Gg −
14	5.71	−86	68	γ_L	53	167	−159	gtt	−39	−159	−125	171	61	−61	g − tttGg −
15	6.64	−86	74	γ_L	−65	122	168	g − -tt	−72	−156	−130	172	55	74	g − tttGg
16	8.96	−86	70	γ_L	46	−171	167	gtt	−73	−88	−175	175	59	−60	g − g − ttGg −
17[c]	4.93	−86	69	γ_L	53	166	−158	gtt	−38	−159	−126	170	167	52	g − tttTg
18	7.20	73	−50	γ_D	171	90	−174	tgt	−54	−161	−119	175	−58	53	g − tg − tG − g
19	7.63	−108	−1	γ_L	55	132	−172	gtt	−52	−156	−141	176	−50	75	g − tttG − g
20 → 17[b]	4.93														

[a] The energy of the most stable rotamer 1 in Table 2 is used as reference: −62.31 kcal mol^{-1}. The MM2* energy of the most stable rotamer 1 in this Table is −57.22 kcal mol^{-1}. For the definition of the torsional angles refer to Figure 2. [b] The notation 1 → 2 means that rotamer 1 provided by the MM2* method is not stable according to the HF/6-31G(d) method and during the geometry optimization it was transformed to rotamer 2. [c] The most stable rotamer according to the ab initio HF/6-31G(d) method, total energy = −1306.99316 hartree. The energy of the most stable rotamer 6 in Table 2 is used as reference: −1307.00775 hartree.

does not have a similar strain as χ_2 is about 173° (cf. rotamers 2, 4, and 14 in Table 3). In these rotamers of **2** the same interresidue $(N2_{Carb})H\cdots O2_{Ser}$ interaction occurs as in the most stable rotamers of **1**. This is the origin of the surprising stability of the non-exoanomeric rotamers of **2**. However, the overall shape of these rotamers of **2** is quite different from the shape of **1** due to the difference between α- and β-anomers. In the most stable rotamers of **3** χ_2 is about 127° (cf. rotamers 1, 3, and 6 in Table 4) this again yields an eclipsed position for the H-C$^{\beta}$-O1-C1 torsional angle. Oddly enough a $(C1_{Carb})$-$H\cdots O2_{Ser}$ interaction stabilizes this conformer. Such interactions are missing from the most stable rotamers of the conformational space of other compounds. In the most stable rotamers of **4** χ_2 is about 170° (cf. rotamers 4, and 17 in Table 5). In these rotamers the usual $(N2_{Carb})H\cdots O2_{Ser}$ interaction occurs; however, this leads to a rather bent structure in **4**. In rotamer 4 of **4** $(O6_{Carb})H\cdots O2_{Ser}$ supplements this interaction, yielding an $O2_{Ser}$ with two hydrogen bonds. While this structure is among the most stable structures according to MM2* results, the ab initio results show the rotamer 17 of **4** to be more stable.

The χ_3 torsional angle does not show a large variation. With very few exception this angle takes an *anti* position in accordance with the *anti* anomeric rule for α-anomers. We

0947-6539/02/0820-4728 $ 20.00+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 20

Table 6. B3LYP/6-31+G(d)//HF/6-31G(d) relative energies ΔE, relative enthalpies ΔH_{298}, and relative Gibbs free energies ΔG_{298} [kcal mol^{-1}] of the selected low energy rotamers.[a]

Rotamer	αGalNAcSer (**1**)			βGalNAcSer (**2**)			αManNAcSer (**3**)			βManNAcSer (**4**)			αGlcNAcSer (**5**)			βGlcNAcSer (**6**)		
	ΔE	ΔH_{298}	ΔG_{298}	ΔE	ΔH_{298}	ΔG_{298}	ΔE	ΔH_{298}	ΔG_{298}	ΔE	ΔH_{298}	ΔG_{298}	ΔE	ΔH_{298}	ΔG_{298}	ΔE	ΔH_{298}	ΔG_{298}
1	0.35	0.42	0.66		**3.07**	**3.22**	**3.25**	5.38	5.07	3.17			**−0.56**	**−0.66**	**−0.89**	4.96	4.96	4.46
2	**0.00**	**0.00**	**0.00**	3.22	3.47	4.08							−0.13	−0.33	−0.80	**2.53**	**2.74**	3.06
3				4.47	4.67	5.48	**5.16**	**4.94**	3.39				0.48	0.30	−0.27	2.83	2.95	**2.97**
4	2.06	1.94	1.58	3.35	3.54	3.73				5.20	4.93	4.54				3.63	3.76	3.75
6	0.42	0.58	0.58				5.85	5.56	**3.00**									
9	4.26	4.45	5.42															
10	4.19	4.33	4.71															
12	8.52	8.32	9.46															
13	8.32	8.22	7.21															
14				3.66	4.01	4.27				5.55	5.20	4.52						
15	4.41	4.28	2.39															
16	3.28	3.14	2.68															
17				4.55						**4.87**	**4.65**	**4.44**						

[a] The B3LYP/6-31+G(d)//HF/6-31G(d) energy of the most stable rotamer of **1**, −1314.66760 hartree, is chosen as reference. For this rotamer the corrections for the enthalpy, H_{298}, and Gibbs free energy, G_{298}, at 298 K are 0.43021, and 0.34478 hartree, respectively. The scaled (scaling factor: 0.8929) ZPE corrections and thermal corrections to 298 K were obtained from HF/6-31G(d) frequency analysis. ΔH_{298} and ΔG_{298} corresponds to relative enthalpies and relative Gibbs free energies at 298 K, respectively. The most stable rotamers are in bold.

Table 7. HF/6-31G(d), ΔE(HF), and B3LYP/6-31+G(d)//HF/6-31G(d) relative energies, ΔE(DFT) of the selected low energy rotamers of threonine containing compounds.[a]

Rotamer	αGalNAcThr (**7**)		βGalNAcThr (**8**)		αManNAcThr (**9**)		βManNAcThr (**10**)		αGlcNAcThr (**11**)		βGlcNAcThr (**12**)	
	ΔE(HF)	ΔE(DFT)	ΔE(HF)	ΔE(DFT)	ΔE(HF)	ΔE(DFT)	ΔE(HF)	ΔE(DFT)	ΔE(HF)	ΔE(DFT)	ΔE(HF)	ΔE(DFT)
1	0.96	**−0.11**	**3.62**	**2.43**					**−0.32**	**−0.96**		
2	0.34	−0.06									**5.70**	**3.79**
3					**5.04**	**4.72**						6.16
4.14												
6	**0.00**	0.00										
17							**4.55**	**4.18**				

[a] Rotamer 6 of **1** is chosen as reference. The HF/6-31G(d) and B3LYP/6-31+G(d)//HF/6-31G(d) reference energies are −1346.04521 and −1353.98472 hartree, respectively. The most stable rotamers are in bold.

should state that the τ_1 and χ_3 torsional angles are not independent ($\tau_1 \cong \chi_3 + 240^\circ$) and we show both angles just for convenience.

Carbohydrate residue conformations: The orientation of the τ_1 torsion angle is always g ($\approx 60^\circ$ relative to O5) in the α-anomers **1** and **3** as it is expected from the exo-anomeric effect. In the β-anomers **2** and **4** this τ_1 torsion angle can be g or g− in the most stable rotamers (cf. Tables 2–6); thus, the non-exoanomeric rotamers are surprisingly stable. The origin of this is explained in the discussion of χ_2 torsion angle (see below). The orientation of the OH groups of the carbohydrate residue in the most stable rotamers shows the typical pattern frequently found in monosaccharides: a chain of O···H interactions. Several consecutive intraresidue O···H interactions can be observed in the carbohydrate residue (these are gas phase structures, we note that solvation would weaken the intramolecular interactions, and these interactions are also often absent from crystal structures.) In the stable rotamers of the four compounds studied in this paper the $(O3_{Carb})$-H···$O7_{Carb}$ hydrogen bond occurs frequently (cf. Table 8, the O3H donor to oxygen atom of the NAc group). Due to this interaction the hydrogen atom of the NAc group turns toward the serine residue. As noted earlier the $(O3_{Carb})$H···$O7_{Carb}$ hydrogen bond is missing from some of the O-glycosidic core structures. That is because in these core structures the O3 of GalNAc because is linked to other carbohydrates.

As already observed in monosaccharides, the different rotamers of primary alcohol groups yield nearly the same relative energies. The feature that the most stable rotamers have in common is gtttXy where *Xy* symbolizes the various orientations of the primary alcohol group characterized by the τ_5 and τ_6 torsional angles (e.g. G−g, Gg−, Gt or Tg, cf. Tables 2–6.) Figure 7 illustrates the conformational space of the primary alcohol groups in the GalNAc and ManNAc residues (the alternative axial and equatorial orientations are denoted in the Figure). Figure 7 shows that G−g and the Gg− orientations make possible the O6H···O5 interaction. This interaction usually provides some stabilization as it was mentioned earlier. The G−g− orientation in the GalNAc residue provides a possibility for O6H···O4 interaction (this is not possible for ManNAc, cf. Figure 7). The Tg orientation makes the O6H···O4 interaction possible for the GalNAc, GlcNAc, and ManNAc residues. The Gt, G−t, and Tt orientations do not allow O6H intraresidue interactions. These are among the preferred orientations when O6 is linking another carbohydrate residue, although this requires further study.

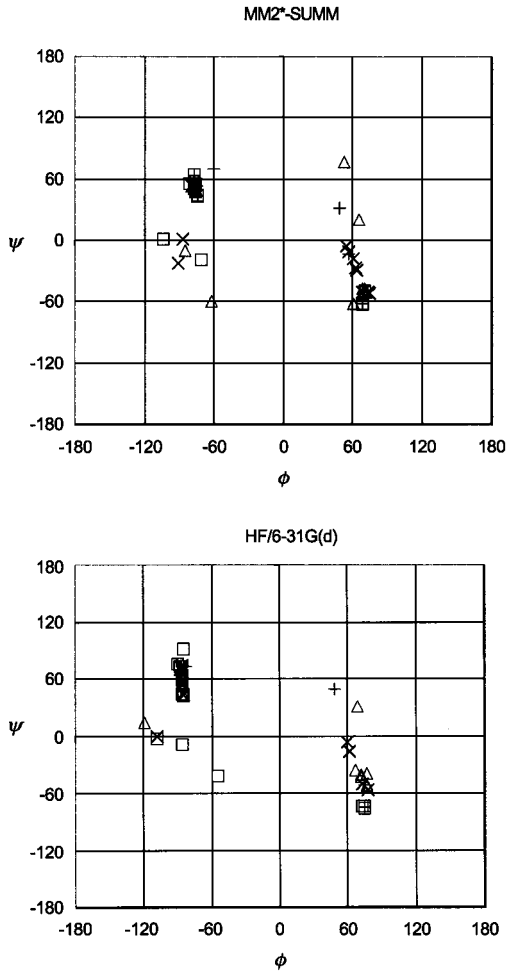


Figure 5. Backbone conformations of αGalNAcSer (+), βGalNAcSer (△), αManNAcSer (□), and βManNAcSer (×), on the Ramachandran map based on MM2* and HF/6-31G(d) calculations.

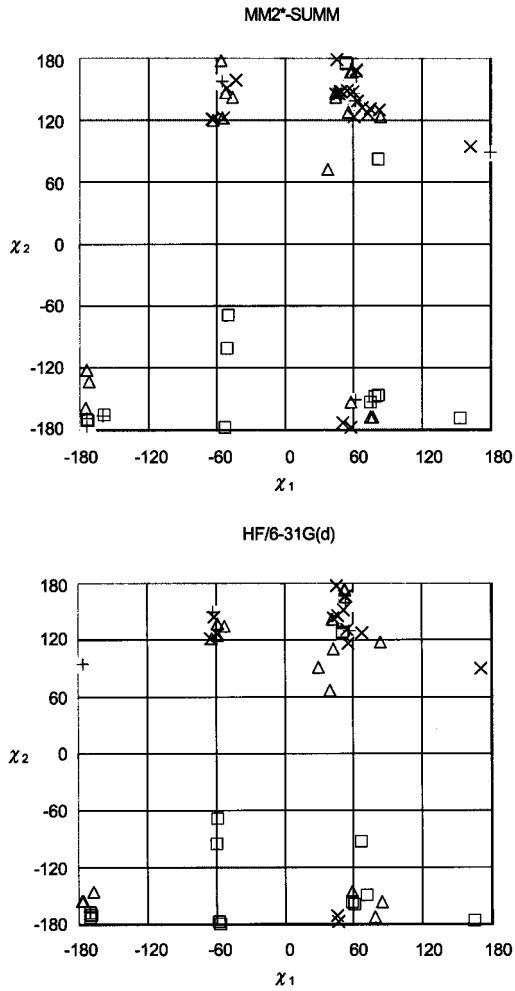


Figure 6. Side chain rotamers of αGalNAcSer (+), βGalNAcSer (△), αManNAcSer (□), and βManNAcSer × on the Ramachandran map based on MM2* and HF/6-31G(d) calculations.

Table 8. Relative frequencies of various types of hydrogen bonds [%] in the most stable HF/6-31G(d) and the MM2* structures.[a]

Type of H-bond	HF/6-31G(d)				MM2*			
	αGal (**1**)	αMan (**3**)	βGal (**2**)	βMan (**4**)	αGal (**1**)	αMan (**3**)	βGal (**2**)	βMan (**4**)
inside the carbohydrate residue:								
$(O3_{Carb})H\cdots O7_{Carb}$	**94.4**	**100.0**	**90.5**	**75.0**	**94.4**	**100.0**	**90.5**	**70.0**
$(O6_{Carb})H\cdots O5_{Carb}$	11.1	23.5	19.0	0.0	11.1	29.4	19.0	0.0
between serine and the carbohydrate residues:								
$(N1_{Ser})H\cdots O5_{Carb}$	0.0	5.9	14.3	15.0	0.0	5.9	9.5	10.0
$(N1_{Ser})H\cdots O6_{Carb}$	0.0	5.9	0.0	0.0	5.6	29.4	9.5	10.0
$(N1_{Ser})H\cdots O7_{Carb}$	5.6	0.0	0.0	25.0	5.6	23.5	0.0	25.0
$(N2_{Ser})H\cdots O6_{Carb}$	0.0	5.9	0.0	0.0	0.0	11.8	0.0	0.0
$(N2_{Carb})H\cdots O1_{Ser}$	11.1	0.0	9.5	10.0	11.1	0.0	9.5	20.0
$(N2_{Carb})H\cdots O2_{Ser}$	**77.8**	**0.0**	**38.1**	**25.0**	**55.6**	**0.0**	**38.1**	**15.0**
$(O6_{Carb})H\cdots O1_{Ser}$	0.0	11.8	9.5	5.0	0.0	11.8	9.5	5.0
$(O6_{Carb})H\cdots O2_{Ser}$	**0.0**	**5.9**	**19.0**	**35.0**	**0.0**	**5.9**	**19.0**	**60.0**
inside the amino-acid residue:								
$(N2_{Ser})H\cdots O1_{Ser}$	**88.9**	**88.2**	**90.5**	**95.0**	**94.4**	**88.2**	**90.5**	**95.0**

[a] The most frequently observed types of hydrogen bonds in the most stable rotamers are in bold.

0947-6539/02/0820-4730 $ 20.00+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 20

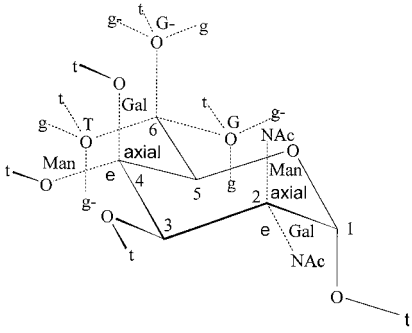


Figure 7. Illustration of the conformational space of the primary alcohol group in the GalNAc, and ManNAc residues, corresponding to the τ_5, and τ_6 torsional angles. The axial (ax) and equatorial (eq) positions are noted in the Figure. The numbering of the carbon atoms follows the convention.

Interresidue hydrogen bonds: Table 8 shows the possible interresidue interactions for four selected molecules studied in the present paper. The Glc analogues **5** and **6** contain the same type of interresidue interactions as **1** and **2**. The Thr analogues **7**–**12** behave similarly to the Ser containing molecules in this respect. The MM2* and the HF/6-31G(d) results differ considerably, the MM2* shows a preference for various seemingly less important hydrogen bonds as $(N1_{Ser})H\cdots O5_{Carb}$, $(N1_{Ser})H\cdots O6_{Carb}$, or $(N2_{Ser})H\cdots O6_{Carb}$. The HF/6-31G(d) results show a preference for $(N2_{Carb})H\cdots O2_{Ser}$ hydrogen bond for **1** (cf. Table 8), providing a well-equilibrated orthogonal position for the saccharide residue. This interresidue interaction is possible if τ_2 is in the g position ($\approx 60^\circ$) for both anomers. Figures 8 and 9 show that this interaction stabilize the most stable rotamers of **1**, **5**, and **6**. Figure 8 shows the essential difference between the most stable rotamer of α- and βGlcNAcSer is in the relative orientation of the carbohydrate relative to serine moiety. In the former the carbohydrate moiety is orthogonal to the serine moiety; however, in the latter it is parallel in a sense shown in Figure 8. The shortest $(N2_{Carb})H\cdots O2_{Ser}$ hydrogen bond distance was calculated for most stable rotamer of **6** to 2.04 Å; this short distance probably contributes to the extra stability of **6**. The most stable rotamers of α- and βGalNAcSer behave similarly. The other interresidue hydrogen bonds usually lead to considerably bent structures in which the distant part of the carbohydrate moiety is linked to the serine residue (e.g. $(O6_{Carb})H\cdots O2_{Ser}$ interaction in ManNAcSer in Table 8). These results suggest that in the most stable GalNAcSer molecule the GalNAc residue is anchored to the serine residue in a way that it takes a special orientation, which provides a suitable non-conflicting foundation for the large carbohydrate antenna. The most stable threonine analogues are similar to the most stable serine rotamers. These results support the experimental observations cited in the introduction.[8, 9]

Conclusion

An MM2*-SUMM conformational search was performed for twelve model compounds of O-glycosides of different constitution in order to understand the origins of the natural preference for *N*-acetyl-*O*-(2-acetamido-2-deoxy-α-D-galactopyranosyl)-L-seryl-*N*′-methyl amide abbreviated as αGalNAcSer. The other structures can be derived from αGalNAcSer by simply changing the corresponding axial (ax) or equatorial (eq) positions either on C1 (ax in the α, and eq in the β-anomer), on C2 (eq in galactose, or glucose and ax in

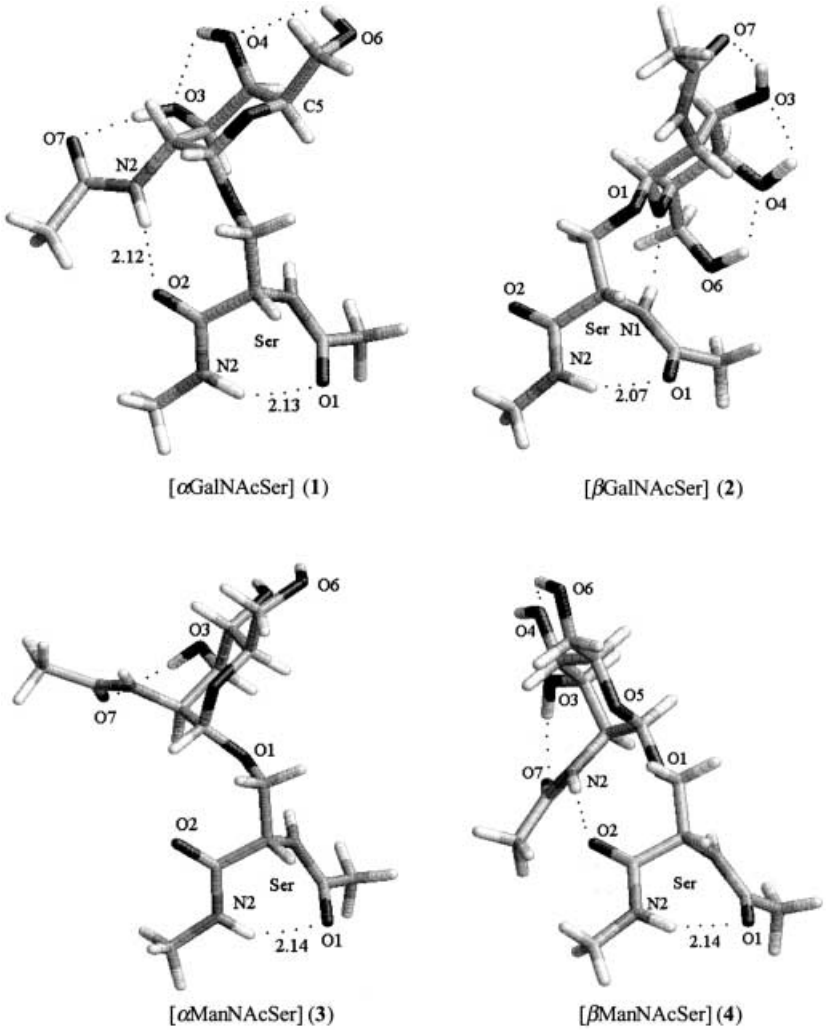


Figure 8. Three-dimensional representation of the fully optimized HF/6-31G(d) global minima for compounds **1**–**4**.

  0947-6539/02/0820-4731 $ 20.00+.50/0

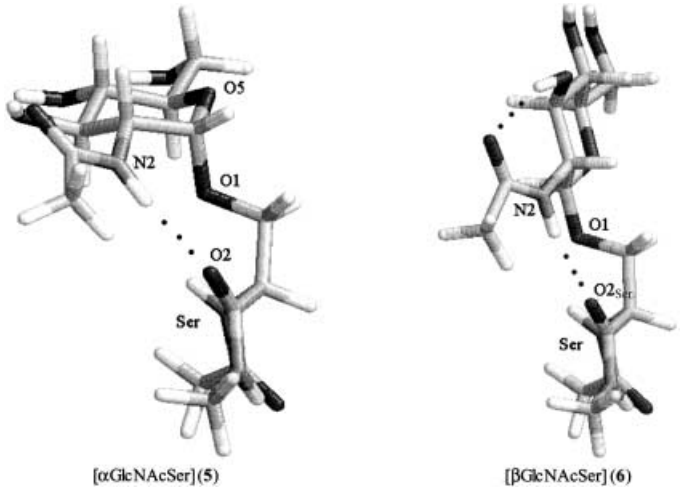


Figure 9. Three-dimensional representation of the fully optimized HF/6-31G(d) global minima for compounds **5** and **6**.

mannose), and on C4 (ax in galactose, and eq in mannose, or glucose) of the carbohydrate residue and replacing the serine by L-threonine. The geometries of nineteen, twenty-one, seventeen, and twenty-four low-energy rotamers of αGalNAcSer, βGalNAcSer, αManNAcSer, βManNAcSer molecules, respectively, were further optimized at HF/6-31G(d) level of theory. Vibrational frequency analyses confirmed each stationary point as a true minimum on the potential energy surface. The glucose analogues were derived from the most stable rotamers of galactose containing derivatives. Single point B3LYP/6-31+G(d)//HF/6-31G(d) energy calculations that include correlation energy were also performed. Electronic energies were converted into relative enthalpies and Gibbs free energies (at 298 K) using the zero-point vibration and thermal corrections in the gas phase. The individual rotamers of the conformational spaces obtained at the different levels of theory were analyzed and compared.

Following conclusions can be drawn from the results:

- There is a substantial difference between the MM2* and HF/6-31G(d) relative energies of the various conformers. Even the most stable structures obtained by MM2* and HF/6-31G(d) methods show considerable differences in the molecular geometry for αGalNAcSer, αManNAcSer, and βManNAcSer.
- The most stable ab initio rotamers of αGalNAcSer can be summarized as γ_L[gtt | gtttXy], where γ_L symbolizes the serine conformation, gtt the serine side chain rotamer positions, the next gtt the first three torsion angles in αGalNAc, and *Xy* symbolizes the various orientations of the primary alcoholic group of αGalNAc characterized by the τ_5, and τ_6 torsional angles (e.g. G – g, Gg – , Gt or Tg). The ab initio total energy difference between these three most stable rotamers is less than 1 kcal mol^{-1}.
- The common feature of the most stable γ_L[gtt | gttt*Xy*] rotamer family of αGalNAcSer is the preference for AcNH $\cdots$ O=C-NHMe, noted as $(N2_{Carb})H \cdots O2_{Ser}$, type hydrogen bonds. This anchors the galactose residue in an orthogonal position with respect to the serine residue (cf. the gtt side chain orientation with two *anti* torsional angles).
- Almost all conformers contain intraresidue the hydrogen bonds: $(N2_{Ser})H \cdots O1_{Ser}$, and $(O3_{Carb})H \cdots O7_{Carb}$.
- Comparison of the calculated energies of the compounds studied in this paper provides that axial position is preferred over equatorial position on C1 (α-anomers are preferred over β-anomers), and equatorial position is preferred over equatorial position on C2 (galactose, or glucose is preferred over mannose). The energy difference caused by the epimerization of C4 is relatively small, consequently the galactose and glucose containing derivatives show similar stability, the latter being slightly more stable. αGlcNAcSer and αGalNAcSer are the most stable compounds according to any method used in this study. The threonine analogues provide similar results. This is in agreement with the natural occurrence of αGalNAcSer/Thr in O-glycosidic core structures; however, it does not explain the lack of αGlcNAcSer/Thr type linkages. To explain this the difference between 3- or 6-substitution of GalNAc and GlcNAc should be studied. The most stable β-anomer is the βGlcNAcSer. The energetic differences are in the range of 4-6 kcal mol^{-1} according to the HF/6-31G(d) method. The single point B3LYP/6-31+G(d)//HF/6-31G(d) calculations provide only slightly different, 2.5 – 5 kcal mol^{-1} energy difference.
- ΔH_{298} and ΔG_{298} data for gas phase provide that the effect of the ZPE and thermal corrections are relatively small ($<$ 0.4 kcal mol^{-1}), alhtough the $T\Delta S_{298}$ term can be as large as 2.6 kcal mol^{-1}. The entropy terms stabilize the α-anomers relative to β-anomers, and ManNAc relative to GalNAc. The largest stabilization effect can be observed for one of the rotamers of α-anomer of ManNAc; however, αGalNAcSer remains the most stable by about 3 kcal mol^{-1}.
- The common feature of the most stable γ_L[gtt | gtttXy] rotamer families of αGalNAcSer and αGlcNAcSer is the preference for AcNH $\cdots$ O=C-NHMe (noted as $(N2_{Carb})$-$H \cdots O2_{Ser}$) type hydrogen bond. This hydrogen bond anchores the αGalNAc or αGlcNAc residues in an orthogonal position with respect the serine residue (cf. gtt side chain orientation with two *anti* torsional angles). The same hydrogen bond anchores the βGalNAc or βGlcNAc residues in a "parallel" position with respect the serine residue in the gas phase.

This theoretical work with other experimental studies could provide some explanation of why mutation and alteration are rare in the linker region of glycoproteins, although further studies are necessary in order to understand the preference for αGalNAc over αGlcNAc. Additional studies for solution or enzymatic catalysis can use these gas phase energies, enthalpies, and Gibbs free energies as a starting point.

Acknowledgement

Alfred French is thanked for comments on the preprint of this work. István Kolossváry is thanked for the MMFF94 data. The financial support of the Hungarian Scientific Research foundation is gratefully acknowledged (OTKA T034764, and T034299).

  0947-6539/02/0820-4732 $ 20.00+.50/0

[1] H. Helzner, T. Reipen, M. Schultz, H. Kunz, *Chem. Rev.* **2000**, *100*, 4495.

[2] a) R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683; b) E. F. Hounsell, M. J. Davies, D. V. Renouf, *Glycoconjugate J.* **1996**, *13*, 19; c) J. F. G. Vliegenthart, F. Casset, *Curr. Opin. Struct. Biol.* **1999**, *8*, 565.

[3] R. R. Koganty, M. A. Reddish, B. M. Longenecker in *Glycopeptides and related compounds* (Eds.: D. G. Large, C. D. Warren), Marcel Dekker, New York, **1997**.

[4] O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins, see: R. Gupta, H. Birch, K. Rapacki, S. Brunak, J. E. Hansen, *Nucleic Acids Res.* **1999**, *27*, 370.

[5] a) P. Hallgren, A. Lundblad, S. Svensson, *J. Biol. Chem.* **1975**, *250*, 5312; b) A. L. Rosenthal, I. M. Noedin, *J. Biol. Chem.* **1975**, *250*, 5295; c) D. H. Miller, D. T. A. Lamport, M. Miller, *Science* **1972**, *176*, 918.

[6] R. S. Haltiwanger, W. G. Kelly, E. P. Roquemore, M. A. Blomberg, L.-Y. D. Dong, L. Kreppel, T.-Y. Chou, K. Greis, G. W. Hart, *Biochem. Soc. Trans.* **1992**, *20*, 264.

[7] J. Broddefalk, J. Backlund, F. Almqvist, M. Johansson, R. Holmdahl, J. Kihlberg, *J. Am. Chem. Soc.* **1998**, *120*, 7676.

[8] D. H. Live, L. J. Williams, S. D. Kuduk, J. B. Schwarz, P. W. Glunz, X.-T. Chen, D. Sames, R. A. Kumar, S. J. Danishefsky, *Proc. Nat. Acad. Sci. USA* **1999**, *96*, 3489.

[9] S. E. O'Connor, B. Imperiali, *Chem. Biol.* **1998**, *5*, 427.

[10] L. Stryer, *Biochemistry*, W. H. Freeman, New York, **1995**.

[11] M. Hollósi, A. Perczel, G. D. Fasman *Biopolymers* **1990**, *29*, 1549.

[12] Perczel, E. Láng, M. Hollósi, G. D. Fasman, *Biopolymers* **1993**, *33*, 665.

[13] W. C. Still, MacroModel 5.0, Columbia University, USA.

[14] a) N. L. Allinger, *J. Am. Chem. Soc.* **1977**, *99*, 8127; b) different versions are available from Quantum Chemistry Program Exchange, University of Indiana, Bloomington, IN 47405 (USA), http://ccl.osc.edu/ccl/qcpe/QCPE/catalog.html

[15] K. Gundertofte, T. Liljefors, P.-O. Norrby, I. Pettersson, *J. Comput. Chem.* **1996**, *17*, 429.

[16] T. A. Halgren, R. B. Nachbar, *J. Comput. Chem.* **1996**, *17*, 587.

[17] J. M. Goodman, W. C. Still, *J. Comput. Chem.* **1991**, *12*, 1110.

[18] a) H. B. Schlegel, *J. Comput. Chem.* **1982**, *3*, 214; b) C. Peng, P. Y. Ayala, H. B. Schlegel, M. J. Frisch, *J. Comput. Chem.* **1996**, *17*, 49.

[19] *Gaussian 98*, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, J. A. Pople, Gaussian, Inc., Pittsburgh, PA, **1998**.

[20] G. I. Csonka, I. Kolossváry, P. Császár, K. Éliás, I. G. Csizmadia, *J. Mol. Struct. (THEOCHEM)* **1997**, *395–396*, 29.

[21] J. H. Lii, B. Y. Ma, N. L. Allinger, *J. Comput. Chem.* **1999**, *20*, 1593.

[22] a) G. I. Csonka, http://web.inc.bme.hu/mols/gal/hf6geom.htm; b) G. I. Csonka, http://web.inc.bme.hu/mols/man/hf6geom.htm; c) G. I. Csonka, http://web.inc.bme.hu/mols/glc/hf6geom.htm

[23] Y. Nishida, H. Ohrui, H. Meguro, *Tetrahedron Lett.* **1984**, *25*, 1575.

[24] Y. Nishida, H. Hori, H. Ohrui, H. Meguro, *J. Carbohydr. Chem.* **1988**, *7*, 239.

[25] I. Tvaroska, F. R. Taravel, J. O. Utille, J. P. Carver, *Carbohydr. Res.* **2002**, *337*, 353.

[26] S. E. Barrows, J. W. Strorer, C. J. Cramer, A. D. French, D. G. Truhlar, *J. Comput. Chem.* **1998**, *19*, 1111.

[27] G. I. Csonka, http://web.inc.bme.hu/mols/glc/hf6geom.htm

[28] G. M. Brown, H. A. Levy, *Acta Crystallogr. Sect. B* **1979**, 656.

[29] G. I. Csonka, K. Éliás, I. G. Csizmadia, *Chem. Phys. Lett.* **1996**, *257*, 49.

[30] G. I. Csonka, K. Éliás, I. G. Csizmadia, *J. Comput. Chem.* **1997**, *18*, 330.

[31] G. I. Csonka, K. Éliás, I. Kolossváry, C. P. Sosa, I. G. Csizmadia, *J. Phys. Chem. A* **1998**, *102*, 1219.

[32] W. Damm, A. Frontera, J. Tirado-Rives, W.-L. Jorgensen, *J. Comput. Chem.* **1997**, *18*, 1955.

[33] S. Reilig, M. Schlenkrich, J. Brinkmanmn, *J. Comput. Chem.* **1996**, *17*, 450.

[34] H. Senderowitz, C. Parish, W. C. Still, *J. Am. Chem. Soc.* **1996**, *118*, 2078.

[35] J. R. Kneisler, N. L. Allinger, *J. Comput. Chem.* **1996**, *17*, 757.

[36] I. Kolossváry, W. C. Guida, *J. Am. Chem. Soc.* **1996**, *118*, 501. The LMOD procedure is available in the 6.0 or later release of MacroModel.

[37] G. I. Csonka, http://web.inc.bme.hu/mols/galnacser/

[38] F. H. Allen, J. E. Davis, J. J. Galloy, O. Johnson, O. Kennard, C. F. Macrae, E. M. Mitchell, G. F. Mitchell, J. M. Smith, D. G. Watson, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187.

[39] A. D. French, personal communication.

[40] a) A. Perczel, M. Hollósi, V. Fülöp, A. Kálmán, P. Sándor, G. D. Fasman, *Biopolymers* **1990**, *30*, 763; b) R. A. Shaw, A. Perczel, G. D. Fasman, H. H. Mantsch, *Int. J. Peptide Protein Res.* **1996**, *48*, 71.

[41] a) J. N. Scardsdale, C. Van Alsenoy, V. J. Klimkowski, L. Schafer, F. A. Momamany, *J. Am. Chem. Soc.* **1983**, *105*, 3438; b) A. Perczel, R. Daudel, J. G. Ángyán, I. G. Csizmadia, *Can. J. Chem.* **1990**, *68*, 1182, c) A. Perczel, Ö. Farkas, I. Jákli, I. G. Csizmadia, *J. Mol. Struct. (TEOCHEM)* **1998**, *455*, 315; d) I. Jákli, A. Perczel, Ö. Farkas, A. G. Császár, C. P. Sosa, I. G. Csizmadia, *J. Comput. Chem.* **2000**, *21*, 626; e) A. Perczel, I. G. Csizmadia in *The amide linkage, structural significance in chemistry, biochemistry and materials sciences* (Eds.: A. Greenberg, C. M. Breneman, J. F. Liebman), Wiley, New York, **2000**, pp. 409.

[42] A. Perczel, Ö. Farkas, I. G. Csizmadia, *J. Am. Chem. Soc.* **1996**, *118*, 7809.

Received: February 6, 2002
Revised: April 12, 2002 [F3859]

  0947-6539/02/0820-4733 $ 20.00+.50/0