JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

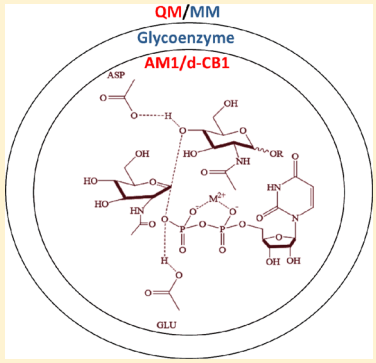
Evaluating AM1/d-CB1 for Chemical Glycobiology QM/MM Simulations

Krishna K. Govender and Kevin J. Naidoo*

Scientific Computing Research Unit and Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

Supporting Information

ABSTRACT: The newly parametrized AM1/d-CB1 is evaluated for its performance in modeling monosaccharide structure, carbohydrate ring pucker, amino acid proton transfer, DNA base pair interactions, carbohydrate−aromatic π interactions, and phosphates that are prominent in glycosyltransferases. The accuracy of the method in these computations is compared to a comprehensive range of NDDO methods commonly used to study glycan structure and reactivity in chemical biology. AM1/d-CB1 shows significant improvement over existing NDDO type methods in the computation of five and six membered carbohydrate ring pucker free energy surfaces. Moreover, the computation of carbohydrate amino acid interactions commonly present in catalytic domains and binding sites are improved over existing NDDO methods. AM1/d-CB1 shows slight improvement for carbohydrate−aromatic π interactions compared to a commonly used NDDO method (AM1). The method is applied to a glycosyltransferase reaction, where it is the only NDDO method able to achieve an optimized reaction profile. Moreover, a comparison of the geometry optimized computations of the reaction scheme give a transition state energy barrier that best compares with DFT (MPW1K). Overall, AM1/d-CB1 is shown to significantly improve on existing NDDO methods in modeling chemical glycobiological events.



1. INTRODUCTION

Carbohydrates or glycans, as they are popularly termed, play a central role in much of biology. The increased appreciation for the ubiquity of glycans and their importance to human health has spawned the field of chemical glycobiology.[1] These molecular systems challenge the limits of experimental investigations making the investigation of molecular mechanisms underlying the many physical and chemical processes in chemical glycobiology via computer simulations a necessity.

Of central interest to chemical glycobiologists is the biosynthesis of complex oligosaccharides that are bonded to glycoproteins and glycolipids. Glycosyltransferases (GTs) constitute a large family of enzymes that are responsible for the biosynthesis of these oligosaccharides/glycans.[2] GTs are highly regio- and stereo- selective enzymes occurring in the cytoplasm of the cell and can be isolated from milk, serum, and organ tissues.[3] Particularly abundant are the GTs that transfer a sugar residue from an activated nucleotide sugar donor, to specific acceptor molecules, forming glycosidic bonds.[2] Transfer of the sugar residue occurs in one of two ways: *retaining*, where the product glycan/glycoside has the same stereochemistry as the activated leaving group, and *inverting*, where the reaction proceeds with inversion of the stereochemistry at the anomeric carbon. A number of studies using substrate analogues have suggested that inverting glycosyltransferase-catalyzed reactions, such as those of glycosidases, proceed via an oxocarbenium ion-like transition state.[4] These enzymes are present in both prokaryotes and eukaryotes and generally display exquisite specificity for both the glycosyl donor and the acceptor substrates.[2] To date, there are limited possibilities to investigate glycosyltransferase reactions experimentally, ranging from X-ray structural analysis of mutated enzymes or enzymes bound to inhibitors,[5] to kinetic isotope effect (KIE) studies.[6] More importantly the nature of the transition state remains inaccessible to experiments. As a result, theoreticians make use of hybrid quantum classical (QM/MM) methods, to try and better understand the mechanistic pathway of these enzymes as well as the conformational and electronic nature of the transition state.[7]

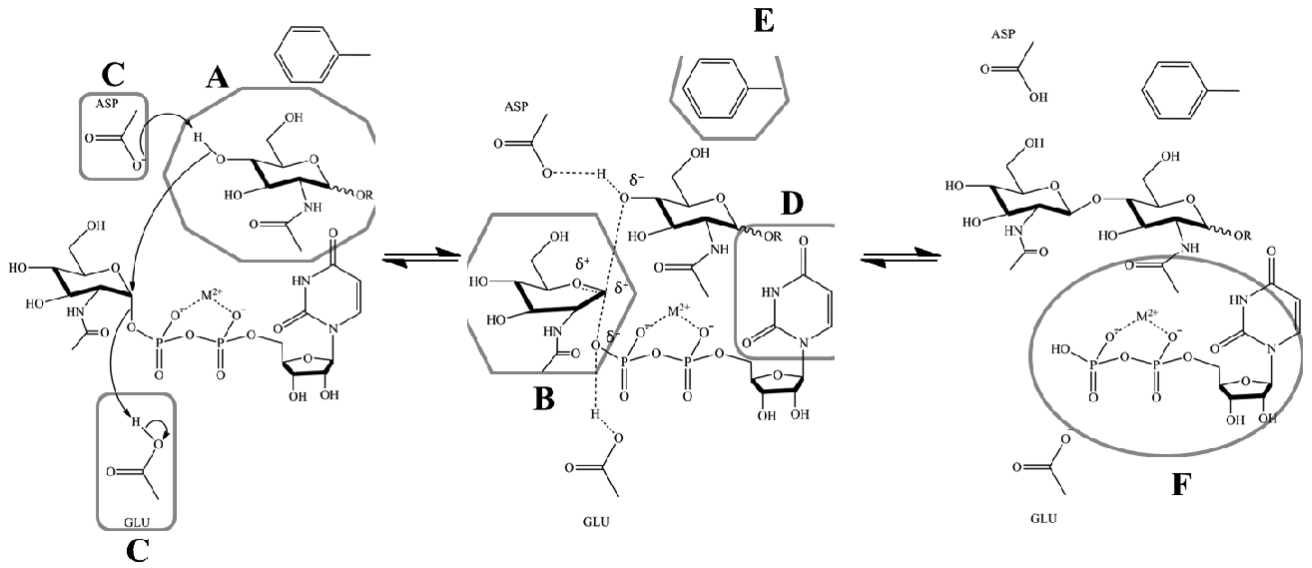
In Scheme 1, we illustrate, using the mechanistic pathway followed by an inverting glycosyltransferase-catalyzed reaction, the key features that must be accurately modeled. These features are illustrated in areas A−F. These correspond to monosaccharide structures and heats of formation, ring puckering and the formation of an oxocarbenium ion in the transition state, proton affinities as observed in amino acid acid/base catalysis, base pairing in both hydrogen bond and π-stacked base pairs, π-stacking of carbohydrate-aromatic stacked systems, and the structure, interaction, and conformation of a nucleoside phosphate.

The accuracy of computer calculations relies on accurate models. When modeling reactions in chemical glycobiology a semiempirical (SE) method must accurately model; (A) the molecular structure of the monosaccharide, (B) conformational (particularly ring puckering) and electronic transition (for-

Received: April 30, 2014
Published: September 4, 2014

dx.doi.org/10.1021/ct500373p | J. Chem. Theory Comput. 2014, 10, 4708−4717

Scheme 1. Mechanism for Inverting Glycosyltransferase Reaction Involving a UDP-GlcNAc Donor and GlcNAc-Containing Acceptor Substrates[a]



[a]Labels represent: A, carbohydrate structure; B, carbohydrate ring pucker; C, proton accepting and donating amino acids; D, base pair interactions; E, carbohydrate-aromatic π stacking; and F, phosphate leaving group.

mation of oxocarbenium ion) of monosaccharides, (C) the ability of amino acids to accept and donate protons, (D) the interactions involved when a base hydrogen bonds and/or π-stacks with molecules in the catalytic domain or with other bases, (E) the interactions of a sugar with neighboring aromatic rings, and (F) the barrier heights required in order for a phosphate group to leave resulting in extended carbohydrate chain formation (oligosaccharides). We have developed a semiempirical (SE) AM1/d-CB1 method to specifically model glycans and more generally biochemical processes of interests in chemical glycobiology taking the above-mentioned features into account.[8] Here, we evaluate the performance of AM1/d-CB1 using these metrics and draw comparisons to NDDO methods (AM1,[9] PM3,[10] PM3CARB-1,[11] PM3^{MS},[12] and RM1[13]) historically used to model carbohydrates. Further, as most of these methods are incapable of correctly modeling hypervalent atoms such as phosphorus we include AM1/d-PhoT,[14] SCC-DFTB,[15] and M06-2X[16] to further measure the relative performance of AM1/d-CB1.

2. THEORY

The theoretical background contextualizing the AM1/d-CB1 development has been detailed previously.[8] The important difference between AM1/d-CB1 and previous methods specifically parametrized for carbohydrate simulations are the inclusion of *d*-orbitals onto hypervalent phosphorus and a core−core repulsion term which is given as

$$E_{AB}^{\mathrm{AM1/d\text{-}CB1}} = E_{AB}^{\mathrm{MNDO}} + \frac{Z_A Z_B}{R_{AB}} \times G_{\mathrm{scale}}^{A} G_{\mathrm{scale}}^{B} \left[\sum_k a_{Ak}\, \mathrm{e}^{-b_{Ak}(R_{AB}-c_{Ak})^2} + \sum_k a_{Bk}\, \mathrm{e}^{-b_{Bk}(R_{AB}-c_{Bk})^2} \right] \quad (1)$$

where G_{scale}^{A} and G_{scale}^{B} are scaling parameters for atoms A and B. These vary from zero to one (values of 0 give the conventional MNDO core−core repulsion, whereas values of 1 give the AM1 core−core model).

The method was developed for chemical glycobiological systems. Here we probe the features as shown in Scheme 1 to illustrate efficacy of AM1/d-CB1 in modeling glycans and catalytic reactivity as they occur in enzymes that act on carbohydrates (glycoenzymes).

3. RESULTS AND DISCUSSION

A. Carbohydrate Structure. There are nine monosaccharides that form the basic alphabet upon which the mammalian glycome (all the carbohydrates in an organism) is constructed. We used the M06-2*X*/6-311++G(3df,2p) (DFT) level of theory to get optimized structures for these nine monosaccharides (Figure S1, Supporting Information). The relative heats of formation acquired from DFT, where all energies were computed relative to β-D-glucose, are compared to those generated from the SE calculations (Tables S1−S2, Supporting Information). The structures obtained from DFT were used in both single point and geometry optimization calculations from which heats of formation were calculated. Methods that have been specifically parametrized for carbohydrates (PM3CARB-1 and PM3^{MS}) give rise to the largest Mean Unsigned Errors (MUEs), with single point errors of 104.29 and 194.45 kcal/mol, respectively (Table S1, Supporting Information). The MUEs, 104.79 and 196.31 kcal/mol respectively, do not improve following geometry optimization (Table S2, Supporting Information). AM1/d-CB1 substantially outperforms all of the methods (Figure 1) exhibiting an error of 3.64 kcal/mol for single point energies (Table S1, Supporting Information). In the case of geometry optimized structures, AM1/d-CB1 is second only to RM1 (Figure 1) with a MUE of 5.02 kcal/mol (Table S2, Supporting Information).

A comparison of the optimized coordinates with those obtained from the DFT simulations were used to compute the root-mean-square deviations (RMSD) of the individual monosaccharides (Table S3, Supporting Information). AM1, PM3, RM1, and AM1/d-CB1 minimize to geometries that are very similar to those of DFT. Interestingly the methods

dx.doi.org/10.1021/ct500373p | J. Chem. Theory Comput. 2014, 10, 4708−4717

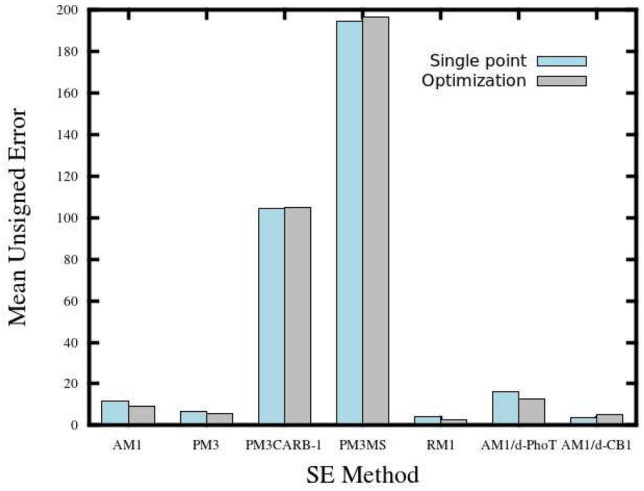


Figure 1. Relative mean unsigned errors in heats of formation (kcal/mol) for nine monosaccharaide using both single point and geometry optimization, on the DFT optimized structures.

parametrized for carbohydrates (PM3CARB-1 and PM3^{MS}) give the poorest optimized structures.

B. Carbohydrate Ring Pucker from Free Energy Simulations. The generalized free energy approach termed Free Energy from Adaptive Reaction Coordinate Forces (FEARCF)[17] was used with various SE based methods (including AM1/d-CB1) to calculate the free energy of ring pucker[17b] surface for β-D-ribofuranose and free energy ring pucker volume for β-D-glucopyranose. The accuracy of the computed reaction mechanism depends on the ability of the model to simulate ring pucker (Scheme 1, label B). The simulations conducted in this work followed the same methodology as described earlier[18] with the only differences being that CHARMM[19] v35b5 was used, instead of v33b2, and the QM MD simulations were conducted with the CHARMM/MNDO97[20] interface.

B.1. Ribofuranose. To compute the furanose ring pucker conformations, the ring is subdivided into a reference plane and two rotatable planes using a triangular tessellation method.[21] The free energy of pucker is then computed as a function of the angles (θ_0, θ_1) that the rotatable planes make with the reference plane.[17b,18,22] Canonical conformers of the furanose ring, that is, envelopes (E) and twists (T), are illustrated in terms of nodes (θ_0, θ_1) in Figure 2a significantly distanced away from the center of the surface (0,0) that corresponds to a flattened ring.

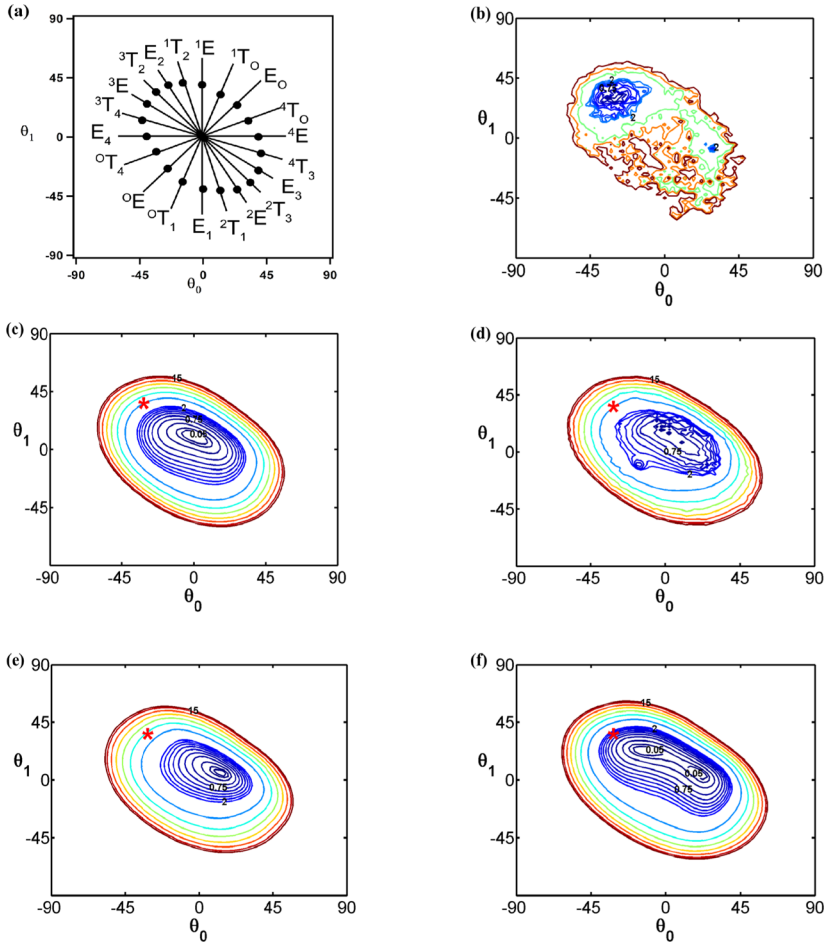


Figure 2. (a) Triangular tessellation pucker space for five-membered rings with canonical conformer coordinates shown as nodes. Ribofuranose free energy of puckering shown as two-dimensional contour plots for (b) HF/6-31G, (c) PM3CARB-1, (d) PM3^{MS}, (e) AM1/d-PhoT, and (f) AM1/d-CB1. Energy is mapped to color from 0 kcal/mol (blue) to 15 kcal/mol (red). Contours are shown at 0.05 to 0.1 kcal/mol, then from 0.1 to 2 kcal/mol in steps of 0.25 kcal/mol and every 2 kcal/mol thereafter. The HF global energy minimum (shown as red stars) is marked on each SE FES.

dx.doi.org/10.1021/ct500373p | J. Chem. Theory Comput. 2014, 10, 4708−4717

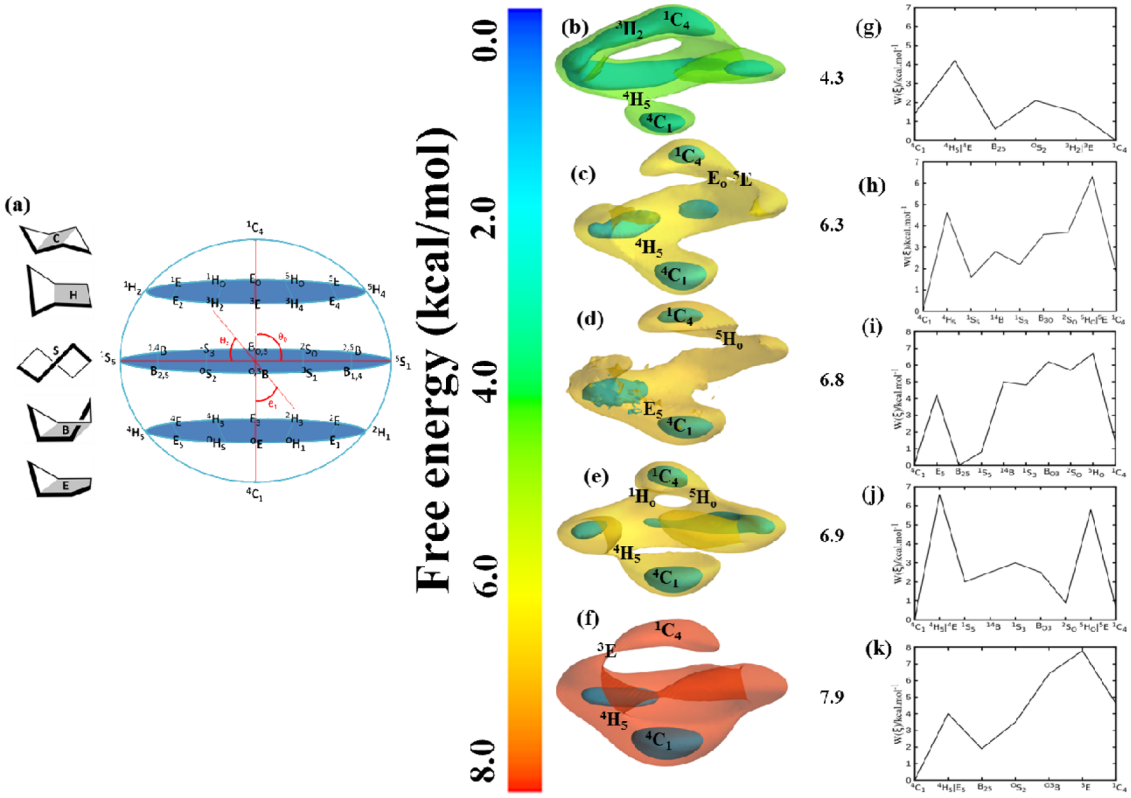


Figure 3. (a) Canonical conformers projected onto the triangular tessellated pucker coordinates (θ_0, θ_1, θ_2) for six-membered rings. The free energy $W(\theta_0, \theta_1, \theta_2)$ volumes for (b) AM1/d-PhoT, (c) AM1/d-CB1, (d) PM3^{MS}, (e) SCC-DFTB, and (f) PM3CARB-1 are shown on color. The free energy values are mapped in color from 0 kcal/mol (blue) to 8 kcal/mol (red). The inner isosurface is at 3 kcal/mol and the outer isosurface indicates the minimum free energy to connect the $^1C_4$ and $^4C_1$ conformers which occurs at 4.3, 5.3, 6.8, 6.9, and 7.9 kcal/mol, respectively. The one-dimensional minimum free energy paths have been extracted from the free energy volumes and are shown for (g) AM1/d-PhoT, (h) AM1/d-CB1, (i) PM3^{MS}, (j) SCC-DFTB, and (k) PM3CARB-1.

The Hartree−Fock (HF) free energy of pucker surface had been computed using 6-31G basis set (Figure 2b).[18] HF free energy computations require vast amounts of compute cycles as a result the surface is not converged. Nonetheless, the HF free energy of pucker for the furanose ring does reveal distinct minima interpretable as canonical puckered conformers. The global minimum (−35°, 30°) is a $^3T_2/^3E$ conformer. A second minimum exists at (27.5°, −7.5°) that is approximately 0.95 kcal/mol higher in energy than the global minimum.

Previously, we compared AM1, PM3, PM3CARB-1, and SCC-DFTB free energy of pucker surfaces.[18] We showed that common to all NDDO methods is a large minimum energy well with no distinct global minimum. More serious is the lack of energetic differentiation on the free energy surface (FES) that is evidence of discrimination between different puckering conformers. The shapes of NDDO (AM1, PM3, or PM3CARB-1) pucker FES' are indicative of furanose ring models that are flexible and can pucker relatively easily at room temperature and conformationally indiscriminate. For example PM3^{MS} (Figure 2d) has a global energy minimum of (−2.5°, 17.5°) that is close to the planar conformer with the nearest canonical conformer being ^1E. As with other SE methods PM3^{MS} bowl shaped FES can access several conformers where the canonical ^1T_o, E_o, ^4T_o, ^4E lie on the periphery within 0.5 kcal/mol of the near flattened favored ring conformer. This continues to be the case for AM1/d-PhoT (Figure 2e). Here, the global minimum (15°, 5°) is closest to a ^4T_o conformer with ^1E, ^1T_o, E_o, ^4T_o, and ^4E being 0.5 kcal/mol away at rim of the FES. NDDO methods therefore lead to models incapable of accurately describing distinct transitions states for furanose rings.

The AM1/d-CB1 (Figure 2f) global minimum (−12.5°, 22.5°) represents $^1T_2/E_2$ conformers that are in close proximity to the reference HF $^3T_2/^3E$ global minima conformers as well as the ^3E conformer that is observed in the solid state for the substituted ribose ring in phenylalanine tRNA.[23] A second minimum (20°, 5°) corresponding to a $^4T_o/^4E$ conformation with an energy of 0.01 kcal/mol can be accessed via the $^1T_o/E_o$ conformer (7.5°, 17.5°) of energy 0.09 kcal/mol of its global minimum. Approximately 0.8 kcal/mol is required to reach the planar conformation on the AM1/d-CB1 FES. The AM1/d-CB1 furanose 5-membered ring is more discriminate of canonical conformers and improves on NDDO methods commonly used to model carbohydrates.

B.2. Glucopyranose. For glucopyranose, the free energy volumes are a function of (θ_0, θ_1, θ_2) with reference and rotatable planes chosen using a method of triangular tessellation as described previously.[17b,18,21a] The 38 canonical conformers for six-membered rings comprising chairs (C), boats (B), twists/skews (S), half-chairs (H), and envelopes (E) are shown in Figure 3a.

The free energy of glucopyranose as a function of reaction coordinates $W(\theta_0, \theta_1, \theta_2)$ is visualized in the three dimensions of the reaction coordinates representing the free energy in color where low free energies (0 kcal/mol) are blue and very high free energies (8 kcal/mol) are red (Figure 3b-k). In each of the free energy landscapes, there are two isovolumes. The first is an inner (turquois) surface at 3 kcal/mol (~5kT) representing

dx.doi.org/10.1021/ct500373p | J. Chem. Theory Comput. 2014, 10, 4708−4717

pucker conformers observed at equilibrium dynamics. The second (outer) surface encases minimum energy pathways from the $^4C_1$ conformation (south pole), where all hydroxyl groups are equatorial, to the $^1C_4$ conformation (north pole), where all hydroxyls are axial. The lowest energy conformer predicted for all methods, other than AM1/d-PhoT, is $^4C_1$. At 3 kcal/mol AM1/d-PhoT shows the existence of skew boats and boats ($^{O,3}B$, $^OS_2$, $B_{2,5}$, $^1S_5$, $^{1,4}B$, $^1S_3$, $^{2,5}B$, and $^5S_1$) as well as the $^3H_2$ and ^3E conformers (Figure 3b).

AM1/d-CB1, PM3^{MS}, PM3CARB-1, and SCC-DFTB provide more restricted minimum energy paths (Figure 3c–f) compared to AM1/d-PhoT. The barrier heights separating $^4C_1$ from $^1C_4$ for AM1/d-PhoT are the lowest (4.3 kcal/mol) while AM1/d-CB1, PM3^{MS}, SCC-DFTB, and PM3CARB-1 have barrier heights at least 1 kcal/mol are higher (6.3, 6.8, 6.9, and 7.9 kcal/mol, respectively).

Minimum free energy paths have been extracted and plotted as line diagrams (Figure 3g–l). In the absence of an ab initio computed free energy pucker volume, the SCC-DFTB volume and derived minimum path $^4C_1 \rightarrow {}^4H_5/{}^4E \rightarrow {}^1S_5/{}^{1,4}B \rightarrow {}^1S_3 \rightarrow B_{O,3} \rightarrow {}^2S_O \rightarrow ({}^5H_O/{}^5E \text{ or } {}^5H_O) \rightarrow {}^1C_4$ (Figure 3j) is used as a reference conformational mechanism for the $^4C_1$ to $^1C_4$ ring pucker transition. We do this as we had previously established that SCC-DFTB, of any SE method, best models carbohydrate ring pucker.[18] AM1/d-CB1 directly matches the SCC-DFTB minimum free energy pathway (Figure 3h) from the $^4C_1$ to $^1C_4$ conformer although the energy profile differs between the two methods. In the AM1/d-CB1 case the *H* and *E* conformers populating the southern "tropic" commonly associated with TS structures in GT catalyzed reactions have a barrier height of 4 kcal/mol. Whereas the same SCC-DFTB computed $^4H_5/{}^4E$ conformers are more than 6 kcal/mol higher on the free energy pucker volume than AM1/d-CB1. Along the northern "tropic" the H and E conformers have a barrier heights of more than 6 kcal/mol that are close to the 5.8 barrier predicted by SCC-DFTB.

Ideally, it would be preferable to benchmark against dynamic ab initio results from which free energy volumes were constructed. This, however, is not within reach of current computational resources. In place of this, we compare the SE results with those obtained by dynamical density functional theory (DFT) calculations with the project augmented-wave (PAW) method.[24] Here, it was found that the transition state for permethylated glucose from the $^1C_4$ chair to $^{1,4}B$ was a ^1E envelope conformation. From the envelope-like transition state, the system evolved into the $^{1,4}B$ minimum (TS observed for AM1/d-CB1, SCC-DFTB, and PM3^{MS}). A $^1S_5$ skew-boat minimum was also discovered (minimum along the 1D surfaces of both AM1/d-CB1 and SCC-DFTB).

The AM1/d-PhoT minimum path $^4C_1 \rightarrow {}^4H_5/{}^4E \rightarrow B_{2,5} \rightarrow {}^OS_2 \rightarrow {}^3H_2/{}^3E \rightarrow {}^1C_4$ (Figure 3g) initially puckers in the same way as SCC-DFTB and AM1/d-CB1 with a $^4H_5/{}^4E$ conformation barrier of 4 kcal/mol but the glucopyranose ring crosses the B and S conformers at the "equator" and transitions to the $^1C_4$ conformer very differently from the former two methods.

The PM3^{MS} minimum path (Figure 3i) deviates perhaps the furthest from the SCC-DFTB both in its puckering as well as the very high barriers (6–7 kcal/mol) it passes through, toward the $^1C_4$ conformer. Moreover, the global minima on the pucker free energy surface is the $B_{2,5}$ conformer. This is inconsistent with accepted stereo electronic understanding of glucopyranose 6-membered ring pucker.[25]

The PM3CARB-1 minimum path (Figure 3k) passes through a $^4H_5/E_5$ 4 kcal/mol barrier to get to a $B_{2,5}$ minimum. From $B_{2,5}$ the ring is transformed into $^OS_2$, $^{O,3}B$ and ^3E conformers with increasingly large pucker energy. The $^1C_4$ conformer (4.6 kcal/mol) is higher in energy than the boats and skew-boats in the PM3CARB-1 free energy volume.

Three-dimensional free energy volumes as well as extracted one-dimensional free energy paths for AM1, PM3, and RM1 are provided in the Supporting Information (Figure S3).

C. Proton Transfer. A GT reaction (Scheme 1) proceeds with the aid of a catalytic acid/base. Amino acid residues within the GT catalytic domain (label C) undergo proton transfers with the saccharide. Given its importance in the catalytic process, the proton affinity of amino acids commonly found in the GT catalytic domain as well as in the binding sites of lectins were computed using AM1/d-CB1 as well as methods commonly used to model carbohydrates. In addition we also consider the interaction energies of a number of base pairs that are applicable to DNA, RNA, and numerous enzymes.

The proton affinities of 18 amino acids (illustrated in Figure S4 of the Supporting Information) were computed in two ways. The first involves the protonation of the nitrogen on the amino group of a neutral amino acid leading to a positively charged species. Here, reference values were obtained from Gronert et al.,[26] as well as experimentally available data posted in the NIST database.[27] The second involves protonating the oxygen of the negatively charged carboxylate leading to a neutral amino acid. Here, we used CBS-QB3 calculations computed using Gaussian 09,[28] that had been shown to be accurate,[14,29] to generate reference proton affinities values. All SE proton affinities were computed with the aid of an experimentally determined heat of formation for the proton (365.7 kcal/mol).[30] In the case of nitrogen protonation, geometry optimization calculations on G3MP2 geometries were computed while in the case of oxygen protonation geometry optimization calculations on the CBS-QB3 geometries were computed. All calculations were done using a modified version of the MOPAC7.0 software package. The PM3CARB-1, PM3^{MS}, AM1/d-PhoT, and AM1/d-CB1 MUEs of both the N- and O-protonated amino acids are shown in Figure 4. AM1/d-PhoT has the smallest error for the protonation of nitrogen with a value of 3.5 kcal/mol (Table S4, Supporting Information), with AM1/d-CB1 having a slightly

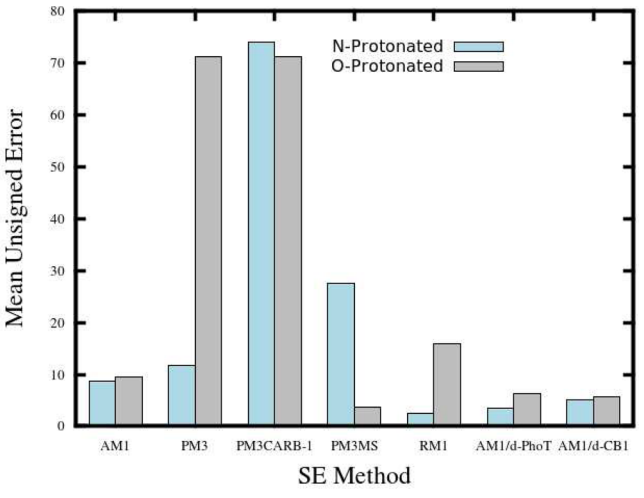


Figure 4. Geometry optimized mean unsigned errors (kcal/mol) for gas phase proton affinities of N- and O-protonated amino acids.

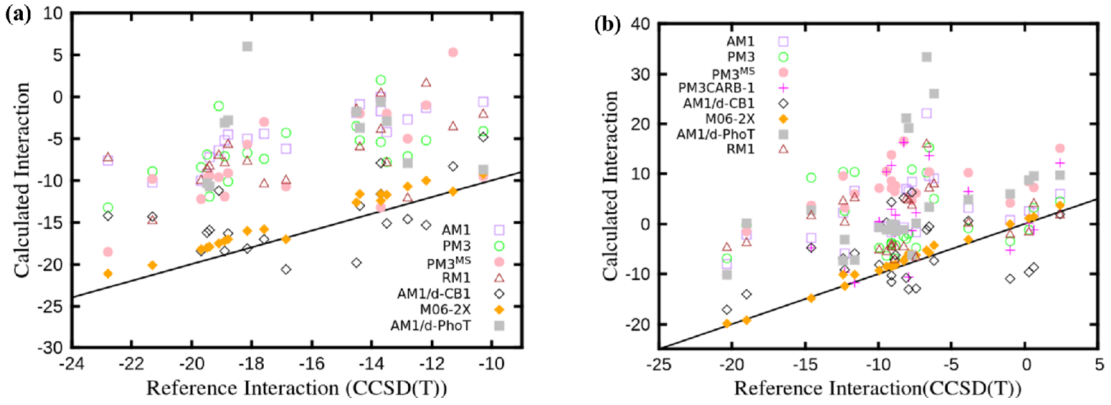


Figure 5. Comparison of NDDO and DFT/M06-2X gas-phase interaction energies for (a) hydrogen bonded and (b) stacked base pairs. The reference interaction energies are from CCSD(T) simulations. All interaction energies (kcal/mol) computed via geometry optimization.

larger error (MUE of 5.1 kcal/mol). PM3CARB-1 and PM3^{MS} are significantly inaccurate with errors of 74.0 and 27.5 kcal/mol, respectively.

AM1/d-CB1 and PM3^{MS} have the lowest errors 5.6 and 3.6 kcal/mol, respectively, for the anionic carboxylate protonation (Table S4, Supporting Information) yielding values closest to the CBS-QB3 reference set. AM1/d-PhoT's performance is reasonable good with MUE of 6.3 kcal/mol. PM3CARB-1, however, has the extremely large MUE of 71.2. It is interesting to note that all SE methods that have too low average errors for the protonation of the amino acid nitrogen have poor performance for the protonation of oxygen. In contrast AM1/d-CB1 produces similar results for both species. This implies that the method is likely to model the amino acids within the GT catalytic domain well.

D. Nucleic Acid Base Stacking and Hydrogen Bonding. A number of hydrogen bonded and stacked nitrogeneous base pairs were obtained from a benchmark database.[31] Both single point and geometry optimization calculations on the CCSD(T) optimized geometries using PM3CARB-1, PM3^{MS}, AM1/d-CB1, and AM1/d-PhoT were computed. The resulting interaction energies were then compared with the CCSD(T) values.

The distribution of computed interaction energies for hydrogen bonded and stacked base pairs (structures shown in Supporting Information Figures S5–S6) are shown in Figure 5. A comparison is made with DFT M06-2X energies computed by Hohenstein et al.[32] In general, AM1/d-CB1 compares well with the CCSD(T) computed HB base pair interactions (MUE 3.63 kcal/mol) although it slightly underestimates some of the interactions while all other NDDO methods significantly overestimates the hydrogen bonding (Figure 5a). Of the NDDO methods, AM1/d-CB1 compares best (MUE 3.87 kcal/mol) with the CCSD(T) stacked base pairs data (Figure 5b). In both the hydrogen bonded and stacked base pairs calculations, AM1/d-CB1 performs optimally, comparing with the CCSD-(T) reference by least 2 orders of magnitude better than PM3CARB-1, PM3^{MS}, and AM1/d-PhoT (Tables S5–S6, Supporting Information).

None of the SE methods produce results as accurate as those of M06-2X/aug-cc-pVDZ[32] with errors of 1.61 and 0.98 kcal/mol for the hydrogen bonded and stacked complexes, respectively (Table S5, Supporting Information). The comparison with of M06-2X with AM1/d-CB1 and other NDDO methods is not quite fair as the former includes corrections for dispersion and HB. In the case of the SE methods, these interactions need to be accounted for empirically. It would be beneficial to include both empirical based dispersion and HB corrections to the AM1/d-CB1 Hamiltonian, as it is expected to result in smaller errors for the complexes considered here and would be more comparable to the M06-2X methodology. This is the subject of work, which we will present, in a later publication.

Numerous single point calculations have also been performed, but not shown here (Supporting Information Figure S7 and Tables S7–S8).

E. Carbohydrate–Aromatic π Interactions. Amino acids with aromatic side chains, such as tryptophan, tyrosine, and phenylalanine form the basis of carbohydrate protein binding interactions and so are frequently found in protein active sites that recognize carbohydrates.[33] That adds a further dimension to the proton acceptor/donor properties to be accounted for in the glycoenzyme catalytic domains. Raju et al. conducted a number of calculations[33,34] where they modeled the carbohydrate-aromatic interactions with high level ab initio computations. They used model complexes comprising carbohydrates interacting with toluene, p-hydroxytoluene, and 3-methylindole (analogues of phenylalanine, tyrosine, and tryptophan, respectively). Here we use the QM optimized carbohydrate-aromatic interacting complexes generated by Raju et al.[33,34] and conduct SE calculations on these complexes (Figure 6).

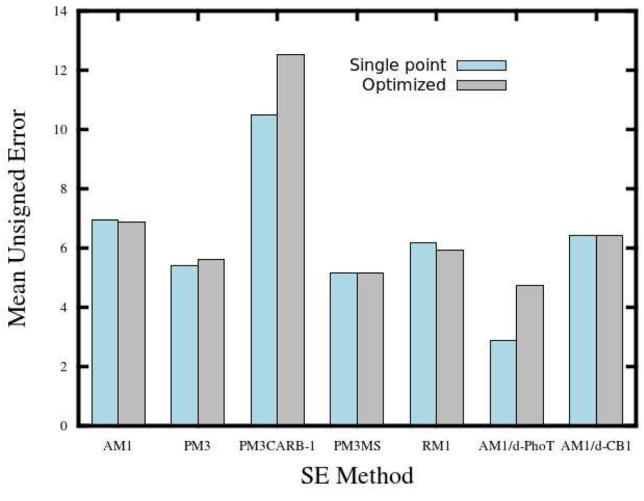


Figure 6. All NDDO MUEs computed for ab initio generated structures[33,34] of carbohydrate–aromatic π interactions (kcal/mol).

dx.doi.org/10.1021/ct500373p | J. Chem. Theory Comput. 2014, 10, 4708–4717

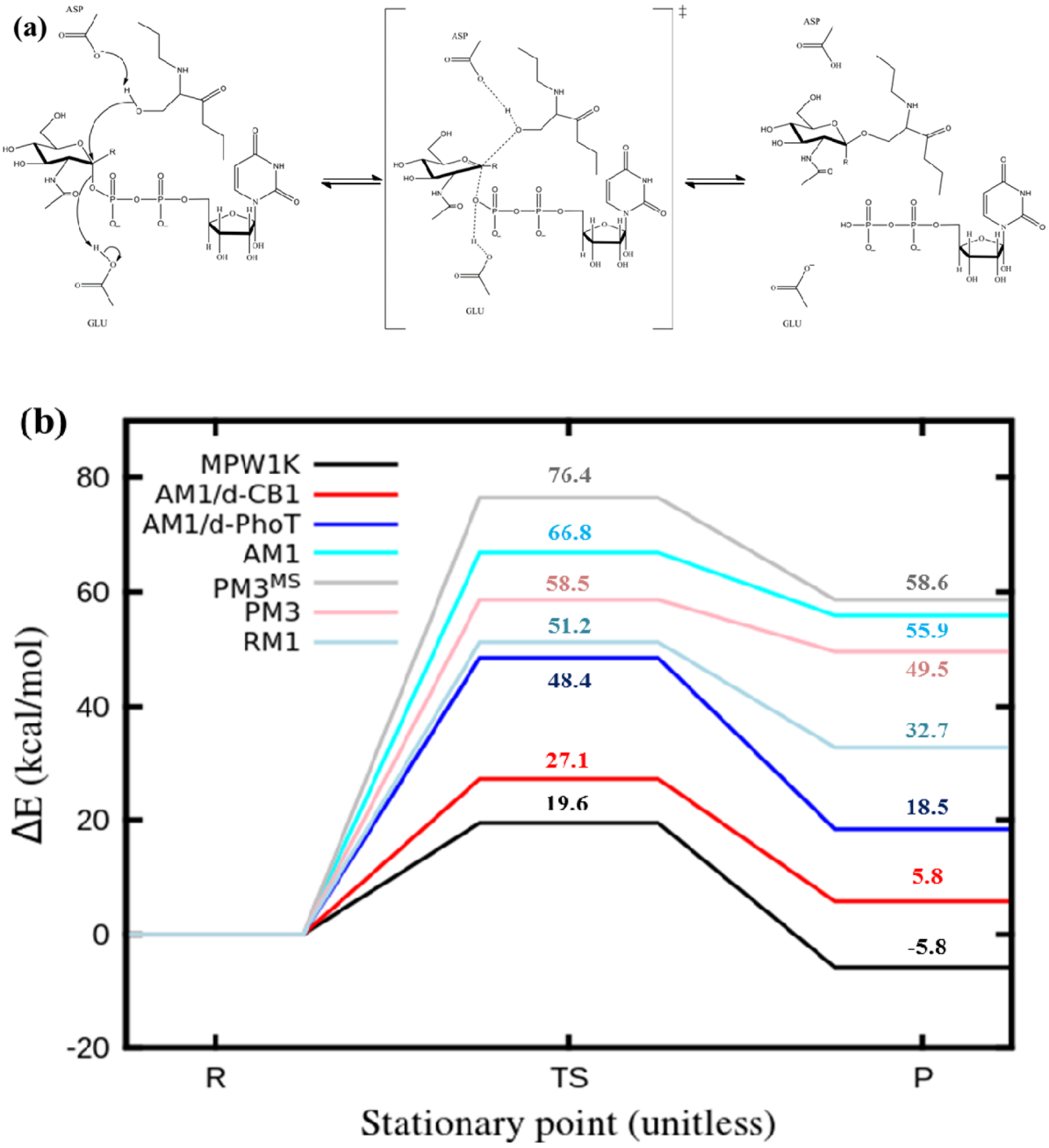


Figure 7. (a) Reaction scheme for enzymatic reaction catalyzed by uridine diphospho-N-acetylglucosamine polypeptide β-N-acetylaminyltransferase and (b) geometry optimized QM/MM 1D reaction profile energy traces. MPW1K profile was obtained from work by Tvaroška et al.[35]

AM1/d-PhoT gives the smallest errors for both single point (2.89 kcal/mol) and geometry optimized (4.76 kcal/mol) structures (Table S9−S10, Supporting Information). AM1/d-CB1 performs poorly with a MUE of 6.44 kcal/mol for both single point and geometry optimized interaction energies. For the complexes considered here, the predominant interaction is that of dispersion and we anticipate a significant improvement, specifically for AM1/d-CB1, gained form an empirical based dispersion correction added to the existing Hamiltonians.

F. GT Reactions. Up until now, we have evaluated AM1/d-CB1's performance for separate components that contribute to the accurate modeling of chemical glycobiological events particularly those central to a glycoenzymatic reactions. A typical glycosyltransferase reaction featuring an inversion or retention mechanism at the anomeric position requires a nucleoside phosphate, nucleoside diphosphate, or lipid phosphate leaving group present as illustrated in Scheme 1, label F.

Recently, a QM(DFT)/MM investigation into the substrate-assisted catalytic mechanism of O-linked N-acetylglucosamine (O-GlcNAc) transferase was reported by Tvaroška et al.[35] A schematic for the reaction that includes surrounding residues (UDP-GlcNAc, Val20, Ser21, Ser22, His498, His558, Gln839, Lys842, Lys898, His901, His920, and three water molecules in the vicinity of UDP-GlcNAc) is given in Figure 7a. We obtained the QM/MM geometries for the reactant (R), transition state (TS), and product (P) that were computed at the MPW1K level of theory using a combination of the 6-31G** and 6-31+G* basis sets,[35] as well as the OPLS force field for the classical region.[36] Both single point and geometry

dx.doi.org/10.1021/ct500373p | J. Chem. Theory Comput. 2014, 10, 4708−4717

optimized QM/MM calculations of these structures were then performed using various SE methods (Table B11) contained within the CHARMM/MNDO97[20] interface. The MM was treated with the OPLS force field[36] in order to ensure that the calculations are in line with those performed by Tvaroška et al.[35] The energy barriers for the reaction are provided in Figure 7b. The energies obtained for the MPW1K level are 19.6 kcal/mol (TS) and −5.8 kcal/mol (P) relative to R. It is important to note that the QM region defined by Tvaroška et al.[35] consisted of a total of 198 atoms, which was too large for the MNDO97 software package.

The SE QM/MM calculations were performed with a smaller QM region comprising 81 atoms (UDG-GlcNAc, Ser21, and His498). The barriers obtained from the geometry optimized QM/MM calculations are provided in Figure 7b as well as Table S11 of the Supporting Information. AM1/d-CB1 surpasses all of the other NDDO methods in predicting both the TS barrier and the relative P energies of 27.1 and 5.8 kcal/mol, respectively. Moreover the optimized structures do not differ much from those obtained by the MPW1K method (see RMSDs in Table S12 Supporting Information). AM1 and AM1/d-CB1 have the smallest RMSD's for the N-acetylglucosamine reactant (0.012345 and 0.013110, respectively) conformation and product (0.011491 and 0.012373, respectively) conformation. AM1/d-CB1 gives the best transition state N-acetylglucosamine conformation with an RMSD of 0.018694. The SE methods had no change in transition state pucker ($^4H_3$) after a geometry optimization.

AM1, AM1/d-PhoT, and AM1/d-CB1 produce bond reaction coordinate distances that are closest to the MPW1K results (Table S13). However, AM1/d-CB1 OG-C_1' (the bond that is being formed) bond in the TS configuration is almost 0.10 Å away from that seen in the MPW1K structure. Single point QM/MM calculations are presented in Table S11 (Supporting Information).

4. CONCLUSION

The applicability of the newly parametrized AM1/d-CB1[8] SE method to chemical glycobiology has been evaluated. The method gives accurate results for molecular structures of monosaccharides that are important in mammalian biology. AM1/d-CB1 also shows considerably different behavior for both 5- and 6-membered ring puckering when compared to other NDDO type methods (AM1, PM3, PM3CARB-1, PM3^{MS}, RM1, and AM1/d-PhoT) producing more than just one minimum energy conformer for the ribofuranose (5-membered) ring. This addresses a common weakness exhibited by NDDO methods where the pucker free energy surface of five membered sugar rings display no stationary points that correlate with canonical conformers. Further NDDO methods favor flattened five membered sugar rings showing no canonical ring pucker preferences. This is unlike AM1/d-CB1, which discriminates between canonical pucker conformers and presents a global minimum distinctly shifted away from the flattened ring. In the case of glucopyranose, the ring is no longer unrealistically flexible as is the case with AM1, PM3, and PM3^{MS}; instead, AM1/d-CB1 yields a $^4C_1$ to $^1C_4$ ring pucker pathway that directly matches the SCC-DFTB $^4C_1$ to $^1C_4$ pathway. Comparing with SCC-DFTB is best since no ab initio free energy pucker volume benchmarks exists.

While the accurate modeling of carbohydrates by AM1/d-CB1 is an important measure of its performance, it is critical to examine its performance when computing key properties of molecules often present in the glycan environment. An example of this is the amino acid proton affinities that are central to the acid/base catalytic mechanism commandeered by GTs and other glycoenzymes. Here, AM1/d-CB1 (as well as AM1/d-PhoT) yields the most accurate proton affinities for the protonation of nitrogen amino group of a neutral amino acid. In the case of oxygen protonation, AM1/d-CB1 gives the most accurate results when compared to other NDDO type methods. Therefore, overall, we expect AM1/d-CB1 to reliably model the acid/base contributions to the glycosidase or glycosyltransferase catalytic mechanisms.

Another important feature of carbohydrate protein interactions is the role that aromatic groups play in the binding of glycans in glycoenzyme catalytic domains as well as in the recognition sites of proteins such as lectins. Here, AM1/d-CB1 is expected to produce more accurate results when coupled with a post-SCF dispersion based correction and then compared with benchmark calculations done with DFT-D/TZVDZ.

The AM1/d-CB1 parameter set is by no means perfect although it achieves better performances for the range of chemical glycobiological important property calculations than other NDDO methods. It suffers from the same perennials shortcomings underlying the NDDO method regarding hydrogen bond and dispersion interaction modeling. Therefore, we believe that AM1/d-CB1 when coupled with specifically optimized post SCF semiempirical HB and dispersion corrections will be capable of delivering optimal semiempirical performance for chemical glycobiological events. These corrections are currently under development.

Not withstanding the current lack of HB and dispersion corrections the overall performance of AM1/d-CB1 as a reliable SE method for chemical glycobiology applications is apparent when we used it to compute the reaction profile of a recently studied GT. Here, it gave the lowest transition state and product barrier compared with DFT (MPW1K) computations.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Provided are figures containing monosaccharide structures; additional 2D and 3D free energy surfaces; various stacked base pairs; single point interaction energies, and tables with errors for molecules used during testing. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Email: kevin.naidoo@uct.ac.za.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is based on research supported by the South African Research Chairs Initiative (SARChI) of the Department of Science and Technology and National Research Foundation (NRF) to K.J.N. K.K.G. thanks the NRF and Equity Development Program (EDP), Department of Chemistry, University of Cape Town for doctoral support. We thank the Center for High Performance Computing (Cape Town, SA) for the provision of computational resources. Parameters for the AM1/d-CB1 Hamiltonian can be found at http://www.scientificomputing.com.

REFERENCES

(1) Kiessling, L. L.; Splain, R. A. Chemical approaches to glycobiology. *Annu. Rev. Biochem.* **2010**, *79* (1), 619−653.

(2) Breton, C.; Šnajdrová, L.; Jeanneau, C.; Koča, J.; Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **2006**, *16* (2), 29R−37R.

(3) Boons, G.-J.; Demchenko, A. V. Recent advances in O-sialylation. *Chem. Rev.* **2000**, *100* (12), 4539−4565.

(4) (a) Murray, B. W.; Takayama, S.; Schultz, J.; Wong, C.-H. Mechanism of human α-1,3-fucosyltransferase V: Glycosidic clevage occurs prior to nucleophilic attack. *Biochemistry* **1997**, *36*, 823−831. (b) Burkart, M. D.; Vincent, A.; Duffels, B. W.; Ley, S. V.; Wong, C.-H. Chemo-enzymatic synthesis of fluorinated sugar nucleotide: Useful mechanistic probes for glycosyltransferases. *Bioorg. Med. Chem.* **2000**, *8*, 1937−1946. (c) Hartman, M. C. T.; Jiang, S.; Rush, J. S.; Waechter, C. J.; Coward, J. K. Glycosyltransferase mechanisms: Impact of a 5-fluoro substituent in acceptor and donor substrates on catalysis. *Biochemistry* **2007**, *46* (41), 11630−11638.

(5) (a) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schülein, M.; Davies, G. J. Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. *Biochemistry* **1996**, *35* (48), 15280−15287. (b) Vasella, A.; Davies, G. J.; Böhm, M. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **2002**, *6* (5), 619−629.

(6) Werner, R. M.; Stivers, J. T. Kinetic isotope effect studies of the reaction catalyzed by uracil DNA glycosylase: Evidence for an oxocarbenium ion−uracil anion intermediate. *Biochemistry* **2000**, *39* (46), 14054−14064.

(7) (a) Biarnés, X.; Ardèvol, A.; Iglesias-Fernández, J.; Planas, A.; Rovira, C. Catalytic itinerary in 1,3-1,4-β-glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* **2011**, *133* (50), 20301−20309. (b) Gómez, H.; Polyak, I.; Thiel, W.; Lluch, J. M.; Masgrau, L. Retaining glycosyltransferase mechanism studied by QM/MM methods: Lipopolysaccharyl-α-1,4-galactosyltransferase C transfers α-galactose via an oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* **2012**, *134* (10), 4743−4752. (c) Gómez, H.; Lluch, J. M.; Masgrau, L. Substrate-assisted and nucleophilically assisted catalysis in bovine α1,3-galactosyltransferase. Mechanistic implications for retaining glycosyltransferases. *J. Am. Chem. Soc.* **2013**, *135* (18), 7053−7063. (d) Tvaroška, I.; André, I.; Carver, J. P. 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. *J. Am. Chem. Soc.* **2000**, *122* (36), 8762−8776. (e) Tvaroška, I.; André, I.; Carver, J. P. Catalytic mechanism of the inverting N-acetylglucosaminyltransferase I: DFT quantum mechanical model of the reaction pathway and determination of the transition state structure. *Glycobiology* **2003**, *13* (8), 559−566. (f) Tvaroška, I. Molecular modeling insights into the catalytic mechanism of the retaining galactosyltransferase LgtC. *Carbohydr. Res.* **2004**, *339* (5), 1007−1014. (g) Raab, M.; Kozmon, S.; Tvaroška, I. Potential transition-state analogs for glycosyltransferases. Design and DFT calculations of conformational behavior. *Carbohydr. Res.* **2005**, *340* (5), 1051−1057.

(8) Govender, K. K.; Gao, J.; Naidoo, K. J. AM1/d-CB1: A semi-empirical method for the QM/MM simulations of chemical glycobiology systems. *J. Chem. Theory Comput.* **2014**, DOI: 10.1021/ct500372s.

(9) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(10) (a) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209−220. (b) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods. 2. Applications. *J. Comput. Chem.* **1989**, *10* (2), 221−264.

(11) McNamara, J. P.; Muslim, A.-M.; Abdel-Aal, H.; Wang, H.; Mohr, M.; Hillier, I. H.; Bryce, R. A. Towards a quantum mechanical force field for carbohydrates: a reparameterized semi-empirical MO approach. *Chem. Phys. Lett.* **2004**, *394*, 429−436.

(12) Mane, J. Y.; Klobukowski, M. New parameterization of the PM3 method for monosaccharides. *Chem. Phys. Lett.* **2010**, *500* (1−3), 140−143.

(13) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101−1111.

(14) Nam, K.; Cui, Q.; Gao, J.; York, D. M. Specific reaction parameterization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.* **2007**, *3*, 486−504.

(15) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. J. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* **2001**, *105* (2), 569−585.

(16) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(17) (a) Strümpfer, J.; Naidoo, K. J. Computing free energy hypersurfaces for anisotropic intermolecular associations. *J. Comput. Chem.* **2010**, *31*, 308−316. (b) Barnett, C. B.; Naidoo, K. J. Free Energies from Adaptive Reaction Coordinate Forces (FEARCF): An application to ring puckering. *Mol. Phys.* **2009**, *107*, 1243−1250. (c) Naidoo, K. J. Multidimensional free energy volumes offer unique insights into reaction mechanisms, molecular conformation, and association. *Phys. Chem. Chem. Phys.* **2012**, *14* (25), 9026−9036.

(18) Barnett, C. B.; Naidoo, K. J. Ring Puckering: A Metric for Evaluating the Accuracy of AM1, PM3, PM3CARB-1, and SCC-DFTB carbohydrate QM/MM simulations. *J. Phys. Chem. B* **2010**, *114*, 17142−17154.

(19) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(20) Thiel, W. *MNDO97, version 5.0*; University of Zurich: Zurich, Switzerland, 1998.

(21) (a) Hill, A. D.; Reilly, P. J. Puckering coordinates of monocyclic rings by triangular decomposition. *J. Chem. Inf. Model.* **2007**, *47*, 1031−1035. (b) Khalili, P.; Barnett, C. B.; Naidoo, K. J. Interpreting medium ring canonical conformers by a triangular plane tessellation of the macrocycle. *J. Chem. Phys.* **2013**, *138* (18), 184110−1−184110−7.

(22) Taniguchi, N.; Honke, K.; Fukuda, M. *Handbook of Glycosyltransferase and Related Genes*; Springer: Tokyo, 2002.

(23) Harvey, S. C.; Prabhakaran, M. Ribose puckering: Structure, dynamics, energetics, and the pseudorotation cycle. *J. Am. Chem. Soc.* **1986**, *108*, 6128−6136.

(24) Ionescu, A. R.; Berces, A.; Zgierski, M. Z.; Whitfield, D. M.; Nukada, T. Conformational pathways of saturated six-membered rings. A static and dynamical density functional study. *J. Phys. Chem. A* **2005**, *109*, 8096−8105.

(25) Stoddart, J. F. *Stereochemistry of Carbohydrates*. John Wiley & Sons, Inc.: New York, 1971.

(26) Gronert, S.; Simpson, D. C.; Conner, K. M. A reevaluation of computed proton affinities for the common α-amino acids. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (11), 2116−2123.

(27) (a) Hunter, E. P.; Lias, S. G. *Nist Standard Reference Database Number 69*; Mallard, W. G., Linstrom, P. J., Eds. National Institute of Standards and Technology (http://webbook.nist.gov): Gaithersburg MD, 2008. (b) Linstrom, P.; Mallard, W. *NIST Chemistry WebBook*, http://webbook.nist.gov/chemistry; NIST Standard Reference Database Number 69: National Institute of Standards and Technology, Gaithersburg MD, 2003.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision C.01*; Gaussian, Inc.: Wallingford CT, 2009.

(29) (a) Moser, A.; Range, K.; York, D. M. Accurate proton affinity and gas-phase basicity values for molecules important in biocatalysis. *J. Phys. Chem. B* **2010**, *114* (43), 13911−13921. (b) Range, K.; Riccardi, D.; Cui, Q.; Elstner, M.; York, D. M. Benchmark calculations of proton affinities and gas-phase basicities of molecules important in the study of biological phosphoryl transfer. *Phys. Chem. Chem. Phys.* **2005**, *7* (16), 3070−3079.

(30) Chyall, L. J.; Squires, R. R. The proton affinity and absolute heat of formation of trifluoromethanol. *J. Phys. Chem.* **1996**, *100*, 16435−16440.

(31) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8* (17), 1985−1993.

(32) Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. Assessment of the performance of the M05-2X and M06-2X exchange-correlation functionals for noncovalent interactions in biomolecules. *J. Chem. Theory Comput.* **2008**, *4* (12), 1996−2000.

(33) Raju, R. K.; Ramraj, A.; Vincent, M. A.; Hillier, I. H.; Burton, N. A. Carbohydrate−protein recognition probed by density functional theory and ab initio calculations including dispersive interactions. *Phys. Chem. Chem. Phys.* **2008**, *10* (43), 6500−6508.

(34) Raju, R. K.; Ramraj, A.; Hillier, I. H.; Vincent, M. A.; Burton, N. A. Carbohydrate−aromatic π interactions: A test of density functionals and the DFT-D method. *Phys. Chem. Chem. Phys.* **2009**, *11* (18), 3411−3416.

(35) Tvaroška, I.; Kozmon, S.; Wimmerová, M.; Koča, J. Substrate-assisted catalytic mechanism of O-GlcNAc Transferase discovered by Quantum Mechanics/Molecular Mechanics investigation. *J. Am. Chem. Soc.* **2012**, *134* (37), 15563−15571.

(36) Rizzo, R. C.; Jorgensen, W. L. OPLS all-atom model for amines: Resolution of the amine hydration problem. *J. Am. Chem. Soc.* **1999**, *121*, 4827−4836.