# Ab Initio Molecular Orbital Study of the Catalytic Mechanism of Glycosyltransferases: Description of Reaction Pathways and Determination of Transition-State Structures for Inverting *N*-Acetylglucosaminyltransferases

**Igor Tvaroska,*,† Isabelle André, and Jeremy P. Carver**

*Contribution from GlycoDesign Inc., 480 University Avenue, Suite 900, Toronto, Ontario, Canada, M5G 1V2*

*Received May 2, 2000*

**Abstract:** The catalytic mechanism of the GlcNAc transfer by inverting *N*-acetylglucosaminyltransferases is explored for the first time using ab initio quantum chemical methods. The structural model describing the active site where the reaction occurs consists of all essential molecules or their fragments assumed to be involved in the mechanism: a complete sugar-donor molecule, UDP-GlcNAc, a hydroxyl group of the oligosaccharide-acceptor modeled by methanol, a divalent metal cofactor represented by $Mg^{2+}$, as well as the essential parts of the catalytic acid (A) and catalytic base (B) modeled by acetic acid and acetate molecules. Different possible mechanisms of reaction have been followed by means of several two-dimensional potential energy maps calculated as a function of predefined reaction coordinates. Potential energy surfaces calculated at the HF/6-31G* level revealed 11 transition states and five intermediates associated with distinct possible reaction pathways. All stationary points, transition states, and intermediates, were characterized at HF/6-31G*, HF/6-31++G**//HF/6-31G*, DFT/B3LYP/6-31G*, and DFT/B3LYP/6-31++G**//DFT/B3LYP/6-31G* levels. A detailed description of the reaction pathways that includes energetic evaluations and the structural modifications of the different participants occurring along the catalytic process is given, followed by a discussion on their feasibility, consequences, and implications for the catalytic mechanism. Among the different reaction pathways, a stepwise reaction pathway assuming the enrolment of only a catalytic base appeared to be the most probable reaction path and consistent with the existing experimental data.

## 1. Introduction

Glycosyltransferases (GT's, a general nomenclature for glycosyltransferases is EC 2.4.*x.y*) comprise a group of enzymes that are involved in the biosynthesis of complex oligosaccharides.[1−4] The result of the reaction catalyzed by these enzymes is the formation of a new glycosidic linkage, and it appears that there is at least one distinct glycosyltransferase for every type of glycosidic linkage. Glycosylation proceeds in a stepwise manner, and therefore, the expression and specificity of the enzymes represent key regulatory factors in defining the repertoire of biosynthesized oligosaccharides. During oligosaccharide processing, oligosaccharides are converted into hybrid and complex oligosaccharides by the addition of *N*-acetylglucosaminyl residues (GlcNAc, 2-acetamido-2-deoxy-α-D-glucopyranosyl). These modifications in the oligosaccharide chains of *N*- and *O*-linked glycoproteins accompany many physiological and pathological cell processes.[5] The transfer of GlcNAc is catalyzed by *N*-acetylglucosaminyltransferases (GlcNAc-T's or GnT's). In such a transfer, the donor of the GlcNAc residue is UDP-GlcNAc [uridine 5′-(2-acetamido-2-deoxy-α-D-glucopyranosyl pyrophosphate)] while the acceptor is one of the hydroxyl groups located at a particular position of a variety of oligosaccharides. *N*-acetylglucosaminyltransferases show a decisive specificity for the oligosaccharide-acceptor and they generally require the presence of a metal cofactor.[6] There are at least eight different GlcNAc-T's involved in the biosynthesis of complex and hybrid *N*-glycans (GlcNAc-T I−GlcNAc-T VIII), five in the biosynthesis of *O*-glycans (Core 2−Core 4 GnT's, Core 1 and Core 2 elongation GnT's), and two in the biosynthesis of antigen determinants (blood group i and blood group I) (Scheme 1).[2,4,7] Though some of these GlcNAc-T's have already been cloned, the origin of their specificity remains unknown due to the lack of experimental structures of GlcNAc-T's or any other mammalian glycosyltransferase.

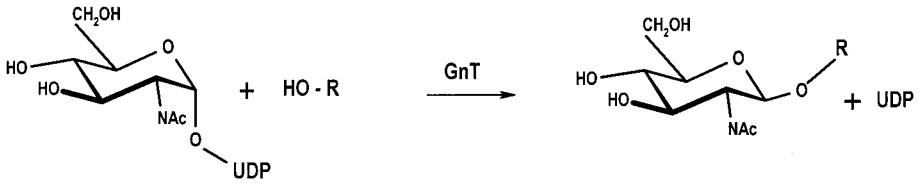
Despite their wide abundance in nature, crystal structures of glycosyltransferases are rare. Until recently, the only available structure of a glycosyltransferase was that of a DNA-modifying β-glucosyltransferase from bacteriophage T4 and its complex with UDP-Glc.[8] However, that enzyme is somewhat different

† On leave of absence from the Institute of Chemistry, Slovak Academy of Sciences, Bratislava, Slovak Republic.

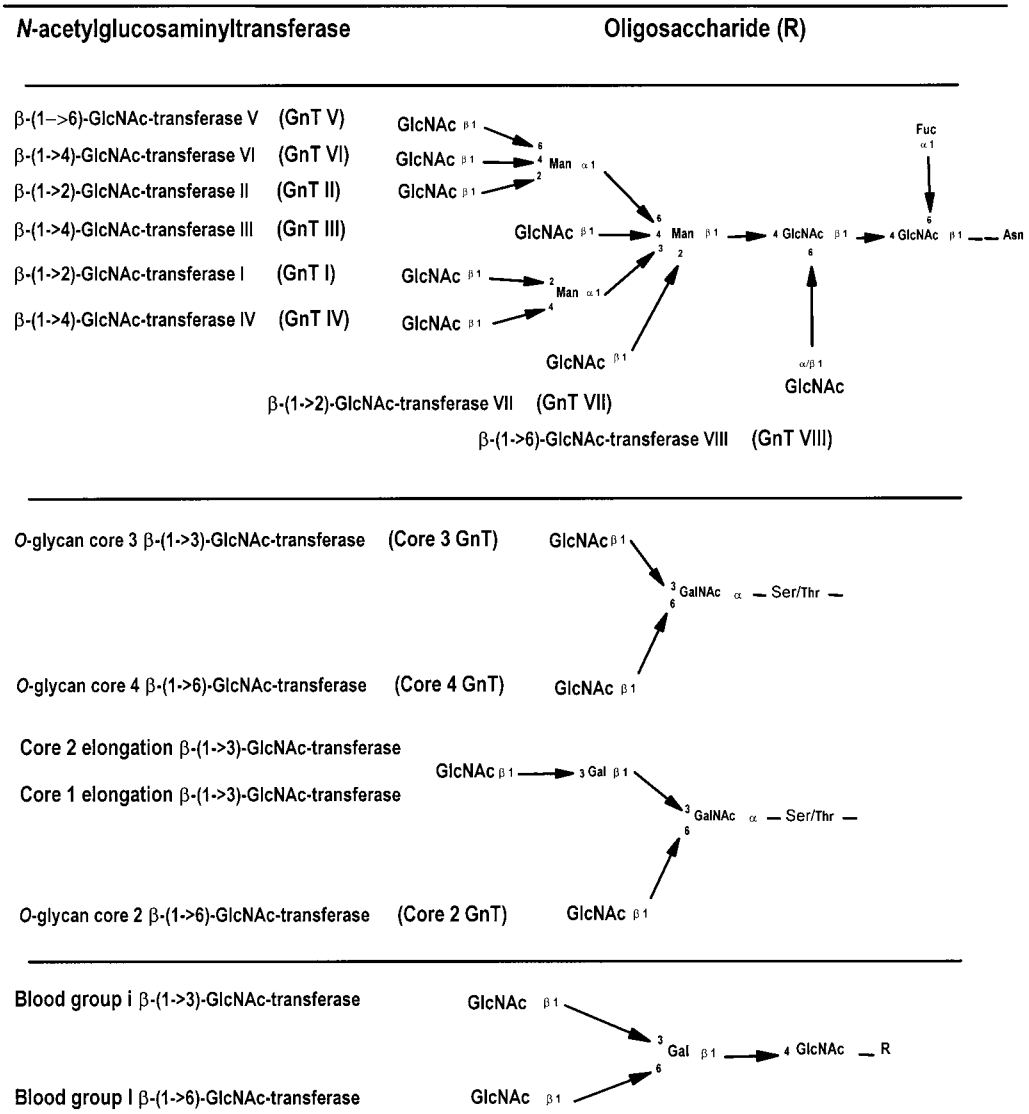
(1) Beyer, T. A.; Sadler, J. E.; Rearick, J. I.; Paulson, J. C.; Hill, R. L. *Adv. Enzymol.* **1981**, *52*, 23−175.
(2) Schachter, H. *Curr. Opin. Struct. Biol.* **1991**, *1*, 755−765.
(3) Kleene, R.; Berger, E. G. *Biochim. Biophys. Acta* **1993**, *1154*, 283−325.
(4) Montreuil, J.; Vliegenthart, J. F. G.; Schachter, H. *Glycoproteins*; Neuberger, A., van Deenen, L. L. M., Eds.; Elsevier: Amsterdam, 1995; Vol. 29a.
(5) Montreuil, J.; Vliegenthart, J. F. G.; Schachter, H. *Glycoproteins and Disease*; Neuberger, A., van Deenen, L. L. M., Eds.; Elsevier: Amsterdam, 1995; Vol. 30.
(6) Bendiak, B.; Schachter, H. *J. Biol. Chem.* **1987**, *262*, 5784−5790.
(7) Raju, T. S.; Stanley, P. *J. Biol. Chem.* **1998**, *273*, 14090−14098.
(8) Vrielink, A.; Ruger, W.; Driessen, H. P. C.; Freemont, P. S. *EMBO J.* **1994**, *13*, 3413−3422.

10.1021/ja001525u CCC: $19.00 
Published on Web 08/23/2000

**Scheme 1.** Schematic Representation of the *N*-Acetylglucosaminyltransferases Involved in the Biosynthesis of *N*-Glycans (GlcNAc-T I−VIII), *O*-Glycans (Core 2−4 and Core 1−2 Elongation GnT's), and Antigen Determinants (Blood Groups i and I)




*(UDP - uridine diphosphate; R - oligosaccharide part of glycoprotein)*




from other glycosyltransferases since the acceptor involved in the reaction with this enzyme is not a carbohydrate. Indeed, this enzyme catalyses the transfer of a glucose moiety from UDP-glucose to hydroxymethylated cytosines of DNA. The DNA-modifying $\beta$-glucosyltransferase from bacteriophage T4 presents no sequence homology with any other glycosyltransferase,[9] although the structure of this enzyme has been used as a template to predict the structure of other glycosyltransferases.[10] In the past few months, a decisive breakthrough in this field has been achieved with the resolution of the X-ray structures of two bacterial glycosyltransferases in their native and nucleotide-complexed forms, the SpsA,[11] for which the substrate specificity is undefined, and the $\beta$-1,4-galactosyltransferase T1.[12]

The reaction catalyzed by GlcNAc-T's can be regarded as a nucleophilic displacement of the UDP (uridine 5′-pyrophosphate) functional group at the anomeric carbon C1 of the GlcNAc (2-acetamido-2-deoxy-α-D-glucopyranose) residue of UDP-GlcNAc by a hydroxyl group of a specific oligosaccharide-acceptor (Scheme 1). The enzymatic reaction of all known GlcNAc-T's, except the α-1,4-GlcNAc-T,[13] leads to an inversion of the anomeric configuration. There is a clear resemblance

(9) Campbell, J. A.; Davies, G. J.; Bulone, V.; Henrissat, B. *Biochem. J.* **1997**, *326*, 929−942.

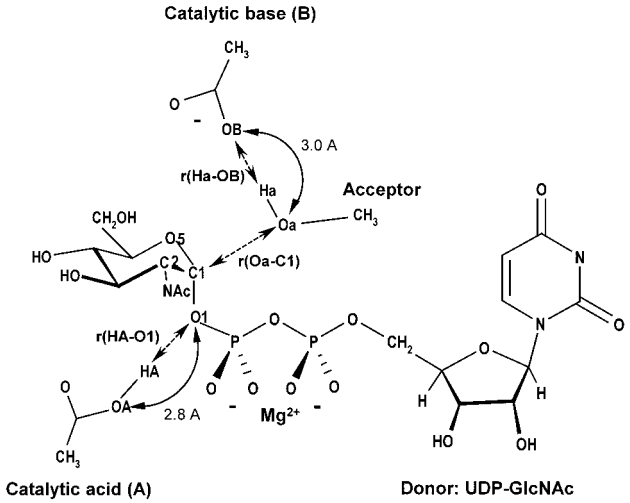
(10) Imberty, A.; Monier, C.; Bettler, E.; Morera, S.; Freemont, P.; Sippl, M.; Flockner, H.; Ruger, W.; Breton, C. *Glycobiology* **1999**, *9*, 713−722.

(11) Charnock, S. J.; Davies, G. J. *Biochemistry* **1999**, *38*, 6380−6385.

(12) Gastinel, L. N.; Cambillau, C.; Bourne, Y. *EMBO J.* **1999**, *18*, 3546−3557.

(13) Coutinho, P. M.; Henrissat, B. Carbohydrate-Active Enzymes server at URL: http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html.

**Scheme 2.** Schematic Representation of the Structural Model Used To Describe the GlcNAc Transfer by Inverting *N*-Acetylglucosaminyltransferases




between the enzymatic action of glycosyltransferases and the enzymatic action of glycoside hydrolases, the mechanisms of which have largely been characterized in detail.[14−21] Many aspects of the functions and catalytic mechanisms of *N*-acetylglucosaminyltransferases are, however, still unknown since only a few mechanistic studies on *N*-acetylglucosaminyltransferases have been reported to date.[6,22] In the absence of experimental data, high-level ab initio calculations can be used to gain some insight into many characteristics of the enzymatic reaction catalyzed by *N*-acetylglucosaminyltransferases. They can provide a description on an atomic level of the discrete intermediates and transition states found along the enzyme-catalyzed reaction pathway. In this work, we report high-level ab initio quantum chemical results on a model of the GlcNAc transfer reaction catalyzed by *N*-acetylglucosaminyltransferases.

## 2. Model and Computational Procedures

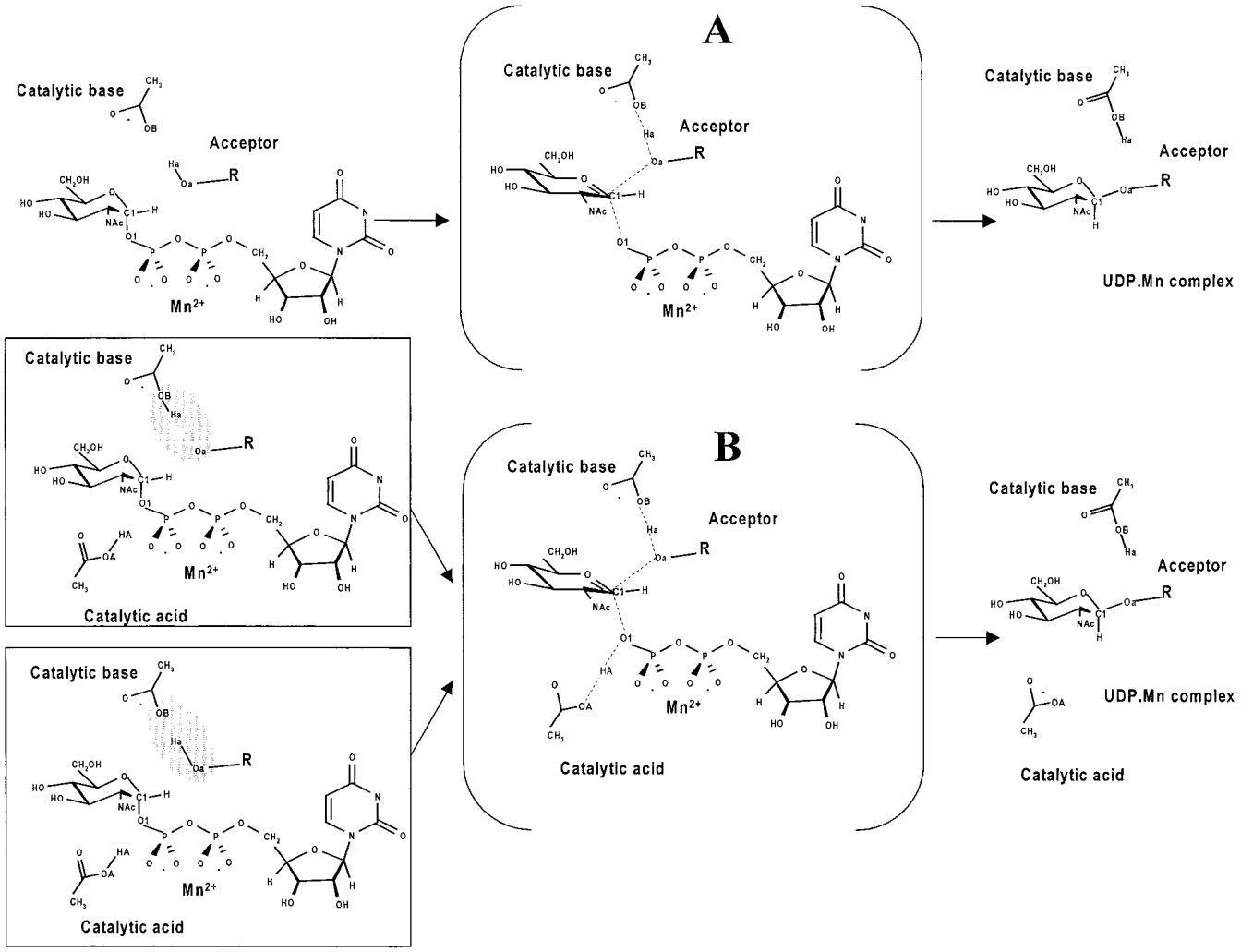
In this investigation, the structural model used to analyze computationally the enzymatic mechanism of the GlcNAc transfer by an *N*-acetylglucosaminyltransferase consists of all the essential molecules, or their fragments, assumed to be involved in the reaction (Scheme 2). The reaction site model contains: the complete sugar-donor molecule, UDP-GlcNAc; the hydroxyl group of the oligosaccharide-acceptor modeled by methanol; a divalent metal cofactor modeled by $Mg^{2+}$; as well as the essential parts of the catalytic acid (A) and catalytic base (B) represented by acetic acid and acetate molecules. Such a model of the active site allows all of the required electronic rearrangements occurring during the enzymatic reaction, such as the proton transfers between the active site components and the substrates. This model consists of 86 atoms and has an overall charge of minus one. In the construction of this model, the relative position of the different participants was an important issue that could not be restricted by the usual means of crystallographic data since no structure of an *N*-acetylglucosaminyltransferase complexed with the entire UDP-GlcNAc substrate was available when we initiated our study. As a consequence, the conformation of the UDP-GlcNAc used in our model is based on previous extensive calculations on sugar−phosphate and diphosphate models.[23−25] The two catalytic amino acids present in our model were placed in an arrangement that emulates their orientation in the active site of inverting glycosyl hydrolases[16] where the two carboxylates are located 6.5 to 9.5 Å apart. In our model, the two amino acids are located about 5.0 Å away from the anomeric carbon C1. The methanol oxygen atom Oa, representing the reactive hydroxyl of the sugar-acceptor, was initially placed at 3.0 Å from the anomeric carbon C1 and at 3.0 Å from the oxygen OB of the amino acid noted B in Scheme 2. Geometrical constraints applied to fix the relative positions of the enzyme residues with respect to the substrates are another important element to consider for building a physically meaningful model. Because the whole structure of the enzyme is not used in our model, these constraints are essential to prevent movement of amino acid residues to unrealistic positions with respect to the substrates. Therefore, the positions of the relevant oxygen atoms and the relative orientation of both the catalytic base and catalytic acid have been restricted in our model. However, the remaining internal geometrical parameters of all molecules present in the model were optimized to reflect geometrical changes occurring during the reaction.

In addition to the nucleophilic attack, the transfer of either one or two protons can be involved in the catalytic reaction of GlcNAc-T's. Consequently, three distances were used as reaction coordinates to follow the mechanism (Scheme 2): the distance $r_{HA-O1}$ between the proton HA of the catalytic acid and the glycosidic oxygen O1; the distance $r_{Ha-OB}$ between the proton Ha of the sugar-acceptor and the oxygen OB of the catalytic base; and the $r_{C1-Oa}$ distance between the anomeric carbon C1 and the oxygen Oa of the acceptor hydroxyl group. These geometrical parameters reflect the proton-transfer process from the catalytic acid to the sugar-donor, the proton-transfer process from the sugar-acceptor to the catalytic base, and the nucleophilic attack of the sugar-acceptor on the anomeric C1−O1 linkage. The energy of the model calculated as a function of these three reaction-coordinates gives the potential energy surface (PES). Each calculated point on the PES corresponds to the optimized structure and arrangement of the model for the given $r_{HA-O1}$, $r_{Ha-OB}$, and $r_{C1-Oa}$ distances. These distances were varied by 0.2 Å increments within the 0.9 to 2.1 Å range for $r_{Ha-OB}$, 0.9 to 1.9 Å range for $r_{HA-O1}$, and within the 3.0 to 1.3 Å range for $r_{C1-Oa}$. During the optimization, all geometrical parameters of the reactants were optimized with the exception of those defining the location and orientation of the two catalytic amino acids. As a result, each point on the PES represented by fixed values of the $r_{HA-O1}$, $r_{Ha-OB}$, and $r_{C1-Oa}$ distances have all their geometrical variables adjusted to their most stable values. Since the calculation of such maps requires an extremely large amount of CPU time, we have divided the calculation of the PES's into two parts. The first PES corresponds to the energy calculated as a function of the $r_{Ha-OB}$ and $r_{C1-Oa}$ distances, whereas for the second PES, the energy was computed as a function of the $r_{HA-O1}$ and $r_{C1-Oa}$ distances. The location of the local minima and transition barriers on the PES's is only approximate, and for that reason a further optimization of the stationary points with no constraints on the $r_{Ha-O1}$, $r_{Ha-OB}$, and $r_{C1-Oa}$ distances is required. These stationary points represent structures of the intermediates and transition states found on the different PES's and along the different reaction pathways. However, to avoid any confusion, we will use throughout the paper the same acronyms: TSi, and INTi respectively for the barriers located on PES's and for the stationary points later refined.

The ab initio calculations were carried out with the Jaguar program.[26] The optimization of the geometry was performed at the SCF level with

(14) Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171−1202.
(15) Legler, G. *Carbohydr. Res.* **1993**, *250*, vii-xx.
(16) Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853−859.
(17) Withers, S. G.; Aebersold, R. *Protein Sci.* **1995**, *4*, 361−372.
(18) Davies, G.; Sinnott, M. L.; Withers, S. G. Glycosyl transfer. In *Comprehensive Biological Catalysis*; Sinott, M. L., Ed.; Academic Press Limited: New York, 1997; Chapter 3, pp 119−208.
(19) Zechel, D. L.; Withers, S. G. Mechanisms of glycosyl transfer. In *Comprehensive Natural Products Chemistry*; Poulter, C. D., Ed.; Elsevier: New York, 1999; Chapter 5, pp 279−314.
(20) Withers, S. G.; Namchuk, M.; Mosi, R. *Iminosugars as glycosidase inhibitors*; Stutz, A. E., Ed.; Wiley-VCH: Weinheim, 1999; pp 188−206.
(21) Heightman, T. D.; Vasella, A. T. *Angew. Chem., Int. Ed.* **1999**, *38*, 750−770.
(22) Nishikawa, Y.; Pegg, W.; Paulsen, H.; Schachter, H. *J. Biol. Chem.* **1988**, *263*, 8270−8281.
(23) Tvaroska, I.; André, I.; Carver, J. P. *J. Mol. Struct. (THEOCHEM)* **1999**, *469*, 103−114.
(24) Tvaroska, I.; André, I.; Carver, J. P. *J. Phys. Chem. B* **1999**, *103*, 2560−2569.
(25) André, I.; Tvaroska, I.; Carver, J. P. *J. Phys. Chem. A* **2000**, *104*, 4609−4617.
(26) Jaguar 3.5, S., Inc., Portland, OR, 1998.

**Scheme 3.** Schematic Representation of the Two Different Types of Mechanism Investigated for the Transfer of GlcNAc by Inverting *N*-Acetylglucosaminyltransferases[a]




[a] Mechanism A involves only a catalytic base, while two catalytic amino acids are implicated in mechanism B.

the 6-31G* (834 basis functions) basis set. Full optimizations were accomplished using the gradient optimization routines of the program without any symmetry constraints. To better characterize the individual reaction paths, the location and structure of the different transition states were calculated using the three nearest points to the particular barrier on the PES using the QST-guided search of the Jaguar software.[26] The geometries of all stationary points on PES's were then fully optimized using the 6-31G* basis set. The effects of electron correlation on the potential energy surface were estimated using the B3LYP density functional method.[27] Ultimately, selected geometries were used to estimate the effect of the basis set by calculating their single point energy with the 6-31++G** basis set (1178 basis functions).

### 3. Results and Discussion

The only experimental data available to date on the mechanism of *N*-acetylglucosaminyltransferases are kinetic studies on GlcNAc-T I and GlcNAc-T II.[6,22] They indicate an ordered sequential mechanism and the prerequisite of a metal cofactor for the enzyme activity. A metal cofactor has been shown to be required by many of these *N*-acetylglucosaminyltransferases, presumably to increase the leaving ability of the pyrophosphate group. This metal ion binds to the enzyme prior to the donor nucleotide-sugar. Then, before any of the products leaves the enzyme, a sugar-acceptor has to be bound to proceed to the transfer of the sugar. For some galactosyltransferases,[28] the metal cofactor was found to be released from the enzyme in the form of a complex with the nucleotide-diphosphate.

By analogy with the reaction mechanism of the inverting glycosyl hydrolases,[18−20] one can assume two different types of mechanism for the inverting GlcNAc-T's (Scheme 3). In the first type (Scheme 3A), only a carboxylate acting as a general base catalyst is involved in the catalytic mechanism. Such a catalytic base assists the nucleophilic attack of the acceptor oxygen Oa on the anomeric carbon C1 of the donor to form a new glycosidic linkage C1−Oa. A pair of carboxylic acids is involved in the second type of catalytic mechanism (Scheme 3B). This mechanism consists of an electrophilic attack of a carboxylic acid on the target oxygen O1 of the donor that cleaves the bond, followed by the nucleophilic attack of the acceptor oxygen Oa on the anomeric carbon C1 of the donor. Here, one catalytic acid behaves as a general acid catalyst protonating the glycosidic oxygen atom O1, while the second carboxylate acts as a general base catalyst deprotonating the nucleophilic oxygen Oa of the acceptor. Both types of catalytic reaction may proceed via one or several transition states and, in terms of the course of events, both mechanisms can proceed either in a concerted or stepwise manner. In the absence of experimental information

(27) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(28) Khatra, B. S.; Herries, D. G.; Brew, K. *Eur. J. Biochem.* **1974**, *44*, 537−560.

**Table 1.** Ab Initio Calculated Geometrical Parameters of the Points observed on PES's Described on Figures 1−3 at the HF/6-31G* and DFT/B3LYP/6-31G* Levels

| | bond lengths | | | | | | | | | | bond angles | | torsional angles | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cl−Oa | | Ha−Oa | | $H_A$−O1 | | Cl−O1 | | Cl−O5 | | Cl−O5−C5 | | Cl−O5−C5−C4 | |
| conformer | HF | B3LYP | HF | B3LYP | HF | B3LYP | HF | B3LYP | HF | B3LYP | HF | B3LYP | HF | B3LYP |
| R | 2.800 | 2.800 | 0.959 | 0.985 | 1.847 | 1.818 | 1.478 | 1.519 | 1.359 | 1.371 | 120.2 | 119.8 | 58.3 | 59.3 |
| INT1 | 1.551 | 1.532 | 0.994 | 1.055 | 1.832 | 1.801 | 2.985 | 2.829 | 1.333 | 1.369 | 125.9 | 124.3 | 9.7 | 46.5 |
| INT2 | 2.800 | 2.800 | 2.026 | 1.981 | 1.848 | 1.848 | 1.492 | 1.529 | 1.368 | 1.381 | 120.1 | 120.5 | 56.4 | 54.2 |
| INT3 | 2.800 | 2.800 | 2.044 | 2.021 | 0.990 | 1.112 | 2.731 | 2.663 | 1.230 | 1.253 | 127.4 | 125.4 | 20.4 | 21.0 |
| INT4 | 2.800 | 2.800 | 0.961 | 0.988 | 0.987 | 1.038 | 3.288 | 3.733 | 1.233 | 1.259 | 125.5 | 122.2 | 42.4 | 48.0 |
| INT5 | 1.502 | 1.653 | 1.000 | 1.032 | 0.988 | 1.035 | 3.306 | 3.467 | 1.350 | 1.330 | 125.1 | 123.6 | 43.4 | 7.5 |
| TS1 | 2.003 | 2.158 | 0.973 | 1.000 | 1.834 | 1.806 | 2.686 | 2.535 | 1.277 | 1.290 | 124.3 | 123.6 | 51.3 | 52.2 |
| TS2 | 1.447 | 1.499 | 1.380 | 1.361 | 1.832 | 1.800 | 3.009 | 3.014 | 1.371 | 1.379 | 123.8 | 124.9 | 47.6 | 27.5 |
| TS3 | 2.800 | 2.800 | 1.599 | 1.638 | 1.848 | 1.848 | 1.489 | 1.524 | 1.364 | 1.378 | 119.6 | 120.0 | 56.5 | 56.6 |
| TS4 | 2.490 | 2.426 | 2.030 | 1.993 | 1.848 | 1.811 | 1.821 | 1.863 | 1.301 | 1.331 | 120.3 | 118.8 | 47.7 | 55.0 |
| TS5 | 1.399 | 1.423 | 2.067 | 2.044 | 1.291 | 1.221 | 3.229 | 3.150 | 1.391 | 1.412 | 122.9 | 121.5 | 47.6 | 48.0 |
| TS6 | 2.800 | 2.800 | 2.035 | 2.013 | 1.280 | 1.228 | 2.454 | 2.668 | 1.251 | 1.289 | 123.6 | 121.3 | 43.6 | 47.2 |
| TS7 | 2.476 | 2.379 | 2.040 | 2.027 | 1.005 | 1.081 | 2.728 | 2.540 | 1.254 | 1.296 | 125.1 | 123.9 | 34.4 | 39.0 |
| TS8 | 1.546 | 1.626 | 0.996 | 1.036 | 1.304 | 1.247 | 2.908 | 2.843 | 1.333 | 1.337 | 125.7 | 123.9 | 13.3 | 5.9 |
| TS9 | 2.800 | 2.800 | 0.961 | 0.987 | 1.239 | 1.169 | 1.846 | 1.950 | 1.282 | 1.286 | 122.1 | 121.9 | 58.0 | 57.3 |
| TS10 | 2.656 | 2.653 | 0.960 | 0.991 | 0.992 | 1.141 | 3.403 | 2.068 | 1.237 | 1.276 | 125.2 | 123.4 | 45.9 | 51.6 |
| TS11 | 1.457 | 1.518 | 1.300 | 1.362 | 1.006 | 1.070 | 3.063 | 2.897 | 1.364 | 1.363 | 124.0 | 123.1 | 52.5 | 20.4 |
| PC1 | 1.397 | 1.424 | 2.067 | 2.044 | 1.833 | 1.802 | 3.206 | 3.260 | 1.394 | 1.413 | 122.2 | 121.0 | 51.5 | 49.4 |
| PC2 | 1.395 | 1.428 | 2.067 | 2.046 | 1.007 | 1.076 | 3.270 | 3.253 | 1.392 | 1.410 | 122.3 | 122.5 | 53.0 | 43.1 |

[a] Lengths in Å, angles in degrees.

on the reaction pathway, calculation of the potential energy surfaces (PES's) describing these mechanisms should provide valuable insights into the kinetic importance of a particular pathway and on the structure and energy of stationary points, intermediates and transition states observed on these PES's.

The HF/6-31G* calculated potential energy surfaces of the catalytic reactions are represented in the form of two-dimensional reaction-coordinate contour diagrams in Figures 1a−3a. The distances along the *x*-axis determine the formation and scission of a new glycosidic linkage C1−Oa, and correspond to the nucleophilic attack of the acceptor Oa on the anomeric C1, while the distances along the *y*-axis characterize the proton transfer processes. Different reaction pathways can generally be identified on these potential energy surfaces. The reaction pathways parallel to the vertical and horizontal axes describe particular steps in a stepwise mechanism, while the reaction pathways following the diagonal across the PES represent a concerted mechanism. The profile of the PES's depends on the relative acidity of the different molecules involved. The calculated two-dimensional PES's only represent a section of the potential energy hypersurface describing the entire complex reaction. Nevertheless, several conclusions can be formulated from the calculated two-dimensional PES's displayed in Figures 1a−3a and they will be discussed below. Optimized structures of the different stationary points found along the reaction pathways, determined at the DFT/B3LYP/6-31G* level, are given in Figures 1b−3b. Analysis of the stationary point structures along the reaction pathways calculated at the DFT/B3LYP/6-31G* level revealed features qualitatively similar to those found on the PES's calculated at the HF/6-31G* level. As previously observed,[23−25] the inclusion of electron correlation results in slightly different magnitudes for the bond lengths, and increase of the basis set decreases the relative energy of the minima. For that reason, the discussion will essentially be based on the structures calculated at the DFT/B3LYP/6-31G* level and their energy estimated with DFT/B3LYP/6-31++G**// DFT/B3LYP/6-31G*. More detailed structural information on each stationary point is given for reference in Table 1 while their relative energies ( *E*), determined at various levels, are listed in Table 2. Figure 4 summarizes the various possible

**Table 2.** Comparison of the ab Initio Relative Energies (kcal/mol) Calculated by Various Methods for the Points Observed on PES's Described on Figures 1−3

| geometry energy | HF/6-31G* 6-31G* | HF/6-31G* 6-31++G** | B3LYP/6-31G* 6-31G* | B3LYP/6-31G* 6-31++G** |
|---|---|---|---|---|
| R | 0.00[a] | 0.00[b] | 0.00[c] | 0.00[d] |
| INT1 | 11.52 | 11.95 | 10.28 | 7.60 |
| INT2 | 35.33 | 34.18 | 32.82 | 30.41 |
| INT3 | 60.10 | 56.65 | 57.86 | 54.36 |
| INT4 | 33.76 | 30.63 | 29.25 | 25.77 |
| INT5 | 30.33 | 28.29 | 19.90 | 17.34 |
| TS1 | 19.32 | 18.38 | 16.06 | 13.35 |
| TS2 | 26.20 | 27.11 | 14.39 | 14.65 |
| TS3 | 50.45 | 50.53 | 37.63 | 36.34 |
| TS4 | 39.17 | 38.55 | 34.94 | 32.20 |
| TS5 | 6.30 | 8.22 | 1.32 | -0.22 |
| TS6 | 81.78 | 77.92 | 66.96 | 59.80 |
| TS7 | 70.84 | 64.89 | 62.43 | 54.68 |
| TS8 | 43.18 | 42.98 | 28.12 | 26.36 |
| TS9 | 54.59 | 53.27 | 39.88 | 37.55 |
| TS10 | 47.68 | 41.42 | 40.96 | 37.64 |
| TS11 | 53.50 | 52.28 | 37.92 | 34.35 |
| PC1 | -26.22 | -23.87 | -21.98 | -22.37 |
| PC2 | -1.35 | -1.81 | 0.98 | -2.99 |

[a] $E = -2225097.02$ kcal/mol. [b] $E = -2225204.97$ kcal/mol. [c] $E = -2235455.59$ kcal/mol. [d] $E = -2235585.22$ kcal/mol.

reaction pathways found for the transfer of GlcNAc as catalyzed by *N*-acetylglucosaminyltransferases.

**3.1. Potential energy surfaces. 3.1.a. PES as a Function of the $r_{Ha-OB}$ and $r_{C1-Oa}$ Distances.** The first reaction mechanism studied (Scheme 3A) characterizes the nucleophilic attack of the methanol (sugar-acceptor model) on the anomeric carbon C1 of UDP-GlcNAc (sugar-donor) either followed or preceded by the proton transfer from the methanol (acceptor) to the catalytic base (B). As mentioned earlier, only one carboxylate (catalytic base noted B in Scheme 2) is involved in this reaction mechanism, but for the sake of energy comparison with other reaction pathways, the second catalytic acid (noted A in Scheme 2) was kept in the model in a constrained position. The PES corresponding to such a reaction mechanism, calculated at the HF/6-31G* level, is given in Figure 1a. Distances plotted along both axes of the contour map describe, horizontally, the nucleophilic attack of the methanol oxygen Oa on the anomeric

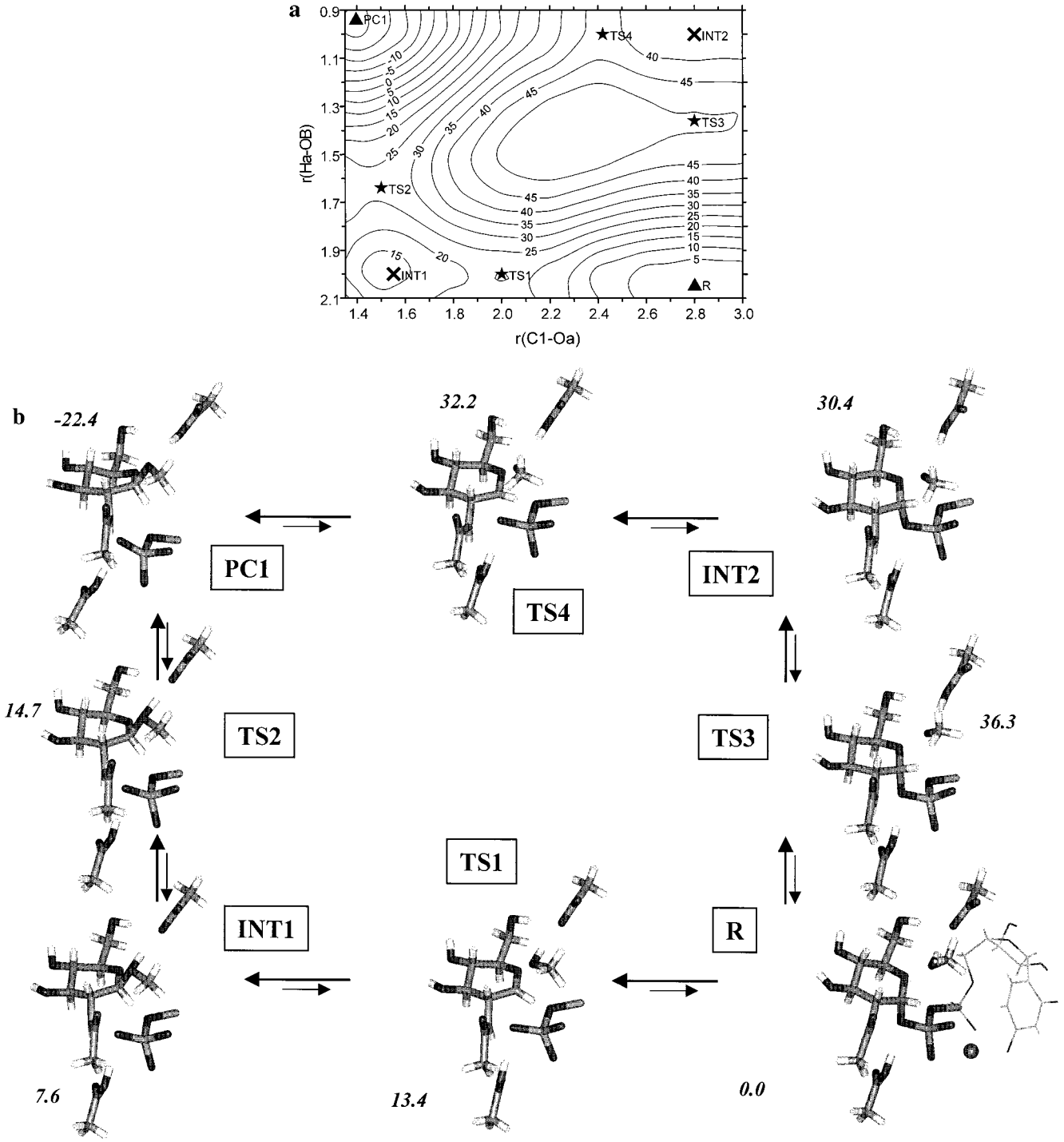



**Figure 1.** (a) Potential energy surface calculated at the HF/6-31G* level and corresponding to the mechanism involving only a catalytic base to assist the nucleophilic attack followed by proton transfer to the base (Scheme 3A). (b) Geometrical representation of the different stationary points calculated at the DFT/B3LYP/6-31G* level. Numbers in italics represent relative energies (in kcal/mol) at the DFT/B3LYP/6-31++G**//DFT/B3LYP/6-31G* level. R, TS, INT, and PC represent the reactants, transition states, intermediates, and products, respectively.

carbon C1 of GlcNAc and, vertically, the proton (Ha) transfer from the hydroxyl group of the methanol to the catalytic base (B). This PES shows two intermediates (INT1 and INT2) and four energy barriers (TS1−TS4) located in valleys along the borders of the contour map. An energy maximum, with no saddle point, is observed in the central region of the map. Thus, a concerted mechanism appears to be impossible in this model, and the reaction must proceed through a stepwise mechanism from the reactants (R) to the product complex (PC1) with the proton transfer and the nucleophilic attack as two distinct steps.The stepwise mechanism observed on the contour map of Figure 1a, offers two distinct pathways leading to the same product complex (PC1) but differing in the sequence of the individual steps. In the first pathway (R → TS1 → INT1 → TS2 → PC1), the enzymatic reaction starts with the nucleophilic attack (along the horizontal axis) of the methanol oxygen Oa on the anomeric carbon C1 of UDP-GlcNAc, followed by proton (Ha) transfer (along the vertical axis) from the methanol to the catalytic base (B). In the second pathway (R → TS3 → INT2 → TS4 → PC1), the order of the steps is reversed with the proton transfer occurring before the nucleophilic attack. Comparison of the energy barriers required to proceed along these two

pathways (Figure 4) reveals that the process starting with the nucleophilic attack is less energy demanding. From Figure 1a and Table 2, both intermediates (INT1 and INT2) appear also to have higher relative energies compared to the reactants or the products. As a consequence, the transition states are located in asymmetric positions next to the intermediates.

The energetics of the proton transfer from methanol to the catalytic base depends on the stage of the nucleophilic attack and reflects the different acidity of both participants. As the nucleophilic attack proceeds, the acidity of the attacking methanol changes from about 15, for methanol, to a value of approximately −5 for the protonated glycosidic oxygen in INT1.[29] The changes in the calculated proton-transfer energy are consistent with this variation in the p$K_a$. When methanol is in the starting position ($r_{C1-Oa}$ = 2.8 Å), not attacking the anomeric carbon of UDP-GlcNAc, the proton-transfer energy is 36.3 kcal/mol, which is in reasonable agreement with the experimental estimate of $\Delta H$ = 44 kcal/mol for the $HCOO^- + C_2H_5OH \rightarrow HCOOH + C_2H_5O^-$ process in the gas phase.[29] The proton transfer energy then gradually decreases as the methanol oxygen Oa attacks the anomeric carbon C1, to finally end up with an energy around −22.4 kcal/mol in the final stage of the nucleophilic attack ($r_{C1-Oa}$ = 1.5 Å). Examination of the PES indicates that along the different pathways, the points located at $r_{C1-Oa}$ = 2.0 Å, but with distinct positions of the proton Ha ($r_{Ha-OB}$ = 1.0 and 2.0 Å respectively), appear to have roughly the same energy. This suggests that at an $r_{C1-Oa}$ distance of 2.0 Å, the acidity of the methanol and of the catalytic base are very similar. The hydroxyl group of the methanol used in our model probably has higher acidity compared to the hydroxyl group that would be present in the real oligosaccharide substrate. It can, therefore, be expected that a weaker acid ROH of this type would increase the transition barrier and move TS1 and TS3 toward the products. Similarly, an increase in the strength of the catalytic base would move the TS's closer to INT2 or INT1.

Since the general reaction can be considered as a simple nucleophilic displacement at the anomeric carbon with inversion of configuration, the reaction path must involve the deprotonation of the acceptor oxygen and a change of the absolute configuration at the anomeric carbon. These changes are clearly seen in the structure of the discrete points along both reaction pathways (Figure 1b). Analysis of these geometrical changes revealed that, while the proton transfer only marginally influences the structure of the reactants, the nucleophilic attack results in a significant alteration of the UDP-GlcNAc structure. Along the reaction path: R → TS1 (13.4 kcal/mol) → INT1 (7.6) → TS2 (14.7) → PC1 (−22.4), the C1−O1 bond length between the anomeric carbon C1 and the leaving group, UDP, gradually elongates from 1.519 to 3.260 Å as the distance between the anomeric carbon and the attacking oxygen $r_{C1-Oa}$ decreases. The transition state for the nucleophilic attack occurring as the first step in the reaction, TS1, undergoes significant geometrical changes compared to the starting structure, R. As the C1−Oa reaction coordinate gets close to 2.16 Å as in TS1, the C1−O1 scissile bond increases drastically by 1 Å going from 1.519 to 2.535 Å and the C1−O5 bond shortens from 1.371 to 1.290 Å. In INT1, the C1−Oa and C1−O1 bonds have values of 1.532 and 2.829 Å, respectively. During the second step of the reaction, the proton Ha of TS2 is positioned closer to the oxygen Oa with $r_{Ha-Oa}$ = 1.361 Å. The C1−O1 distance slightly stretches to 3.014 Å, whereas the C1−Oa bond shortens to 1.499 Å but, overall, only small changes were found between the relevant bonds of TS2 and INT1.

The geometry of the starting active site model (R) is characterized by values of 1.519 and 1.371 Å for, respectively, the C1−O1 and C1−O5 bond lengths. The pyranoid ring of the GlcNAc is initially in the $^4C_1$ chair conformation characterized by ring-puckering parameters $\phi$ = 246.2, $\phi$ = 14.1, and Q = 0.54. Along the reaction coordinate, the conformation of the pyranoid ring continuously changes from a $^4C_1$ chair through a $^4H_3$ half-chair and a $^4E$ envelope conformation where the proton H1 is in a quasi-planar position and finally back to a $^4C_1$ chair conformation. During this process, the H1 atom moves, from an equatorial position through a position in the plane defined by the C2−C1−O5 atoms, to an axial position. Interestingly, no boat conformations of the transferred sugar, as previously described for $\beta$-1,4-xylanases[30] or suggested for chitinases,[31] were found on the PES. These modifications in the six-membered ring conformation of GlcNAc are accompanied by changes in the orientation of the leaving and attacking groups with respect to the six-membered ring. As the ring shape shifts to the envelope conformation, the atoms attached to the anomeric carbon become coplanar with an $sp^2$ character at the reaction center, C1. The delocalization of the ring oxygen lone-pair electrons into the empty p orbital at the C1 atom stabilizes the oxocarbenium ion-like character of GlcNAc. The formation of such an oxocarbenium ion requires alterations of the GlcNAc ring conformation, from chair to half-chair or envelope, to accommodate the partial double-bond character. A consequence of the charge delocalization is the shortening of the C1−O5 bond length from its equilibrium value of 1.371 Å observed in R. This change is more pronounced in TS1, with $r_{C1-Oa}$ = 2.158 Å, where the C1−O5 bond developed a partial double bond character ($r_{C1-O5}$ = 1.290 Å). The orientation of both the leaving and the attacking groups, is also influenced by the tendency to optimize interactions between the C1 carbon p orbital and the lone pairs of the connecting oxygen atoms of these groups. The most efficient interactions clearly occur when these lone pairs are located in the direction of this p orbital, that is, oriented perpendicularly to the O5−C1−C2 plane. A stronger nucleophilic character of the methanolate should result in a larger stabilization of such an oxocarbenium species. Indeed, it appears that in the case of methanol, the O1 and Oa atoms adopt a quasi-orthogonal orientation regarding the O5−C1−C2 plane with the O5−C1−O1/Oa bond angles close to 90° at $r_{C1-Oa}$ = 2.158 Å, whereas in the case of methanolate (the alternative pathway described on the map), this situation occurs earlier at a larger C1−Oa distance ($r_{C1-Oa}$ = 2.426 Å).

For the alternative reaction pathway described on the map, R → TS3 (36.3) → INT2 (30.4) → TS4 (32.2) → PC1 (−22.4), the conversion of the reactants into the intermediate INT2 during the proton transfer occurs through the transition state TS3 at $r_{Ha-Oa}$ = 1.638 Å. The geometry of GlcNAc does not exhibit any important change during this reaction step. The C1−O1 bond length is a good illustration of this behavior since this bond remains almost unchanged with lengths of 1.519, 1.524, and 1.529 Å in, respectively, the R, TS3, and INT2 stationary points.

As in the first pathway, the main geometrical changes are connected with the nucleophilic attack occurring along the horizontal axis of the contour map. Several interesting differences are seen between the TS1 and TS4 structures. In TS4,

(29) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.

(30) Sidhu, G.; Withers, S. G.; Nguyen, N. T.; McIntosh, L. P.; Ziser, L.; Brayer, G. D. *Biochemistry* **1999**, *38*, 5346−5354.

(31) Brameld, K. A.; Goddard, W. A., III. *J. Am. Chem. Soc.* **1998**, *120*, 3571−3580.

the distance C1−Oa of 2.426 Å is longer compared to the values of 2.158 Å found for TS1. By contrast, the length of the C1−O1 scissile bond is considerably shorter in TS4, 1.863 Å, versus 2.535 Å in TS1. Using these distances as a criterion to describe the extent of the transfer reaction, one can assume TS4 is an earlier transition state because its geometry is closer to the reactants, in contrast to TS1 that might be characterized as a late transition state since the structure is nearer to that of the products. Both the TS1 and the TS4 structures have significant $sp^2$ character at the C1 atom.

A comparison of the orientations of the *N*-acetyl group located at C2 shows that for all points on the PES, the acetamido group remains in the most stable conformation, called *Z*-trans.[32] This indicates that the *N*-acetyl group does not participate in the catalytic mechanism through so-called substrate assisted catalysis by stabilizing the developing oxocarbenium character on C1 as proposed earlier for some retaining hydrolases.[33] This is not surprising given the difference in stereochemical outcome. We found that for some early points along the R → TS1 → INT1 → TS2 → PC1 reaction pathway, the *N*-acetyl group is brought closer to the leaving UDP group. However, the weak (N)−H···O1 hydrogen bond formed disappears as the C1−O1 distance increases.

**3.1.b. PES's as a Function of the $r_{HA-O1}$ and $r_{C1-Oa}$ Distances.** The mechanism considered here describes the proton transfer from the catalytic acid (A) to the glycosidic oxygen O1 and the nucleophilic attack of the acceptor oxygen Oa on the anomeric carbon C1 (Scheme 3B). Results from the mechanism described earlier showed that the nucleophilic attack and the proton transfer from the acceptor to the catalytic base (B) proceed in distinct steps. The question then was whether this behavior also remained for pathways where the HA proton of a second catalytic acid (A) is attacking the glycosidic oxygen. Preliminary calculations of the PES as a function of the $r_{HA-O1}$ and $r_{C1-Oa}$ distances indicated a conservation of these features. However, the optimization led to structures with the Ha proton located either at the acceptor oxygen atom (Oa) or at the catalytic base oxygen (OB) depending on the starting position. For that reason, we have calculated two different PES's in order to describe the present mechanism. In the PES in Figure 2a, Ha is initially located at the catalytic base oxygen, OB, while in the PES in Figure 3a, Ha is positioned at the acceptor oxygen, Oa. Analyses of the results support our previous findings and they reveal that on both potential energy surfaces, the Ha proton always remains in its same starting location. Some assumptions are implicitly included in these two models of the reaction mechanisms. For the PES shown in Figure 3a (Ha located at the acceptor), the proton transfer from the acceptor to the catalytic base is the final step completing the reaction. However, in the PES of Figure 2a (Ha located at the catalytic base), the proton transfer from the acceptor to the base (B) precedes the nucleophilic attack and the proton transfer from catalytic acid (A) to glycosidic oxygen O1.

Both PES's corresponding to the type of mechanism described in Scheme 3B and calculated at the HF/6-31G* level, are given in Figures 2a and 3a. The distances $r_{C1-Oa}$ and $r_{Ha-O1}$ represented along both axes of the contour map characterize horizontally the nucleophilic attack of the methanol oxygen Oa on the anomeric carbon C1 of GlcNAc and, vertically, the proton (HA) transfer from the catalytic acid (A) to the glycosidic oxygen O1. Although the proton-transfer processes represented on the vertical axes of Figure 1 and Figures 2−3 refer to a transfer between different molecules, all maps exhibit similar features. They all indicate that a concerted mechanism is impossible in this model. As a consequence, the reaction must proceed by a stepwise mechanism through different reaction channels with the proton transfer and the nucleophilic attack occurring as two distinct steps. In all maps, the transition states are similarly located in asymmetric positions near the intermediates.

Two different pathways are observed on each of the PES's, namely INT2 → TS4 → PC1 → TS5 → PC2 and INT2 → TS6 → INT3 → TS7 → PC2 in Figure 2 and R → TS1 → INT1 → TS8 → INT5 and R → TS9 → INT4 → TS10 → INT5 in Figure 3. Three intermediates and four transition states are encountered in valleys along the borders of each contour map. It should be pointed out that some points of these PES's, corresponding to nucleophilic attack with the HA proton located at the catalytic acid, coincide with points shown in Figure 1. This was possible because in the mechanism illustrated in Figure 1, the second catalytic acid (A) was included in the model in a constrained position even though it is not involved. In this way, a consistent comparison between the energies required by the diverse reaction pathways was achieved.

The calculated PES's show that the HA proton must pass through a relatively large energetic barrier during its transfer from the catalytic acid (A) to the glycosidic oxygen O1. In general, this step is the most energy demanding among all the steps occurring along a particular reaction pathway. When this process is the starting step of a reaction mechanism, the energy barrier calculated at the 6-31G* level is particularly high. For example, in the pathway R → TS9 (37.6 kcal/mol) → INT4 (25.8) → TS10 (37.7) → INT5 (17.3) (Figure 3a), the barrier for the proton transfer from the catalytic acid to the glycosidic oxygen (R → INT4) approaches 37.6 kcal/mol. A similar magnitude for the energy barrier between R→ INT2 (30.4 kcal/mol) was observed in the mechanism described earlier, corresponding to the Ha proton transfer (Figure 1). This suggests that in the present mechanism, the probability of protonation of the glycosidic oxygen might be too low to be kinetically productive. As in Figure 1, the more favorable pathways found in Figures 2 and 3 are those starting with the nucleophilic attack.

The difference between the nucleophilic character of methanolate and methanol oxygen atoms is clearly reflected in the location of the transition barriers on PES's given in Figures 2a and 3a. When the nucleophile is methanolate (Figure 2a), the reaction barriers for the nucleophilic attack are closer to the starting reactants (INT2), whereas, in the case of nucleophilic attack by methanol (Figure 3a), the TS's are closer to the products. Barriers for the proton (HA) transfer from the catalytic acid (A) to the glycosidic oxygen O1 are located asymmetrically on the bottom part of the maps, closer to the final intermediates, and are therefore exhibiting the character of late transition states. The location of these energy barriers results from the fact that the glycosidic oxygen O1 in UDP-GlcNAc might have a p*K*a value[34] of approximately −10 and that the pyrophosphate group is a very strong acid. The activation energy of the reverse reaction in solution (the proton transfer from O1 to the catalytic acid) is an exothermic process assumed to be a diffused-controlled reaction with an activation barrier of about 5 kcal/mol,[35] which is in reasonable agreement with the energy barriers of about 6 kcal/mol calculated on these maps, with however the exception of the proton transfer between INT4 → TS9.

(32) Fowler, P.; Bernet, B.; Vasella, A. *Helv. Chim. Acta* **1996**, *79*, 269−287.

(33) Tews, I.; Perrakis, A.; Oppenheim, A.; Dauter, Z.; Wilson, K. S.; Vorgias, C. E. *Nat. Struct. Biol.* **1996**, *3*, 638−648.

(34) Guthrie, J. P. *J. Am. Chem. Soc.* **1977**, *99*, 3391−4000.

(35) Warshel, A. *Computer modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1997.

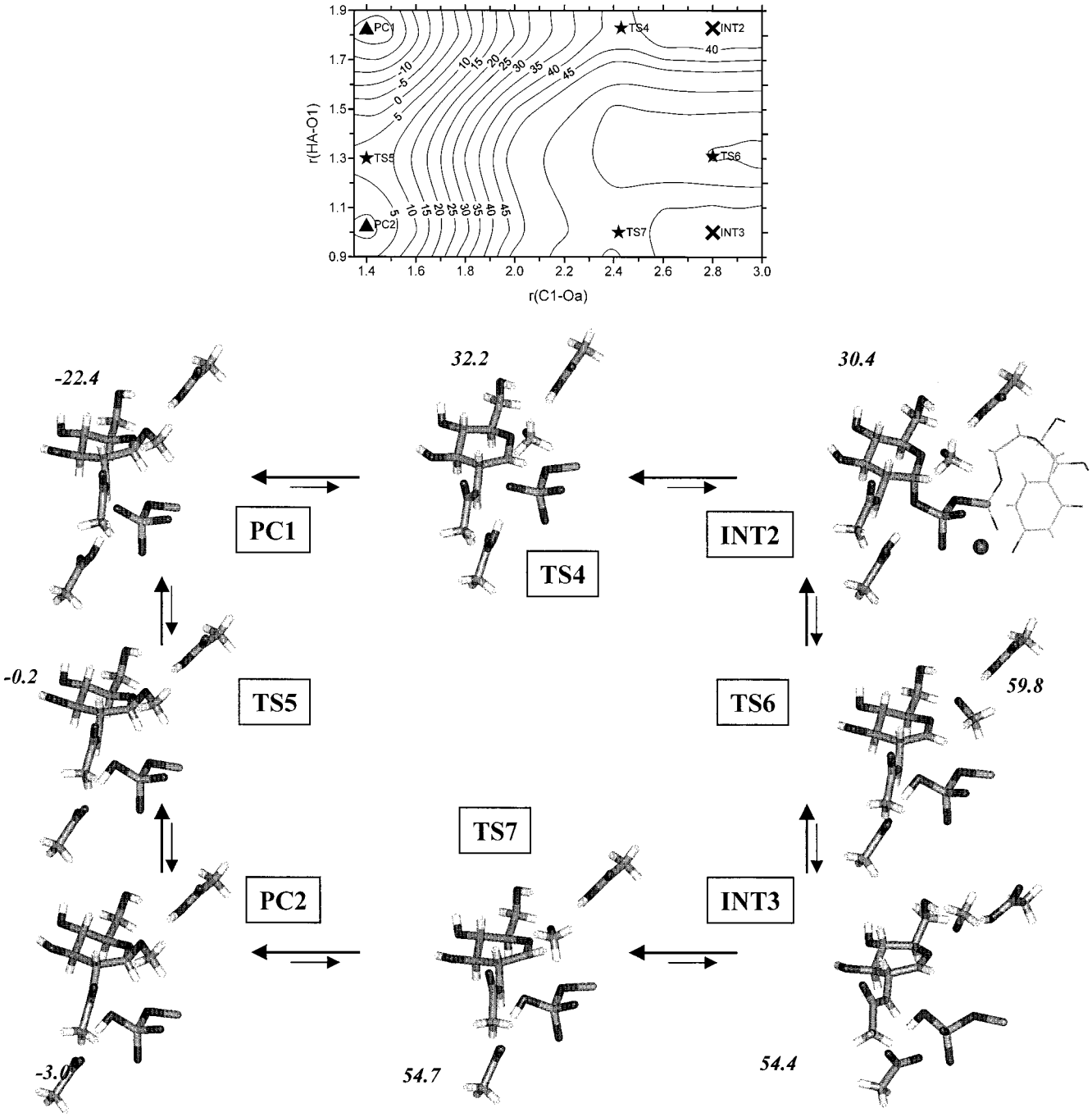



**Figure 2.** (a) Potential energy surface calculated at the HF/6-31G* level and corresponding to the mechanism involving a pair of catalytic amino acids to assist the proton transfer to O1 and the nucleophilic attack. In this mechanism, the proton Ha is positioned at the base (Scheme 3B). (b) Geometrical representation of the different stationary points calculated at the DFT/B3LYP/6-31G* level. Numbers in italics represent relative energies (in kcal/mol) at the DFT/B3LYP/6-31++G**//DFT/B3LYP/6-31G* level. R, TS, INT, and PC represent the reactants, transition states, intermediates, and products, respectively.

As expected, the analysis of the geometrical changes observed along the reaction pathways, shown in Figures 2b and 3b, reveals that the nucleophilic attack alters the structure in a fashion similar to the mechanism described in Figure 1. However, the proton transfer from the catalytic acid to O1 has a more significant influence on the structure of UDP-GlcNAc than was seen in Figure 1, where the proton transfer is from the acceptor to the catalytic base. The modifications in the structure of UDP-GlcNAc caused by the proton transfer can be illustrated by the bond length $r_{C1-O1}$, between the anomeric carbon C1 and the oxygen O1, that gradually elongates as the HA proton approaches the O1 atom. Such a process is observed in Figure 3a during the R → INT4 step, where the value of the C1−O1 bond length changes from an equilibrium position of 1.519 Å through 1.950 Å in TS9 to 3.733 Å in INT4. In the INT2 → INT3 step described in Figure 2a, the elongation of the glycosidic bond gets more pronounced, with the C1−O1 distance increasing from 1.529 to 2.663 Å. The reverse trend is however observed for the C1−O5 bond that shortens from 1.371 to 1.259 Å along R → INT4 and from 1.381 to 1.253 Å along INT2 → INT3. The changes in the bond lengths observed during this proton transfer process are accompanied by an alteration of the six-membered ring conformation. The conformation of the pyranoid ring continuously changes from a $^4C_1$ chair to a $^4H_3$ half-chair and

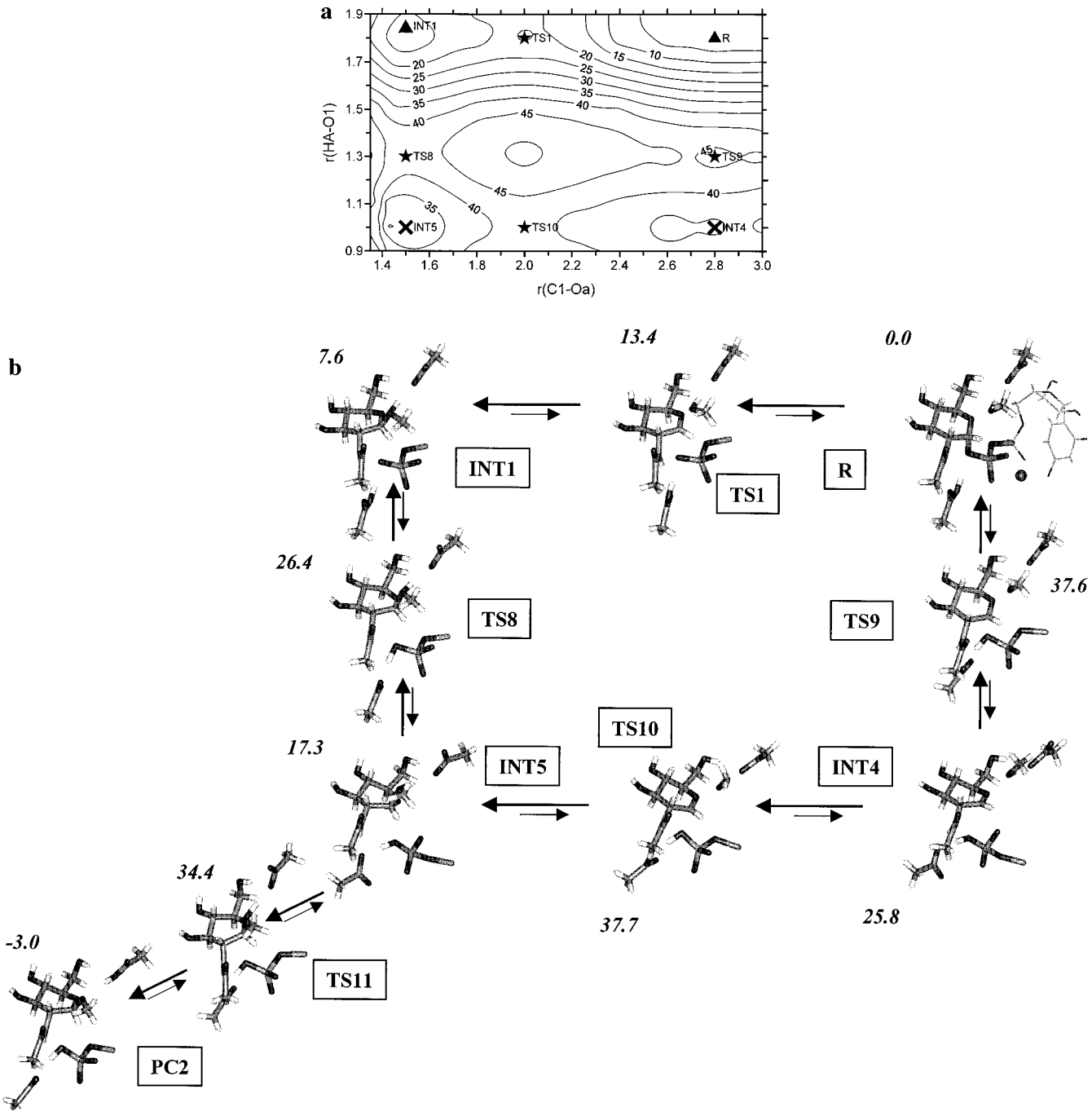



**Figure 3.** (a) Potential energy surface calculated at the HF/6-31G* level and corresponding to the mechanism involving a pair of catalytic amino acids to assist the proton transfer to O1 and the nucleophilic attack. In this mechanism, the proton Ha is located at the acceptor (Scheme 3B). (b) Geometrical representation of the different stationary points calculated at the DFT/B3LYP/6-31G* level. Numbers in italics represent relative energies (in kcal/mol) at the DFT/B3LYP/6-31++G**//DFT/B3LYP/6-31G* level. R, TS, INT, and PC represent the reactants, transition states, intermediates, and products, respectively.

back to the $^4C_1$ conformation. During this process, the H1 atom moves from the equatorial position through the position in the plane defined by the C2−C1−O5 atoms to reach, ultimately, the axial position. These results show that the structure around the reaction center is very sensitive to the nucleophilic attack and the proton transfer to the glycosidic oxygen. Moreover, they suggest that geometrical changes caused by these two processes are comparable, both leading to the cleavage of the C1−O1 bond and the alteration of the GlcNAc ring conformation leading to inversion of the anomeric configuration of the transferred sugar. The energetic requirements for the two processes are very different and they indicate that the nucleophilic attack is the less-demanding operation.

Using information coming from the three calculated PES's, various possible reaction pathways for the transfer of GlcNAc catalyzed by *N*-acetylglucosaminyltransferases could be established. They are schematized in Figure 4. Among them, four distinct stepwise reaction pathways exist to describe the transfer of GlcNAc using mechanisms involving jointly a catalytic acid and a catalytic base. Two pathways, **R → TS3 (36.3) → INT2 (30.4)** → TS4 (32.2) → PC1 (−22.4) → TS5 (−0.2) → PC2 (−3) and **R → TS3 (36.3) → INT2 (30.4)** → TS6 (59.8) → INT3 (54.4) → TS7 (54.7) → PC2 (−3), start with the proton transfer from the acceptor to the catalytic base, **R → TS3 → INT2**, and as such, have their first step (noted in bold) identical to the mechanism previously discussed, involving only the

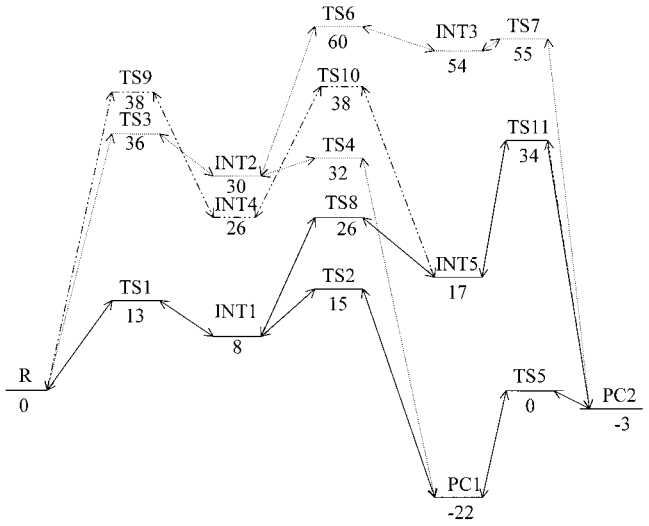



**Figure 4.** Schematic energetic representation (in kcal/mol) of the possible reaction pathways observed in the different PES's for the transfer of GlcNAc catalyzed by inverting *N*-acetylglucosaminyltransferases. Relative energies are calculated at the DFT/B3LYP/6-31++G**// DFT/B3LYP/6-31G* level.

catalytic base. In two other pathways, R → TS1 (13.4) → INT1 (7.6) → TS8 (26.4) → **INT5 (17.3) → TS11 (34.4) → PC2 (−3)** and R → TS9 (37.6) → INT4 (25.8) → TS10 (37.7) → **INT5 (17.3) → TS11 (34.4) → PC2 (−3)**, the final step of the reaction, **INT5 → TS11 → PC2**, is the proton transfer from the acceptor to the catalytic base. In the first of these two pathways, the proton transfer to O1 occurs as the second step after the nucleophilic attack and it proceeds from INT1 to the intermediate INT5 via TS8. Both TS8 and INT5 structures have the GlcNAc ring in a $^4H_3$ conformation and the C1−O5 bond length at around 1.33 Å. The C1−O1 bond length differs though for these structures: 2.843 Å versus 3.467 Å for TS8 and INT5, respectively. In the second pathway, the proton transfer happens as the first step, from R to INT4 via TS9, and it is energetically less favorable. In this case, the C1−O1 and C1−O5 bonds of TS9 are 1.950 and 1.286 Å, respectively. For all points on the PES's, the acetamido group remains in the most stable conformation, called Z-trans.[32]

Several changes described above resemble those assumed in reactions catalyzed by glycosyl hydrolases. The calculated potential energy surface for a general acid-catalyzed reaction of lysozyme[35] has suggested for the transition state [RC1−O1-(H+)-R], C1−O1 bond lengths in the range of 2.5−2.6 Å. We assume that the larger values calculated for the C1−O1 bond lengths of UDP-GlcNAc reflect the better leaving character of the UDP group compared to sugar aglycons. The comparison of the six-membered ring conformational rearrangements calculated in the present study with those observed in reactions catalyzed by glycosidases is of particular interest. Enforced by either the protonation of the glycosidic oxygen or the nucleophilic attack on the anomeric carbon, the changes observed in the GlcNAc ring conformation of UDP-GlcNAc resemble to some extent those calculated for 2-oxanol.[36] This molecule was used as a hexopyranose model in the investigation of transition state structures during the glycoside hydrolysis mechanism. Some major features, however, distinguish glycosidases from glycosyltransferases. For example, it is clear from the analysis of all the points located on the PES's that the GlcNAc ring does not adopt any of the boat conformations sometimes described for glycosidase mechanisms.[30,31] Most likely, the restraints associated with the nucleophilic attack of the acceptor and with the inversion of configuration at the C1 atom prevent any large ring conformational changes along the reaction pathways, which is in agreement with the assumption of the least motion effect.[37] Complexes of glycosidases with a substrate or a product, in which a sugar ring is substantially deformed, have been experimentally observed.[30,38−40] Ring distortion induced by these enzymes in ground states has been assumed to be crucial for their reaction mechanism. On the other hand, circular dichroism studies in solution of *N*-acetylglucosaminyltransferase V and its complex with UDP-GlcNAc [41] suggested that the UDP part alone of UDP-GlcNAc is tightly bound to the enzyme while the GlcNAc residue is simply weakly interacting with the enzyme. These findings are supported by X-ray structural data available on complexes of glycosyltransferases with UDP-sugars[8,11,12] showing only the location of the UDP part in the binding pocket. This suggests that in the ground state, the GlcNAc residue, initially observed in the $^4C_1$ conformation, is only loosely bound to the enzyme, and therefore, unlikely to have a ring distorded.

The inclusion of electron correlation by means of the DFT/B3LYP method at the 6-31G* level usually reduces the relative energy of the stationary points determined on PES's compared to HF calculations (Table 2). The largest shifts are usually found for the structures along the proton transfer process, what indicates the importance of the use of electron correlation to describe systems with hydrogen bonds. The relative energy of the stationary points at our best theory DFT/B3LYP/6-31++G**// DFT/B3LYP/6-31G* is further decreased usually by about 3 kcal/mol and in a few cases by as much as 8 kcal/mol compared to DFT/B3LYP/6-31G*//DFT/B3LYP/6-31G*. The inclusion of electron correlation also affects the geometry of the molecules by increasing the bond lengths by approximately 0.3 Å (Table 1) as has earlier been observed for similar compounds.[23−25]

**3.2. Reaction Pathways.** The calculated PES's show altogether the presence of five intermediates (INT1−INT5) and 11 energy barriers indicating thus the existence of 11 transition states (TS1−TS11). A number of results on the mechanism of inverting *N*-acetylglucosaminyltransferases become immediately apparent on consideration of the structures of the stationary points described in Figures 1b−3b and Table 1 as well as the relative energetic data listed in Table 2 and schematically shown in Figure 4:

A maximum in energy is observed in the central region of all the calculated PES's indicating that a concerted mechanism is impossible in these models of GlcNAc transfer reactions. Therefore, to move from reactants (R) to the product complex, the reaction has to proceed through a stepwise mechanism. To avoid confusion, we would like to emphasize that this concerns only the $r_{C1-Oa}$, $r_{Ha-OB}$, and $r_{HA-O1}$ reaction coordinates used to define the reaction mechanism. The C1−O1 bond has not been considered as a reaction coordinate since we assume here that changes in this bond length are a consequence of the nucleophilic attack or the proton transfer. This is supported by the calculation of a one-dimensional reaction profile using the C1−O1 distance as the reaction coordinate at the DFT/B3LYP/6-31G* level that showed a barrier of ∼27 kcal/mol (data not

(36) Smith, B. J. *J. Am. Chem. Soc.* **1997**, *119*, 2699−2706.
(37) Sinnott, M. L. *Adv. Phys. Org. Chem.* **1988**, *24*, 113−224.
(38) Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schulein, M.; Withers, S. G. *Biochemistry* **1998**, *37*, 11707−11713.
(39) Strynadka, N. C. J.; James, M. N. G. *J. Mol. Biol.* **1991**, *220*, 401−424.
(40) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schulein, M.; Davies, G. J. *Biochemistry* **1996**, *35*, 15280−15287.
(41) Zhang, N.; Peng, K. C.; Chen, L.; Puett, D.; Pierce, M. *J. Biol. Chem.* **1997**, *272*, 4225−4229.

shown). This barrier is considerably higher than the corresponding barrier for the nucleophilic attack and suggests that this process is unlikely to be the first step of the reaction. Our results clearly show that the C1−O1 distance varies in a continuous manner with the $r_{C1-Oa}$ and $r_{HA-O1}$ distances as a result of the nucleophilic attack at C1 or the proton transfer to O1. Whether the prolongation of the C1−O1 bond and the nucleophilic attack at C1 and proton transfer to O1 proceed in a concerted manner remains to be explored. Calculations aimed at addressing this question are ongoing in our laboratory.

There exist a number of transition states and intermediates connected by several pathways associated with proton transfer between the enzyme and the substrates and the nucleophilic attack of the acceptor. From the six possible pathways described here, only one [R → TS3 (36.3) → INT2 (30.4) → TS6 (59.8) → INT3 (54.4) → TS7 (54.7) → PC2 (−3)] appears very unlikely. In this pathway, the catalytic reaction begins with two consecutive energetically unfavorable proton transfers.

The overall activation energy of 38 kcal/mol, calculated for the preferred pathway of the mechanism requiring the participation of a pair of carboxylic acids [R → TS9 (37.6) → INT4 (25.8) → TS10 (37.7) → INT5 (17.3) → TS11 (34.4) → PC2 (−3)], is relatively high. This mechanism involves the proton transfer from a catalytic acid to the glycosidic oxygen. However, the energy results, shown in Figure 4, apparently suggest that for some enzymes the involvement of this catalytic acid in the reaction mechanism might not be essential for the enzymatic catalysis. These conclusions are supported by experimental findings reported for the inverting mechanism of human fucosyltransferase V, where only one carboxylate residue functions as a general base catalyst. It has been shown[42,43] that a single proton is "in flight" at the rate determining transition state. The secondary deuterium kinetic isotope effect for this reaction is consistent with a large degree of $S_N1$ character at the transition state and therefore a largely dissociative mechanism.

The GlcNAc transfer mechanism, assuming the enrolment of only a catalytic base [R → TS1 (13.4) → INT1 (7.6) → TS2 (14.7) → PC1 (−22.4)], appears to be the least energy-demanding pathway represented in Figure 4. The overall activation energy calculated at the DFT/B3LYP/6-31++G** //DFT/B3LYP/6-31G* level is approximately 15 kcal/mol. The alternative pathway with the proton transfer to the catalytic base occurring as the first step, [R → TS3 (36.3) → INT4 (25.8) → TS4 (32.2) → PC1 (−22.4)], requires a considerably higher overall activation energy (about 36 kcal/mol).

A comparison of the calculated reaction barriers with experimental data would be very instructive, but unfortunately, values of the rate constant $k_{cat}$ determined for the various GT's are not available in the literature. To our knowledge, only $k_{cat}$ values for blood group A and B glycosyltransferases[44] and for FucT V[42] have been reported. These $k_{cat}$ values are in the range between 50 and 0.1 $s^{-1}$. Using the phenomenological definition associating $k_{cat}$ with the activation free energy, $\Delta G^{act} = -RT \ln(hk_{cat}/k_BT)$, activation barriers between 15 and 19 kcal/mol have been determined. These estimates are in reasonable agreement with the overall activation energy of about 15 kcal/mol calculated for the GlcNAc transfer mechanism via the [R → TS1 (13.4) → INT1 (7.6) → TS2 (14.7) → PC1 (−22.4)] pathway. This further supports the assumption that inverting *N*-acetylglucosaminylltransferases would prefer the general base-catalyzed mechanism represented by such a pathway. However, without any additional experimental evidence and because this agreement could be only fortuitous, it would be premature to completely exclude the possibility for the reaction to proceed via another pathway. Especially, when similar activation barriers have been calculated for several reaction pathways suggesting that subtle changes in the microenvironment of the active site could change the overall reaction barrier of any of the pathways.

Transition state structures associated with the different reaction pathways exhibit significant variations in their C1−Oa and C1−O1 bond lengths (Table 1). Using these distances as criteria, the multiple transition states can be clustered into three groups based on common structural features. The differences in the C1−Oa and C1−O1 bonds of the TS's, associated with different stages of the reaction pathways, can be as large as 1.2 and 1.5 Å, respectively. Superpositions of the TS's belonging to each group are represented in Figure 5. The first group (Figure 5A) is characterized by structures, such as TS2, TS5, TS8, and TS11, having short C1−Oa bonds within the range of 1.4−1.6 Å and long C1−O1 distances, between 2.8 and 3.2 Å. The geometry of these stationary points is close to the structure of the final products, PC1 or PC2, where the C1−Oa bond created from the nucleophilic attack of the acceptor on C1, almost reached its final length of 1.42 Å as the UDP group is leaving the reaction site. These structures can therefore be described as "late transition states". The second group (Figure 5B) is represented by structures, such as TS3, TS4, TS9, and TS10 with long C1−Oa bonds located within the range of 2.4−2.7 Å and short C1−O1 distances between 1.5 and 2.1 Å. The geometries of these stationary points have not yet been altered by the reaction and thus, are very similar to the initial reactants. These structures can be termed as "early transition states". The third group (Figure 5C) corresponds to intermediate structures such as TS1, TS6, and TS7 where both C1−Oa and C1−O1 are elongated compared to their initial values, but the structures did not yet reach the final arrangement observed in the products, PC1 and PC2. Compared to the $^4C_1$ chair conformation in the reactants, the shape of the GlcNAc ring in most of the transition states changes to a half-chair. These variations in C1−Oa and C1−O1 bond lengths and in the ring shape of the GlcNAc residue, clearly demonstrate that the design of a transition state analogue inhibitor is dependent on the actual mechanism of a particular enzyme.

Although our active-site model consists of all the molecules that may directly be involved in the mechanism, it is only a model of the real active site, and as such, it has its limitations. Several factors can influence the calculated relative energies. The actual arrangement of the relevant molecules, as well as their conformation in the real active site, might differ from the model and also could vary from enzyme to enzyme. The real location of the catalytic acids in a particular enzyme may be different compared to the model. This may result, for example, in a smaller $r_{Oa-OB}$ distance between the catalytic base and the hydroxyl group of the oxygen of the acceptor. As a consequence, the reaction barrier for the proton transfer from the acceptor to the catalytic base might be lower.[45] Reaction barriers for proton transfer and nucleophilic attack processes may also be drastically influenced by the presence in the vicinity (up to 6 Å) of the reaction center of ionized amino acid residues, despite the fact

(42) Murray, B. W.; Takayama, S.; Schultz, J.; Wong, C.-H. *Biochemistry* **1996**, *35*, 11183−11195.
(43) Murray, B. W.; Wittmann, V.; Burkart, M. D.; Hung, S. C.; Wong, C. H. *Biochemistry* **1997**, *36*, 823−831.
(44) Seto, N. O.; Compston, C. A.; Evans, S. V.; Bundle, D. R.; Narang, S. A.; Palcic, M. M. *Eur. J. Biochem.* **1999**, *259*, 770−775.
(45) Lu, D.; Voth, G. A. *J. Am. Chem. Soc.* **1998**, *120*, 4006−4014.

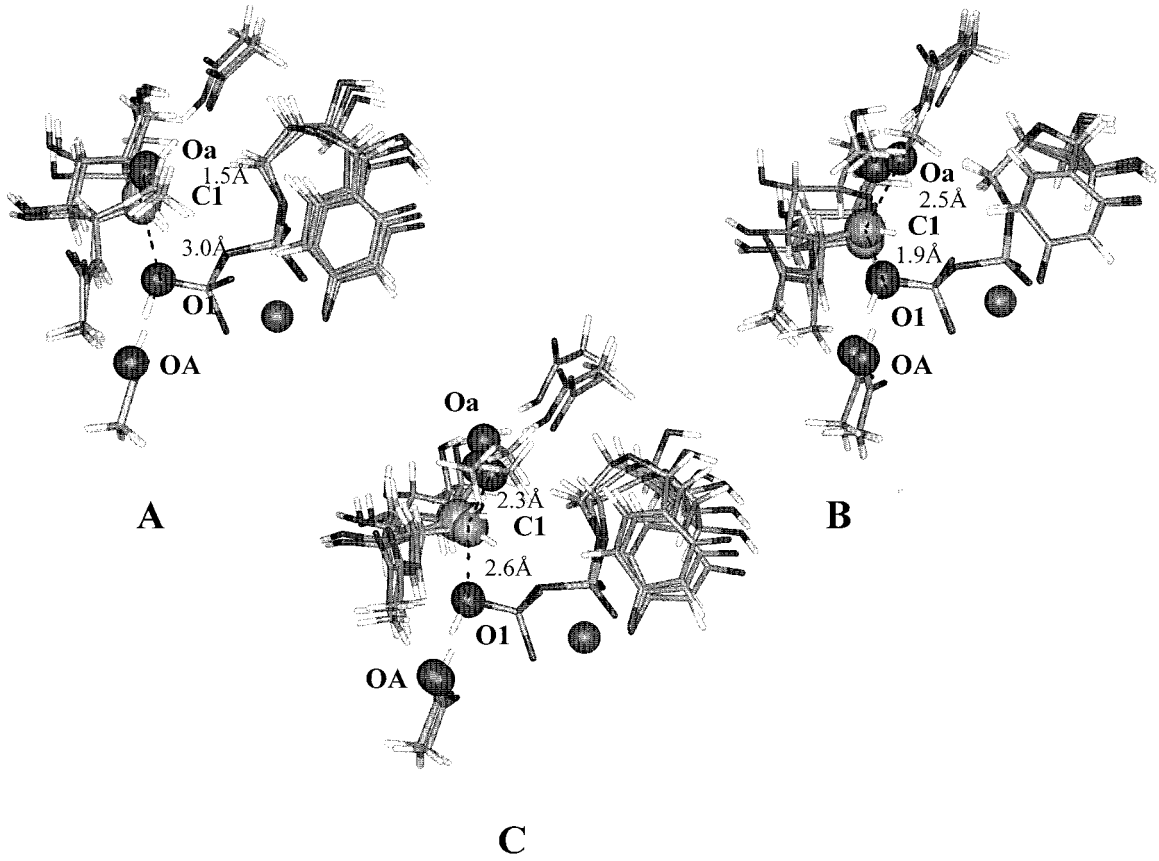



**Figure 5.** Geometrical representation of the transition states, TS1−TS11, calculated at the DFT/B3LYP/6-31G* level. Transition states are clustered by similarities in their C1−Oa and C1−O1 bond lengths. Average C1−Oa and C1−O1 distances, calculated for each group, are noted on the figure. (A) TS2, TS5, and TS8 structures exhibit short C1−Oa (1.4−1.6 Å) and long C1−O1 bond lengths (2.8−3.2 Å). TS11 has been omitted from the structure superposition for clarity purpose. (B) TS3, TS4 and TS9 structures display long C1−Oa (2.4−2.7 Å) and short C1−O1 (1.5−2.1 Å) bond lengths. TS10 has been omitted from the structure superposition for clarity. (C) TS1, TS6 and TS7 structures exhibit elongated C1−Oa (2.1−2.4 Å) and C1−O1 (2.5−2.7 Å) bond lengths.

that they might not participate in the reaction.[46] The influence of the catalytic metal on the conformation of the substrates is also a parameter to consider. An earlier study[25] showed that the relative stability of sugar-pyrophosphate conformations is sensitive to the occupancy of the metal coordination shell by interactions with surrounding elements present in the enzyme active site. This can be observed for instance, in the crystal structure of a nucleotide-complexed form of nucleotide-diphospho-sugar transferase SpsA,[11] where the coordination of the metal present in the active site, $Mn^{2+}$, involves a neighboring aspartate residue. The addition into our theoretical model of molecules coordinating the metal would help in obtaining a more accurate description of the system. This is the reason an investigation of the effects of the metal coordination on the reaction pathways is ongoing in our laboratory.

Although $\Delta G$ values would be more appropriate for description of reaction processes, our conclusions are based on the energy values. For a system as complex as our reaction model, the determination of $\Delta G$ for each individual step of the catalytic reaction would require an enormous computational resource. On the basis of our previous experience working with model compounds[23−25] that suggested that the inclusion of the ZPE and thermodynamic contributions to the calculated energies would only slightly decrease energy differences, we did not attempt such calculations. Therefore, it might be presumed that our calculated $\Delta E$ values could be slightly overestimated compared to $\Delta G$ values. Investigation of all these aspects was beyond the scope of this work. Moreover, it appears that some of the concerns raised above will only be elucidated when crystal structures of nucleotide-sugar·enzyme complexes become available.

**3.3. On the Catalytic Reaction Mechanism of Inverting *N*-Acetylglucosaminyltransferases.** On the basis of the available experimental data and the results of these calculations, we can suggest a mechanism for the inverting *N*-acetylglucosaminyltransferases. Experimental studies have revealed a sequential mechanism for GnT I and GnT II[6,22] where UDP-GlcNAc binds first either to an enzyme·$Mn^{2+}$ complex or as a $Mn^{2+}$·UDP-GlcNAc complex followed by the binding of the acceptor substrate. It has already been observed in crystals[8] and in solution,[41] that the binding of the sugar-donor appears to be followed by a conformational change of the enzyme upon binding of the sugar-acceptor. The binding of the nucleotide-sugar by the enzyme therefore triggers the conformational change that will bring the donor- and acceptor-binding sites into the proper orientation to start the enzymatic reaction. Crystal structures of glycosyltransferases[8,11,12] revealed however, that only the UDP part is observable in the nucleotide-sugar enzyme complex. A solution study on GnT V[41] indicated as well that the sugar residue of the donor appears to be loosely bound in the ground state complex. This brings us to assume that only the UDP-binding domain is present or accessible in the sugar-donor·enzyme complex. This hypothesis can be supported by inhibition studies of GlcNAc-T II[6] indicating that neither

(46) Fuxreiter, M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 183−194.

GlcNAc nor GlcNAc- -1-phosphate binds to the enzyme, whereas UDP-Glc, UDP−Gal, and UDP−GalNAc bind to the enzyme even though they are not substrates for GnT II. These results show how crucial the UDP part is for the binding of the donor to the enzyme. Only during the sugar transfer reaction does the enzyme recognize the GlcNAc residue. Since UDP-GlcNAc is the GlcNAc-donor substrate for all GnT's, we believe that the specificity of the enzyme must ensue from the structure of the transition state-binding domain that also includes specific information on the sugar-acceptor substrate.

In view of these different elements, one can therefore speculate that the sugar donor-binding site in *N*-acetylglucosaminyltransferases consists of two separate pockets: one pocket serving to accommodate the UDP part of the donor and a second for the binding of the sugar residue that will be transferred during the reaction. Only the UDP pocket would be occupied in the ground state. The sugar pocket of the donor-binding site would become accessible only after the reaction starts and the C1−O1 bond is elongated. Only then, in the transition state of the reaction, would the pocket be fully occupied. This pocket could more precisely be termed as the "sugar transition state pocket". The UDP and sugar pockets should be separated by a distance corresponding approximately to the C1−O1 bond length in the transition state, which can be as large as 3.2 Å, based on our calculations.

The architecture of the uridine-binding site, commonly known as a nucleotide recognition domain (NRD),[47] has been well described for many nucleotide-binding enzymes. A network of interactions involving the uracil and ribose rings, with some conserved amino acids, characterizes this region.[8,11,12] Therefore, it can be envisaged that the topology of the UDP-binding site may be fairly similar in all UDP-utilizing glycosyltransferases. An important feature of the UDP pocket is the presence of a metal cofactor, usually $Mn^{2+}$, which is required by most of the UDP-dependent transferases for activity. The exceptions might appear to be $\beta$-1,6-GlcNAc-T's, since their activity does not depend on the addition of a metal cofactor to the medium, however, we suggest that these enzymes already contain a tightly bound metal ion.[1−4] The divalent cation contributes to the binding of UDP-GlcNAc in the binding site through strong interactions with the pyrophosphate group of the nucleotide. The nature of these interactions presumably determines the conformation adopted by the pyrophosphate group during the reaction. Without a doubt, the metal also plays an important role in the stabilization of the leaving group, UDP, which sees its formal charge changing from 0 to −1 as the C1−O1 bond is cleaved. It has been shown that a particular aspartate from the D*X*D (aspartate−*any amino acid*−aspartate) motif, contained in many glycosyltransferases and involved in the binding of the nucleotide-sugar, is crucial for binding the divalent ion associated with the nucleotide.[48] A complex of UDP·$Mn^{2+}$ was found in the products of galactosyltransferase reactions,[28] leading us to the notion that this particular aspartate residue may also play an important role in the removal of the UDP−metal complex from the reaction site.

Calculation of the reaction pathways clearly showed that during the nucleophilic attack of the acceptor on the anomeric carbon, the C1−O1 linkage elongates accompanied by a conformational rearrangement of the glucopyranose ring. As the catalytic reaction proceeds, modifications at the reaction center move the sugar residue closer to the sugar-binding pocket. Interactions between the sugar and the enzyme benefit from this movement. As reactants get closer to the transition state, interactions with the enzyme increase and become crucial for the stabilization of the TS during the rate-limiting step. Relevant structural features within the enzyme active site should reflect the specificity and the structure of the transition state for a particular enzyme. The specificity for a particular sugar residue that differentiates UDP-GlcNAc, UDP-Glc, UDP−GalNAc, and UDP−Gal is likely produced by a sensitive array of hydrogen bonds, hydrophobic, and electrostatic interactions. Though the structure of the "sugar transition state pocket" is not experimentally known, one can hypothesize that this pocket will likely accommodate specific interactions with the proton-rich part of the half-chair sugar ring. Some amino acid residues located in the neighborhood of the *N*-acetyl group at C2 could also be responsible for the specificity of the enzyme, distinguishing for example between UDP-GlcNAc and UDP-Glc substrates. In the case of A/B glycosyltransferases,[44] the difference for the donor specificity, UDP-Gal versus UDP-GalNAc, has been shown to reside in the different nature of a single amino acid, for example methionine versus leucine, interacting favorably with the *N*-acetyl group of the donor. In the same way, properly oriented amino acids should preferentially interact with the equatorially oriented hydroxyl group at the C4 atom of GlcNAc and not with the axially oriented OH4 of GalNAc.

Concerning the acceptor-binding site, it should reflect the differences appearing in the oligosaccharide-acceptor structures, specific for each *N*-acetylglucosaminyltransferase. In the case of a general base mechanism, a catalytic base, presumably an aspartate or a glutamate residue, is likely to be positioned at a proper distance from the hydroxyl group of the oligosaccharide-acceptor where the sugar transfer will occur. Inhibition studies of GlcNAc-T V have revealed[49] that the α-D-mannopyranosyl residue, to which GlcNAc-T V transfers, is not tightly bound to the enzyme prior to the transfer of GlcNAc from the donor. This flexibility may allow the α-D-mannopyranosyl residue to adopt the optimal position for the nucleophilic attack.

## 4. Conclusions

Despite their extreme importance, the mechanism of glycosyltransferases has not yet been determined. However, we believe that the results of this investigation have enlarged our understanding of this process. The present work uses ab initio quantum chemical methods to explore the potential energy surface for the transfer of GlcNAc by inverting *N*-acetylglucosaminyltransferases. The structural model of the reaction site used in this study consists of all the essential molecules assumed to be involved in the mechanism. All stationary points, transition states, and intermediates revealed from the calculated PES's were characterized at the HF/6-31G*, HF/31++G**//HF/6-31G*, DFT/B3LYP/6-31G*, and DFT/B3LYP/6-31++G**//DFT/B3LYP/6-31G* levels. The multiple transition states along the different reaction pathways were grouped into three groups having common structural features which, in turn, related them to different stages of the reaction. These geometrical differences clearly demonstrate that the optimal design of a transition state analogue inhibitor is dependent on a knowledge of the actual mechanism of a particular enzyme. Among the six different reaction pathways analyzed, a stepwise reaction pathway assuming the enrolment of only a catalytic base, [R → TS1 (13.4) → INT1 (7.6) → TS2 (14.7) → PC1 (−22.4)], appeared to be the most probable reaction path, and is consistent with

(47) Kapitonov, D.; Yu, R. K. *Glycobiology* **1999**, *9*, 961−78.

(48) Hagen, F. K.; Hazes, B.; Raffo, R.; deSa, D.; Tabak, L. A. *J. Biol. Chem.* **1999**, *274*, 6797−6803.

(49) Lu, P. P.; Hindsgaul, O.; Li, H.; Palcic, M. M. *Carbohydr. Res.* **1997**, *303*, 283−291.

the existing experimental data. The mechanism described by such a pathway starts with the nucleophilic attack of the acceptor hydroxyl on the anomeric carbon C1 of the transferred GlcNAc, followed by proton transfer from the acceptor to the catalytic base.

The use of ab initio methods to study enzyme reactions presents a great challenge due to the high complexity and dimensionality of the potential energy surfaces describing this type of mechanism. Nevertheless, the recent progress in computational methods and related technology makes it possible to study more and more complex systems, such as those presented in this work. The results obtained by these methods can supplement experimental data and provide unique information about reaction pathways and the structures of the relevant stationary points observed along the reaction coordinate. Needless to say, such information is vital for the design of drugs inhibiting these enzymes.

JA001525U