JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

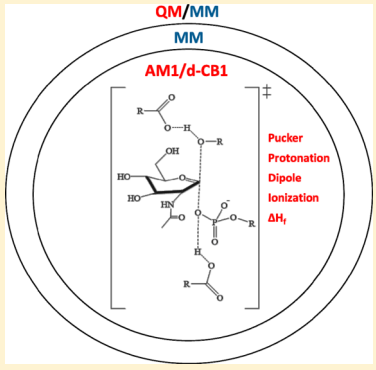
AM1/d-CB1: A Semiempirical Model for QM/MM Simulations of Chemical Glycobiology Systems

Krishna Govender,[†,‡] Jiali Gao,[§,∥] and Kevin J. Naidoo*[,†,‡]

[†]Scientific Computing Research Unit and [‡]Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa
[§]State Key Laboratory of Theoretical and Computational Chemistry, Jilin University, Changchun, Jilin Province 130012, China
[∥]Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455, United States

Supporting Information

ABSTRACT: A semiempirical method based on the AM1/d Hamiltonian is introduced to model chemical glycobiological systems. We included in the parameter training set glycans and the chemical environment often found about them in glycoenzymes. Starting with RM1 and AM1/d-PhoT models we optimized H, C, N, O, and P atomic parameters targeting the best performing molecular properties that contribute to enzyme catalyzed glycan reaction mechanisms. The training set comprising glycans, amino acids, phosphates and small organic model systems was used to derive parameters that reproduce experimental data or high-level density functional results for carbohydrate, phosphate and amino acid heats of formation, amino acid proton affinities, amino acid and monosaccharide dipole moments, amino acid ionization potentials, water-phosphate interaction energies, and carbohydrate ring pucker relaxation times. The result is the AM1/d-Chemical Biology 1 or AM1/d-CB1 model that is considerably more accurate than existing NDDO methods modeling carbohydrates and the amino acids often present in the catalytic domains of glycoenzymes as well as the binding sites of lectins. Moreover, AM1/d-CB1 computed proton affinities, dipole moments, ionization potentials and heats of formation for transition state puckered carbohydrate ring conformations, observed along glycoenzyme catalyzed reaction paths, are close to values computed using DFT M06-2X. AM1/d-CB1 provides a platform from which to accurately model reactions important in chemical glycobiology.



1. INTRODUCTION

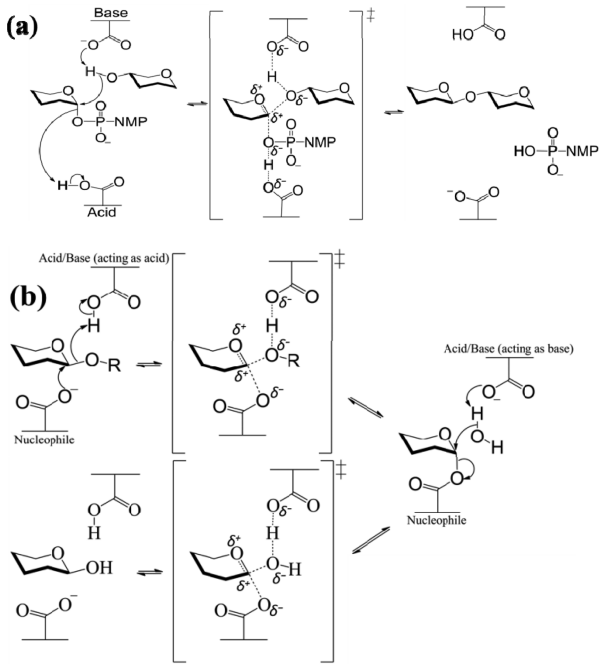
Carbohydrates present in either the five-membered (furanose) or the six-membered (pyranose) ring conformation play central biological and chemical roles in all living organisms. At the core of major biochemical cycles are carbohydrate processing enzymes which we collectively refer to as glycoenzymes. These are principally glycosidases (the enzymes responsible for the breakdown of di-, oligo-, and polysaccharides) and glycosyltransferases (the enzymes which covalently link monosaccharides to oligosaccharides, small molecules, lipids, or proteins).[1] Here the conformational changes for pyranose rings feature prominently about the transition state (TS). There are two possible stereochemical outcomes for reactions involving the glycosidic bond: inversion and retention of the anomeric configuration. Both the inversion and retaining mechanisms have in common the formation of a puckered oxocarbenium ion in the TS where the ring puckers into a conformation other than the standard $^{1}C_{4}$ or $^{4}C_{1}$ chair.

Stabilizing a positively charged oxocarbenium ion is central to the catalytic function of glycoenzymes where some enzymes appear to act primarily by directly stabilizing the oxocarbenium ion.[2] A second common feature of these enzymatic reactions are the acid/base roles played by carboxylic acid containing residues. In, for example, inverting glycosyltransferases the reaction occurs via a single-displacement mechanism wherein one protein residue containing a carboxylate acts as a general base and another carboxylic acid containing residue acts as a general acid (Scheme 1a).[3] Similar ring pucker and acid/base catalytic features are found in retaining glycosidases where the reaction proceeds via a double displacement mechanism (Scheme 1b).[4]

Furanose rings play a significant role in base excision repair (BER), which is the major system responsible for the expulsion of corrupted DNA bases and their repair.[5] The challenge for BER enzymes is to recognize and remove multiple types of DNA damage while avoiding reactions with the millions-fold excess of normal DNA.[2] One of the energetically most challenging chemical reactions is that of glycosidic bond hydrolysis in damaged pyrimidine nucleotides. A commonly occurring damage to a pyrimidine base in DNA is the spontaneous deamination of cytosine to generate uracil. Such events are unavoidable in the aqueous cellular environment, and are repaired by the highly conserved activity of enzyme uracil DNA glycosylase (UDG).[6] There are two possible TS extremes for the UDG catalyzed reaction; (i) the formation of a discrete oxocarbenium intermediate with an sp^2 hybridized anomeric carbon (Scheme 2, upper pathway), and (ii) the concerted association and dissociation of the nucleophile

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

Scheme 1. General Glycoenzymes Mechanisms for (a) Stereochemical Inversion of Glycosyltransferase and (b) a Retaining Glycosidase, Proceeding through a Transition State/Intermediate That Is Puckered Away from the $^4C_1$ Chair



(H_2O) and the leaving group (uracil) at the anomeric carbon (Scheme 2, lower pathway).[6]

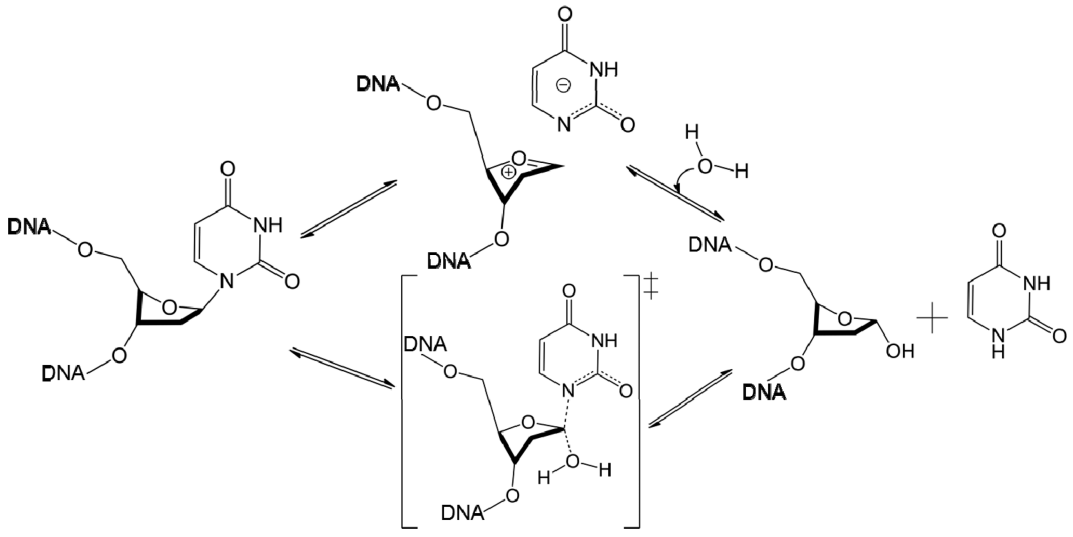
There have been numerous studies of carbohydrate processing enzymes, including X-ray crystallography[7] and kinetic isotope effects.[6] However, a detailed characterization of the transition states for these enzymatic reactions is still lacking,[8] and computational studies using combined QM/MM techniques provide a powerful tool to elucidate the reaction mechanisms. To achieve computational efficiency needed for sufficient conformational dynamic sampling, semiempirical (SE) methods are often used.[1b]

It is important to validate the computational accuracy of SE models in the description of the ring conformation puckering of pyranoid and furanoid carbohydrates. The conformational and electronic transitions of both furanose and pyranose monosaccharides, or the mechanistic details and TS have not been specifically addressed to correctly simulate reactions such as those shown in Schemes 1 and 2 using SE methods. In this work we develop a new SE model, AM1/d-Chemical Biology 1 or AM1/d-CB1 that takes into account the physical, biochemical and chemical features central to chemical glycobiology, which are carbohydrate ring puckering, acid/base or nucleophilic ability of amino acids and phosphate interactions. The method is compared to other well-known SE methods, such as AM1,[9] PM3,[10] PM3CARB-1,[11] PM3^{MS},[12] RM1,[13] and SCC-DFTB[14] to evaluate its accuracy.

2. PARAMETERIZATION STRATEGY

The theory supporting the development of AM1/d-CB1 and critical factors taken into account when conducting the parametrization are detailed below. The properties used for parameter optimization in this work are ring relaxation rates, proton affinities, dipole moments, ionization potentials, interaction energies and heats of formation. We employ a diverse and large training set that aims to improve on current NDDO model deficiencies in the simulation of chemical glycobiological systems. In particular we aimed to improve the modeling of carbohydrate ring flexibilities, carbohydrate ring puckers that occur along commonly observed glycoenzymes catalyzed reaction pathways as well as amino acids that bind the glycans inside the enzyme catalytic domain and protein binding pockets. Since some properties are more significant than others, in the context of chemical glycobiology, a differentiated weighting scheme was used (Table 1). We term the selective molecular property optimization that comprises glycans, amino acids and phosphates coupled with an uneven property optimization weighting to achieve overall improvement for

Scheme 2. Possible Reaction Mechanisms for Uracil DNA Glycosylase (UDG)[a]



[a]Top: Limiting stepwise mechanism involving the formation of a discrete oxocarbenium ion intermediate that is subsequently trapped by the water nucleophile. Bottom: An associative S_N2 mechanism involving concerted addition of the water nucleophile and expulsion of the uracil leaving group. The transition state in both cases is significantly puckered.

dx.doi.org/10.1021/ct500372s | J. Chem. Theory Comput. 2014, 10, 4694−4707

Table 1. Weighting Factors Used for Reference Data

reference data	weight	unit
proton affinities	25.0	mol·kcal^{-1}
dipole moments	25.0	debye^{-1}
interaction energies	20.0	mol·kcal^{-1}
ionization potential	15.0	eV^{-1}
pucker relaxation rate	15.0	ps^{-1}
ΔH_f	1.0	mol·kcal^{-1}

chemical glycobiological modeling the *variable property optimization* (VPO) parameter approach. We show here and in a later report[15] that this strategy produces a parameter set (AM1/d-CB1) that best models chemical glycobiological events.

2.1. Semiempirical Methods. The neglect of diatomic differential overlap (NDDO) approximation is the most widely used semiempirical method, in which three and four center electronic repulsion integrals are neglected. Hamiltonians based on the NDDO formalism and still widely used today include, MNDO,[16] AM1,[9] PM3, PM6, and RM1.[10a] The only differences between these Hamiltonians stem from the way in which the core−core repulsion interactions are treated. In MNDO the repulsion between two nuclear cores (A and B) is described by,

$$E_{AB}^{\mathrm{MNDO}} = Z_A Z_B \langle s_A s_A | s_B s_B \rangle [1 + \mathrm{e}^{-\alpha_A R_{AB}} + \mathrm{e}^{-\alpha_B R_{AB}}] \quad (1)$$

where Z_A and Z_B are the effective core charges, $\langle s_A s_A | s_B s_B \rangle$ is a Coulomb repulsion integral between two s-orbitals centered on atoms A and B, α is an adjustable element specific parameter that decreases the screening of the nuclear charge by the electrons at small interatomic distances, and R_{AB} is the interatomic distance between atoms A and B.

An additional set of Gaussian functions were introduced in methods such as AM1 and PM3 to provide a weak attractive force:

$$E_{AB}^{\mathrm{AM1/PM3}} = E_{AB}^{\mathrm{MNDO}} + \frac{Z_A Z_B}{R_{AB}} \left[\sum_k a_{Ak}\, \mathrm{e}^{-b_{Ak}(R_{AB}-c_{Ak})^2} + \sum_k a_{Bk}\, \mathrm{e}^{-b_{Bk}(R_{AB}-c_{Bk})^2} \right] \quad (2)$$

where a, b, and c are adjustable spherical Gaussian function parameters, for which the AM1 Hamiltonian has up to four per atom, whereas for PM3 there is a maximum of only two Gaussian parameters per atom.

In the parametrization of AM1 only very few molecules was used. This was a natural constraint imposed by the software and equipment available at the time.[17] As such, apart from the difference in the number of Gaussian parameters, PM3 had a mathematical philosophy for the parametrization procedure that involved the derivation and implementation of formulas to arrive at a suitable error function with respect to the parameters.[10] The error function used during the parametrization was given as,

$$S = \sum_i (x_i^{\mathrm{calc}} - x_i^{\mathrm{ref}})^2 \quad (3)$$

where S is defined as the sum of squares of differences between calculated and reference values. This function is considered *optimized* when, for a set of parameters, the value of S is a minimum.

With these modifications a total of 12 elements (H, C, N, O, F, Al, Si, P, S, Cl, Br, and I) were optimized simultaneously during the original PM3 parametrization.[10a] Various new parameters based on the original AM1 and PM3 formalism have been developed over the years, these include PM3CARB-1,[11] PM3$^{\mathrm{MS}}$,[12] and RM1[13] (all of which have been utilized in this work). PM3CARB-1 was developed to accurately model a set of small carbohydrate analogues, PM3$^{\mathrm{MS}}$ was designed for studies of saccharides, and RM1 was a reparameterization of AM1 for a training set representative of organic and biochemical containing atoms (C, H, N, O, P, S, F, Cl, Br, and I).

For hypervalent species such as organic phosphate and sulfate, it becomes necessary to include *d*-orbitals explicitly in the semiempirical formalism.[18] A method in which the *d*-orbital extension has been applied is MNDO/d[18b,c] and this has shown to perform reliably in the study of transphosphorylation under basic conditions.[19] However, MNDO/d was found to perform poorly for transphosphorylation reactions in neutral and acidic conditions where hydrogen bonding and proton transfer play major roles.[20] Nam et al.[20] developed a new model, called AM1/d-PhoT, in which the phosphate element was treated by the MNDO/d formalism, whereas the first row elements are represented by AM1. With this method the authors kept the core−core terms for hydrogen bonding, but turned these interactions off for phosphorus bonding. The resulting method had a core−core repulsion term defined as,

$$E_{AB}^{\mathrm{AM1/d\text{-}PhoT}} = E_{AB}^{\mathrm{MNDO}} + \frac{Z_A Z_B}{R_{AB}} \times G_{\mathrm{scale}}^A G_{\mathrm{scale}}^B \left[\sum_k a_{Ak}\, \mathrm{e}^{-b_{Ak}(R_{AB}-c_{Ak})^2} + \sum_k a_{Bk}\, \mathrm{e}^{-b_{Bk}(R_{AB}-c_{Bk})^2} \right] \quad (4)$$

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

AM1/d-PhoT can provide significant improvement over MNDO/d, AM1, and PM3 for transphosphorylation reactions.[20] It has also been successfully applied to the alkaline hydrolysis of a phosphodiester (methyl *p*-nitrophenyl phosphate) in the active site of *Escherichia coli* alkaline phosphatase.[21] In this study, we employ the same core−core interactions as that used in the AM1/d-PhoT model, and reparameterize the method so that it is applicable to systems which are relevant to chemical glycobiology, thereby giving rise to a new method (AM1/d-CB1).

2.2. Key Molecular Properties to Consider in Chemical Glycobiology. To accurately model the glycans in cellular systems the computation of specific properties important in chemical glycobiology must be used as metrics in the parameter optimization process. The following summarizes chemical features and molecular characteristics that we have taken into account during the parametrization process.

Ring Flexibility and Pucker. Reactions in chemical glycobiology involve the presence of either a 5- or 6-membered carbohydrate ring. Carbohydrate rings are conformationally flexible[1b] although they are not as flexible as cycloalkanes e.g., cyclohexane.[8a] It is this flexibility that leads to an exploration of ring pucker conformational space during the progression of

dx.doi.org/10.1021/ct500372s | J. Chem. Theory Comput. 2014, 10, 4694−4707

Table 2. Optimized AM1/d-CB1 Parameters for Hydrogen, Carbon, Nitrogen, Oxygen, and Phosphorus

params.	hydrogen	carbon	nitrogen	oxygen	phosphorus
U_{ss}	−11.960909	−50.301531	−69.842739	−96.951432	−45.405057
U_{pp}		−38.793389	−55.880457	−77.905354	−41.533302
U_{dd}					−26.708704
ζ_s	1.052925	1.822969	2.351342	3.128851	2.058999
ζ_p		1.801099	2.033642	2.585827	2.214770
ζ_d					0.816679
β_s	−5.792945	−15.298988	−20.881617	−29.843262	−11.435826
β_p		−8.001275	−16.165663	−29.460458	−10.694210
β_d					−2.580718
α	3.026944	2.819744	3.185782	4.207192	2.050087
G_{ss}	13.808409	12.967197	11.847719	13.865036	14.263381
G_{pp}		11.063079	13.339745	14.481686	12.379996
G_{sp}		11.231690	12.735692	15.108816	5.769559
G_{p2}		9.872842	11.665818	12.449295	9.531268
H_{sp}		1.502380	4.683588	3.915720	1.272332
ζ_{sn}					2.069613
ζ_{pn}					1.485597
ζ_{dn}					1.139956
ρ_{core}					1.085029
G_{scale}	1.000000	1.000000	1.000000	1.000000	0.388294
FN_{11}	0.101830	0.075134	0.057001	0.228672	−0.334497
FN_{21}	5.891927	5.898126	4.339867	5.225437	3.202253
FN_{31}	1.175830	1.026976	1.283016	0.914621	1.020740
FN_{12}	0.065851	0.012140	0.023972	0.058956	−0.024098
FN_{22}	6.368911	6.956238	4.760398	7.537833	1.758030
FN_{32}	1.941724	1.664940	2.011604	1.516886	2.731363
FN_{13}	−0.034689	0.036407	−0.023463		−0.035212
FN_{23}	2.856686	6.263881	2.028720		4.902280
FN_{33}	1.625337	1.658710	1.961806		2.045419
FN_{14}		−0.002767			
FN_{24}		9.001121			
FN_{34}		2.817645			

hydrolysis, glycosylation and phosphorylation reactions. A ring that is too *stiff* or too *floppy* will not adapt to the important conformers needed in glycobiological reactions (see Schemes 1 and 2). We therefore monitor the effect of the new parameters on the 5- and 6-membered carbohydrate ring relaxation times. Further since ring puckering is a major driving force for chemical glycobiological reactions we place high priority on the proton affinities and electrostatic character (dipole moments) of the TS and other rings that are puckered away from the $^4C_1$ or $^1C_4$ chair conformers.

Bond Polarization. During a hydrolysis or glycosylation reaction, the carbohydrate is not only puckered away from its equilibrium (e.g., pyranose chair (C)) ring conformer, but localized partial charges evolve on the oxygen and carbon atoms. This is the positively charged oxocarbenium ion. Moreover the oxocarbenium ion positions, nucleophilic residues and leaving groups in the catalytic site. To improve the accuracy of modeling the oxocarbenium ion and the nucleophilic residues bond polarity has to be computed as closely to an ab inito result as possible. This is done by better calculating the molecular dipole moments and ionization potentials of pyranose half chairs (H), envelopes (E), boats (B), and skew (S) ring conformers as well as furanose twist (T) and envelop (E) ring conformers. To increase existing NDDO modeling of the nucleophilic interactions we follow more closely their molecular dipole moments and ionization potentials.

Amino Acid Contributions to Glycan Reactivity. Scheme 1 and 2 show that in glycosyltransferase and glycosidase, or glycosylase proton transfer plays an essential role in glycosyl transfer. It is important to accurately model proton affinities of acid and basic groups involved in these reactions using a QM/MM method. The generic mechanism, in for example glycosyltransferases, following nucleophilic attack is to have one residue act as acid catalyst in promoting the departure of the leaving group while the other acts as a base catalyst to abstract a proton from the acceptor substrate. With this in mind we track the proton affinities of amino acid residues, common to catalytic domains, as an important property that AM1/CB1 is to model as accurately as possible.

3. METHODS

Here we describe the methods used during the development of the AM1/d-CB1 model as given in Table 2. The original set of RM1 and AM1/d-PhoT parameters used to launch the parametrization is given in Table S1 (of the Supporting Information). We lay out the quantum chemical database used as the reference data to fit the SE parameters and provide the details of the parametrization itself.

Structures for all DFT simulations were obtained from previous work by Barnett and Naidoo[1b,22] (β-D-glucopyranose), Jalbout et al.[23] (β-D-ribofuranose), the QCRNA database[24] (phosphoric systems), and from the training set used in the development of the SE PM6 Hamiltonian (amino acids).[25] The

dx.doi.org/10.1021/ct500372s | J. Chem. Theory Comput. 2014, 10, 4694−4707

α-D conformations of the carbohydrates were constructed from their DFT optimized β-D counterparts prior to undergoing an optimization themselves. All DFT based simulations were performed using the M06-2X functional[26] along with the 6-311++G(3df,2p) basis set. The functional was chosen because it has been shown to produce good thermochemistry, kinetics, and noncovalent interactions.[26] All structures were optimized with the M06-2X functional and the 6-311++G(3df,2p) basis set. Frequency calculations were conducted on all optimized structures to verify the stationery points and to obtain relevant thermochemical data. The calculations described above have all been run with the Gaussian 09 software package.[27]

3.1. AM1/d-CB1 Parametrization Strategy. The parametrization strategy adopted in this work differs from those of previous methods.[20,28] Instead of using a specific reaction parametrization (SRP) strategy, we parametrize the semiempirical model with the aim of studying a specific class of molecules (glycans) as well as the environment (amino acids and/or amino acid base pairs) within which they are known to function across the chemical glycobiological landscape. The chemical training set used in this work consists of (i) carbohydrate based conformations (α- and β-conformers of glucopyranose and ribofuranose) that are the basic building blocks for glycobiological systems, (ii) carbohydrate-phosphate conformers (α- and β-conformers of glucose-6-phosphate and ribose-5-phosphate) important in biologically significant systems such as DNA and RNA, (iii) a number of amino acids (aspartic acid, asparagine, glutamic acid, glutamine, histidine, argenine, phenyalanine, tyrosine, and tryptophan) participating in catalytic reactions, and (iv) various other organic molecules (1,2-ethanediol, methoxymethanol, H_2O, CH_3OH, C_2H_5OH, C_6H_5OH, CH_3CO_2OH, CH_3OCH_3, P-$(CH_3)_3$, $(CH_3)_3PO$, and H_3PO_4) that are relevant to chemical glycobiology. The VPO parametrization approach aims to get optimal performance of a particular property for a molecular class based on the role that the molecular system plays within the glycan reaction scheme. We seek to optimize root mean square deviation (RMSD) relaxation, dipole moments, proton affinities, and heats of formation to model carbohydrate out-of-chair ring pucker, oxocarbenium ion formation, amino acid acid/base and nucleophilic catalytic activity, as noted in Schemes 1 and 2 and outlined section 2.2 above. The objective of a VPO approach is therefore an AM1/d-CB1 model that can be used to simulate carbohydrate chemical reactivity in chemical glycobiology.

Of particular importance in glycobiology is the dynamic behavior of carbohydrate ring conformer pucker. This is because the ease and specificity of furanose and pyranose ring pucker away from the equilibrium ring conformers is central to the formation of transition states found along the reaction paths catalyzed by phosphorylases, ribosyl transferases, hydrolases, and dehalogenases.[1b] SCC-DFTB was used as a reference data set because ab initio molecular orbital and DFT dynamics are computationally intensive and not feasible on current hardware platforms. Besides this, SCC-DFTB has been shown to provide an accurate description of carbohydrate ring pucker.[1b] While β-D-glucopyranose was used to measure pucker dynamics for six membered rings the extensive intramolecular hydrogen bonding occurring between the secondary and primary hydroxyls in β-D-ribofuranose (Figure S1 in Supporting Information) impedes the observation of gas phase ring puckering for the five membered ring. Consequently, tetrahydrofuran was used as a model for the five membered ring to improve on the excessive ring flexibility of the AM1 model.[1b] Several molecular dynamics simulations were performed using SCC-DFTB (in the gas phase) on tetrahydrofuran and β-D-glucopyranose.[14] Langevin dynamics at 298.15 K and group based cutoffs of 16.0, 14.0, and 12.0 Å were used. Carbohydrate ring flexibility can be measured using triangular tessellation of the ring.[29] We use the same definition for the central reference planes for five and six membered rings as previously reported and shown in Figure 1.[1b]

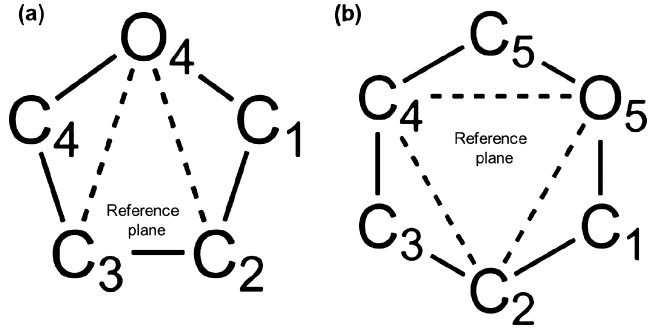


Figure 1. Reference planes for (a) tetrahydrofuran/β-D-ribofuranose and (b) β-D-glucopyranose, as they feature in the triangular tessellation definition of ring pucker.

Relaxation times for the atoms in the ring (C_1 and C_4 for the five membered ring and C_1, C_3, and C_5 for the six membered ring) were computed from their relative motion to the reference plane. The average relaxation times were then computed for the pyranose and furanose rings and included in the parameter-training database. We determined a number time correlation functions (TCFs) for the ring relaxation times of tetrahydrofuran and β-D-glucopyranose. The first 10 ps of the TCFs were fitted using the exponential function,

$$f(x) = a\, e^{(-x/\tau)} + b \tag{5}$$

where τ is the relaxation time that is used during the AM1/d-CB1 parametrization.

3.2. Parameter Optimization. All properties where computed using a modified version of MOPAC along with the CHARMM35/MNDO97[30] interface. SE computations using MOPAC were run as single point calculations on the DFT optimized geometries to preserve conformations such as the half-chair of β-D-glucopyranose in so doing preventing a conversion to the physically incorrect boat conformation upon optimization in vacuum. The error function

$$S = \sum_{i}^{mol} \sum_{j}^{\mathrm{prop}(i)} w_{ij}\left(x_{ij}^{\mathrm{calc}} - x_{ij}^{\mathrm{ref}}\right)^2 \tag{6}$$

was used where the first summation with index i runs over all molecules, and the second summation with index j runs over the properties associated with the ith molecule. w_{ij} is a weighting factor, x_{ij}^{calc} is the calculated molecular property j for molecule i using the generated set of parameters, and x_{ij}^{ref} is the corresponding target value (either M06-2X/6-311++G(3df,2p), SCC-DFTB or experimental data).

Although we do not include molecular structure properties, such as bond lengths, bond angles, and dihedrals, in the parametrization, we do check for the geometric performance and find that AM1/d-CB1 to achieve closer correlation to DFT results than other SE methods.[15] The weights used in the

Table 3. Ring Relaxation Times for Molecules Used in Parameterization (Picoseconds)[a]

	SCC-DFTB[b]	PM3CARB-1	PM3[MS]	AM1/d-PhoT	AM1/d-CB1
			tetrahydrofuran		
τ	0.47214	0.10223	0.10090	0.14734	0.22579
			β-D-glucopyranose		
τ	0.17384	0.17237	NONE	2.10800	0.14219

[a]Correlation time could not be established since exponential fit was not possible with data generated from the dynamics run. τ corresponds to relaxation time for carbohydrate ring described in section 3.1. NONE implies that a relaxation time could not be obtained within the simulation time frame. [b]Theoretical values obtained with gas-phase SCC-DFTB[14] MD simulations.

current work for various properties (Table 1) have been selected (i) to render properties unit less and (ii) to increase the significance of certain properties, relevant to the chemical glycobiology, over others during function evaluation.

The error function given above was used in a genetic algorithm (GA) optimization based on the work of Goldberg.[31] The population in each iteration is called a generation. Within each generation the fitness of every individual in the population is evaluated, were the fitness in the current work is based on eq 6. The more fit individuals are stochastically selected from the current population, and each individual is modified (mutated) to form a new generation. The new generation of candidate solutions is then used in the next iteration of the GA. Termination of a GA is typically achieved when either the maximum number of generations has been reached or a satisfactory fitness level has been acquired for the population. This GA was chosen since it has shown significant promise when establishing SE based parameter sets,[18a,20,32] and can sample a wide range of parameter space.

The parametrization procedure used here comprised two steps. First, we obtained a parameter set for H, C, O, and P, which can be used to accurately model the ring puckering of the carbohydrates and produce reliable results for the phosphates. Next, we refine the parameters of N and P while keeping those of H, C, and O fixed to provide a good energetic description for the amino acids and phosphates. Following this, a final refinement of H, C, N, O, and P is carried out.

Having examined a range of semiempirical methods (AM1, PM3, PM3CARB-1, PM3[MS], RM1, and AM1/d-PhoT), we chose to use RM1 and AM1/d-PhoT parameters as the initial guess in our optimization procedure. RM1 generally produced good results for the organic systems and reasonable results for phosphorus, whereas AM1/d-PhoT produced good results for the phosphates. AM1/d-PhoT accounts for the important *d*-orbitals present on phosphorus atom by applying an *spd* basis onto the atom. As a starting point, we optimized H, C, O, and P parameters first while the parameters for nitrogen were fixed at the original RM1 values. The GA involved using a population size of 100, mutation rate of 0.50, and elitism rate of 0.05. The parameter sets with the lowest fitness were then evaluated for a total of 25 generations (or iterations). During the optimization, parameters were allowed to vary within 5−6% of their initial values. This process was repeated, adjusting the parameter bounds, until the mean unsigned errors (MUEs) for each property were minimized and consistent with the VPO strategy.

The parameters optimized for H, C, O and P performed well for carbohydrate systems, but there were greater errors for phosphates and amino acids. To correct this, a second parametrization exercise was undertaken where the parameters for H, C, and O were fixed at the values generated above and only the parameters for N and P were allowed to vary. In this case, the GA runs had a population size of 512, mutation rate of 0.50, and elitism of 0.05 for a total of 25 generations. Parameters for phosphorus were allowed to vary within 2−5% of the parameters generated above, whereas the parameters for nitrogen were allowed to vary within 5−6% of the original RM1 parameters. Once again, this process was repeated, adjusting the parameter bounds. A final optimization of only the one-center two-electron repulsion integrals (G_{ss}, G_{pp}, G_{sp}, G_{p2}), one-center two-electron exchange integral (H_{sp}), and Gaussian terms (FN) was conducted for all atoms. The H, C, and O parameters were allowed to vary within 1% of the values generated above, whereas the N and P parameters were allowed to vary within 3% of the values generated thus far. The GA in this case consisted of a population size of 100, mutation rate of 0.50, and elitism of 0.05. The process was repeated, adjusting the parameter bounds, until a set of results consistent with our VPO strategy was obtained.

4. RESULTS AND DISCUSSION

We now analyze properties computed for all molecular classes using AM1/d-CB1 optimized parameters (Table 2). The AM1/d-CB1 results are compared with experimental data or high-level DFT simulations (M06-2X/6-311+G(3df,2p)). We pay special attention to the performance of properties of selective molecular classes, for example the proton affinities of amino acid residues, and dipole moments of nonchair puckered pyranose rings. These molecular properties are computed for a range of SE methods (AM1, PM3, PM3CARB-1, PM3[MS], RM1, and AM1/d-PhoT) to measure the success of the AM1/d-CB1 parametrization. Unless otherwise stated, all SE simulations are conducted as single point calculations on the DFT optimized geometries.

4.1. Ring Relaxation Times. The ring relaxation time is a measure of the dynamic performance of the ring indicating the ease with which the carbohydrate may access a transition state conformation during the reaction (see Schemes 1 and 2). Previously, we had shown that the generalized NDDO methods aimed at modeling organic systems (AM1 and PM3) produce carbohydrate rings that have very low ring conformational free energy barriers resulting in monosaccharide rings that are too flexible compared with the Hartree−Fock level of theory or the SE SCC-DFTB method.[1b] We concluded there that the SCC-DFTB method presents the most reliable models for carbohydrate ring pucker of all SE methods. In this context and in the absence of ab initio free energy pucker conformational volumes/dynamics, we compared the minimum pathway for the pyranose ring from C to H or E ring conformations for the NDDO methods to the SCC-DFTB method. Overall the available NDDO methods' minimum free energy puckering paths do not map well to the SCC-DFTB minimum free energy puckering paths. A case in point is the PM3CARB-1 method that had been specifically parametrized for carbohydrates,

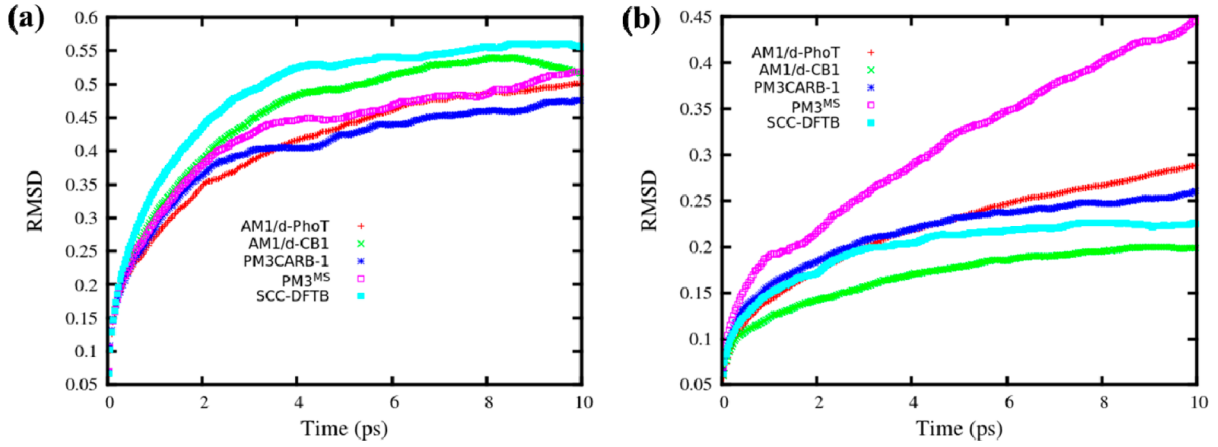


Figure 2. Average RMSD (degrees) for (a) tetrahydrofuran and (b) β-D-glucopyranose.

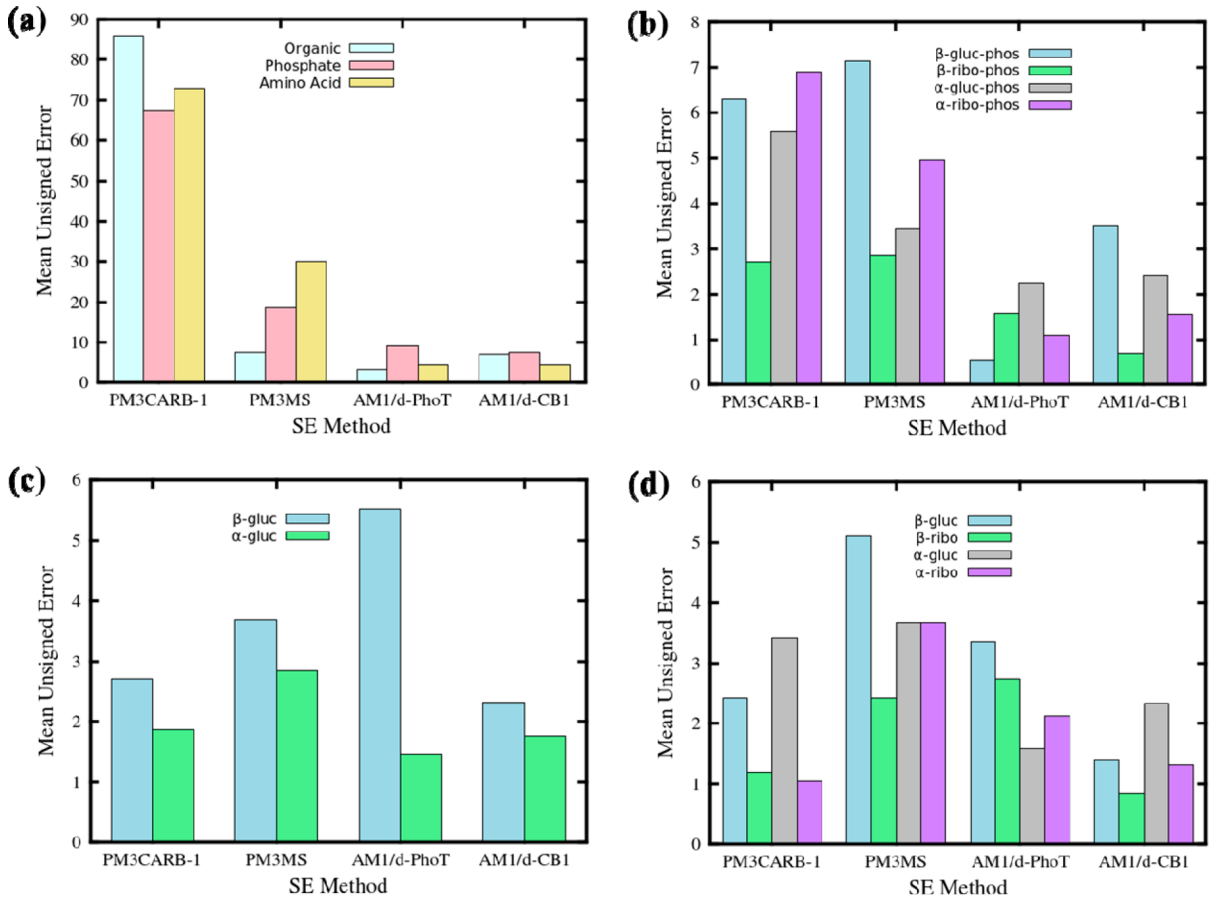


Figure 3. Mean unsigned errors (kcal/mol) for gas phase proton affinities of (a) organic molecules, phosphates and amino acids, (b) carbohydrate chair phosphorylated conformers, (c) carbohydrate chair conformers of glucopyranose and (d) puckered carbohydrate transition state conformers of glucopyranose and ribofuranose.

performed equally poorly for furanose rings and marginally better for pyranose rings.

The carbohydrate ring relaxation times of the 5- and 6-membered sugar rings were included in the AM1/d-CB1 parameter optimization. The two sugar rings used for the parametrization were tetrahydrofuran (5-membered ring) and β-D-glucopyranose (6-membered ring). The relaxation times were obtained by fitting of the TCFs using eq 5. In the absence of long time ab initio simulations, we used TCFs obtained from SCC-DFTB simulations as reference. The relaxation times for SCC-DFTB, the specially designed NDDO carbohydrate methods PM3CARB-1 and PM3^{MS}, AM1/d-PhoT, and AM1/d-CB1 are listed in Table 3 while AM1, PM3, and RM1 data are given in Table S2 (Supporting Information).

In addition to the ring relaxation times, we calculated the root-mean-square deviations (RMSDs) of atoms that lie adjacent to the reference plane (Figure 1). This gives an indication of the time it takes for a carbohydrate system to equilibrate. Average RMSDs for SCC-DFTB, PM3CARB-1, PM3^{MS}, AM1/d-PhoT, and AM1/d-CB1 are provided in Figure

dx.doi.org/10.1021/ct500372s | J. Chem. Theory Comput. 2014, 10, 4694−4707

2. Individual atomistic RMSDs are shown in Figures S2−S3 (Supporting Information).

For tetrahydrofuran, all SE methods have much shorter relaxation times (∼0.1 ps) than SCC-DFTB (Table 3); however, AM1/d-CB1's relaxation time (0.20484 ps) is closest to SCC-DFTB's ring relaxation time (0.47214 ps). The tetrahydrofuran AM1/d-CB1 model takes longer (6 ps) to equilibrate than the SCC-DFTB does (4.5 ps). The other methods do not reach equilibration within 10 ps. This is an indication that methods such as AM1/d-PhoT, PM3CARB-1, and $PM3^{MS}$ have 5-membered rings that are too flexible.

For β-D-glucopyranose, both AM1/d-CB1 and PM3CARB-1 relaxation times (0.17237 and 0.14219 ps, respectively) correlate perfectly with the dynamics of the SCC-DFTB ring model (0.17384 ps). However, AM1/d-CB1 equilibrates within 6.5 ps that is 2 ps is longer than that of SCC-DFTB (4.5 ps) indicating that the former produces a more flexible six membered ring. AM1/d-PhoT and $PM3^{MS}$ six membered ring dynamics produces even more flexible ring pucker models than AM1/d-CB1. We were unable to compute a ring relaxation time for $PM3^{MS}$, since the data obtained from the simulation was too unstable to fit. The RMSD plots (Figure 2b) confirm the poor behavior of both AM1/d-PhoT and $PM3^{MS}$ with a continually increasing average RMSD resulting in models that are not able to equilibrate within 10 ps.

4.2. Gas Phase Proton Affinities. The protonation states of biological systems play an important role in both structure and reactivity. There has been considerable effort devoted to the prediction of pK_a values with quantum chemistry.[33] However, this is an area that remains challenging due to the small differences in free energy that give rise to pK_a shifts (1 pK_a unit = 1.364 kcal/mol at 298.15 K). At the same time, reliable prediction of pK_a shifts using known experimental data may not always be possible since the determination of experimental pK_a values of an appropriate reference state might not be available.[33g] Here, we include both absolute (experimentally available data), and relative (DFT results) proton affinities (PAs). The molecules used in the training set were grouped into molecular subclasses. A summary of these subclasses is provided in Table S3 of the Supporting Information. Table S4 (Supporting Information) provides a comparison of calculated PAs and experimental data. For the reference proton affinities, we make use of the results from M06-2X calculations, since it has been shown that this functional yields excellent PA results in comparison with experiments.[26]

We pointed out earlier that acid and base residues in the catalytic domain of glycoenzymes are key to the success of glycan hydrolysis and glycosylation reactions. Therefore, the parameters that produce accurate amino acid PAs is a priority in our VPO parametrization strategy. PAs of amino acids present in the training set correspond to the enthalpy change from a neutral amino acid species to an N-protonated species (+1 charge). Data for this protonation state was acquired from both high-level calculations[8a] as well as experiment.[34] Single point SE calculations were done on the G3MP2 optimized geometries. AM1/d-CB1 and AM1/d-PhoT give the best performance for amino acids with MUEs of 4.3 and 4.4 kcal/mol, respectively (Table S4, Supporting Information). Table S4 shows that for H_2O AM1/d-PhoT produces the smallest error of 4.7 kcal/mol and $PM3^{MS}$ has an error of 7.2 kcal/mol, while all other SE methods have errors that are larger than 9 kcal/mol. For methanol, it is AM1, PM3, and AM1/d-PhoT that have the smallest errors (2.2, −1.8, and 2.0 kcal/mol, respectively). Figure 3 illustrates the MUE of proton affinities (PAs) for the subclasses of molecules used during parametrization. The errors of AM1/d-CB1 are shown in comparison for SE methods that have been used for carbohydrate modeling (PM3CARB-1, $PM3^{MS}$) or may be used in simulations of chemical glycobiology, as shown in Schemes 1 and 2 (AM1/d-CB1). Further, since the AM1/d-PhoT Hamiltonian is foundational to the development of AM1/d-CB1, these results were of relevance in gauging the convergence toward the optimal AM1/d-CB1 parameter set. AM1/d-PhoT produces the smallest error for the organic molecules (Figure 3a) with an error of 3.3 kcal/mol (Table S4, Supporting Information). In the case of phosphates, AM1/d-PhoT and AM1/d-CB1 give the smallest errors (9.0 and 7.5 kcal/mol, respectively) underscoring the importance of including *d*-orbitals on phosphorus. The molecule contributing the largest error for the phosphates is $(OCH_3)_2(OH)PO$ with an error of 42.5 and 30.4 kcal/mol for AM1/d-PhoT and AM1/d-CB1, respectively.

AM1/d-PhoT and AM1/d-CB1 yield the smallest errors for carbohydrate-phosphate PAs (Figure 3b). A more detailed look at the individual systems reveals that the lowest MUE for the β-D-glucopyranose-phosphate and α-D-ribofuranose-phosphates is given by AM1/d-PhoT with values of 0.54 and 1.09 kcal/mol, respectively (Table S5, Supporting Information). For β-D-ribofuranose-phosphates AM1/d-CB1 gives the smallest errors with a value of 0.69 kcal/mol (Table S5, Supporting Information). The MUE of α-D-glucopyranose-phosphate is similar for AM1/d-PhoT and AM1/d-CB1 with values of 2.26 and 2.41 kcal/mol, respectively. Methods that have previously been parametrized for carbohydrates (PM3 and PM3CARB-1) appear less suitable for modeling of a proton acceptance by a phosphate, which is an important facet for chemical glycobiological reactions (Scheme 1).

The protonation of carbohydrates is important in glycan reactions at the anomeric carbon (for example reactions shown in Scheme 1). PAs for minimum energy conformers ($^4C_1$ and $^1C_4$) of glucopyranose for AM1/d-CB1 (Figure 3c) surpasses all other methods. For the α-anomers the method gives an error (1.76 kcal/mol) that is slightly higher than that of AM1/d-PhoT (1.47 kcal/mol).

Computing the PAs of nonchair conformers that are often present in transition state configurations is a priority molecular property in this VPO strategy. To maintain the structural integrity of the carbohydrate conformers, only single point calculations were performed. AM1/d-CB1 models of β-D-glucopyranose and β-D-ribofuranose give PAs of 1.41 and 0.85 kcal/mol, respectively that are better than other NDDO methods (Table S6, Supporting Information). In the remaining cases, AM1/d-CB1 is second only to AM1/d-PhoT (1.59 kcal/mol) and PM3CARB-1 (1.05 kcal/mol) for α-D-glucopyranose (2.34 kcal/mol) and α-D-ribofuranose (1.31 kcal/mol) conformers, respectively (Table S7, Supporting Information).

4.3. Dipole Moments. During the parametrization of AM1/d-CB1, the DFT dipole moments were used as reference data. The molecular electrostatic potential correlates strongly with dipole moment as a result we placed high value to the accuracy of this property (see Table 1). Therefore, overall for each molecular subclass, the errors for AM1/d-CB1 are small although not the smallest in every case.

The amino acid residue dipole moments were a priority property in the VPO strategy as an accurate representation of the electrostatic potential within the catalytic domain of the

dx.doi.org/10.1021/ct500372s | J. Chem. Theory Comput. 2014, 10, 4694−4707

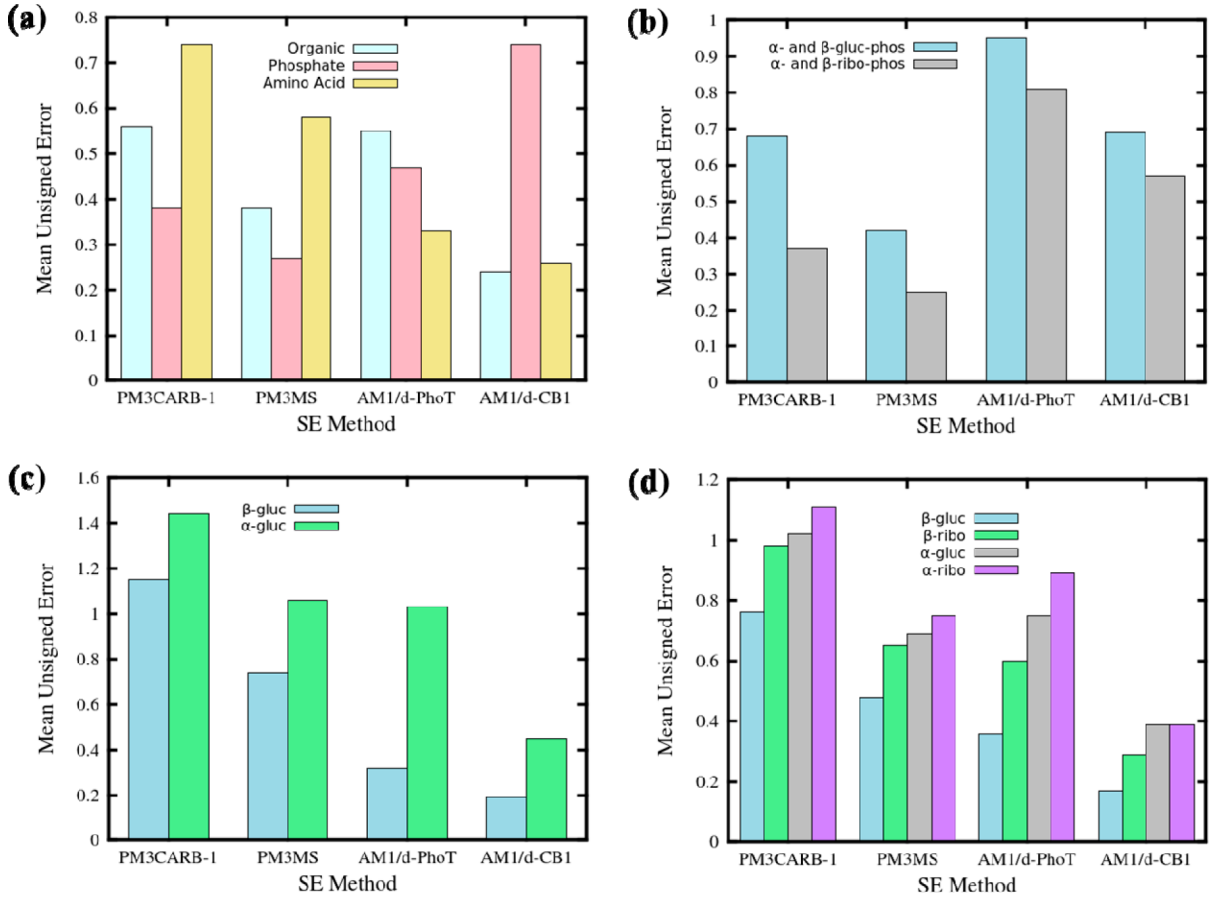


Figure 4. Mean unsigned errors (debye) of dipole moments for (a) organic molecules, phosphates, and amino acids, (b) carbohydrate chair phosphorylated conformers, (c) carbohydrate chair conformers of glucopyranose, and (d) no equilibrium puckered carbohydrate ring conformers of glucopyranose and ribofuranose.

glycoenzymes as well as within the protein binding site of carbohydrate binding proteins are essential.

Therefore, AM1/d-CB1 gives the best dipole moments for amino acids as well as organic molecules (Figure 4a) with MUEs of 0.26 and 0.24 debye, respectively (Table S8, Supporting Information). $PM3^{MS}$ performance (MUE of 0.38 debye) is closest to AM1/d-CB1 (Table S8, Supporting Information) for the organics. AM1/d-PhoT and PM3CARB-1 gives larger errors for organics with MUEs of 0.55 and 0.56 debye, respectively. In the case of the amino acids, it is AM1/d-PhoT that yields an error close to AM1/d-CB1 (0.33 debye). The methods parametrized for carbohydrates (PM3CARB-1 and $PM3^{MS}$) by comparison perform less well for amino acids giving MUEs of 0.74 and 0.58 debye, respectively (Table S8, Supporting Information).

Interestingly, the $PM3^{MS}$ and PM3CARB-1 methods that do not incorporate *d*-orbital character into phosphorus, give smaller dipole moment errors for the phosphate molecules than do the AM1/d-PhoT and AM1/d-CB1 methods that do include *d*-orbital character into phosphorus. While the error for AM1/d-CB1 is not the lowest for phosphates the greater MUE (0.74 debye) is due to two species (Table S8, Supporting Information) that are not significant in chemical glycobiology ($P(CH_3)_3$ and $(CH_3)_3PO$) removing these two from the error calculation produces a much smaller MUE for AM1/d-CB1 (0.15 debye).

The MUEs for the carbohydrate−phosphate systems are <1 debye for all four methods shown in Figure 4b. While PM3CARB-1 and $PM3^{MS}$ produce errors that are smaller, they apply only an *sp* basis onto hypervalent atoms. These methods therefore incorrectly model the phosphorus electronic structure and so reduce the accuracy of the chemical interpretation that follows from a QM/MM simulation. This leaves AM1/d-CB1 and AM1/d-PhoT where the former performs better for example the MUEs for glucopyranose-phosphate are 0.69 and 0.95 debye, respectively, and 0.57 and 0.81 debye, respectively, for ribofuranose−phosphate (Table S8, Supporting Information).

Of greater importance to this work is the modeling of carbohydrate conformations that are commonly found in transition state structures or along the reaction coordinate. We computed the dipole moments for C, H, E, B, and S molecular β-D-glucopyranose ring conformers (Tables S9−S10, Supporting Information) and show the MUEs (Figures 4c, d). The carbohydrate H, E, B, and S rings were a particular priority of the VPO strategy, therefore, AM1/d-CB1 shows the smallest deviation from the M06-2X/6-311++G(3df,2p) computed dipoles. For the minimum energy chair α- and β-D-glucopyranose conformers AM1/d-CB1 produces the smallest MUE of 0.45 and 0.19 debye, respectively. A similar trend is found for the transition state glucopyranose conformers with AM1/d-CB1 producing MUEs of 0.39 and 0.17 debye, respectively. Significant improvements in the errors for the dipole moments were achieved from the final optimization of the one-center two-electron repulsion integrals (G_{ss}, G_{pp}, G_{sp}, G_{p2}), one-center two-electron exchange integral (H_{sp}) and

dx.doi.org/10.1021/ct500372s | J. Chem. Theory Comput. 2014, 10, 4694−4707

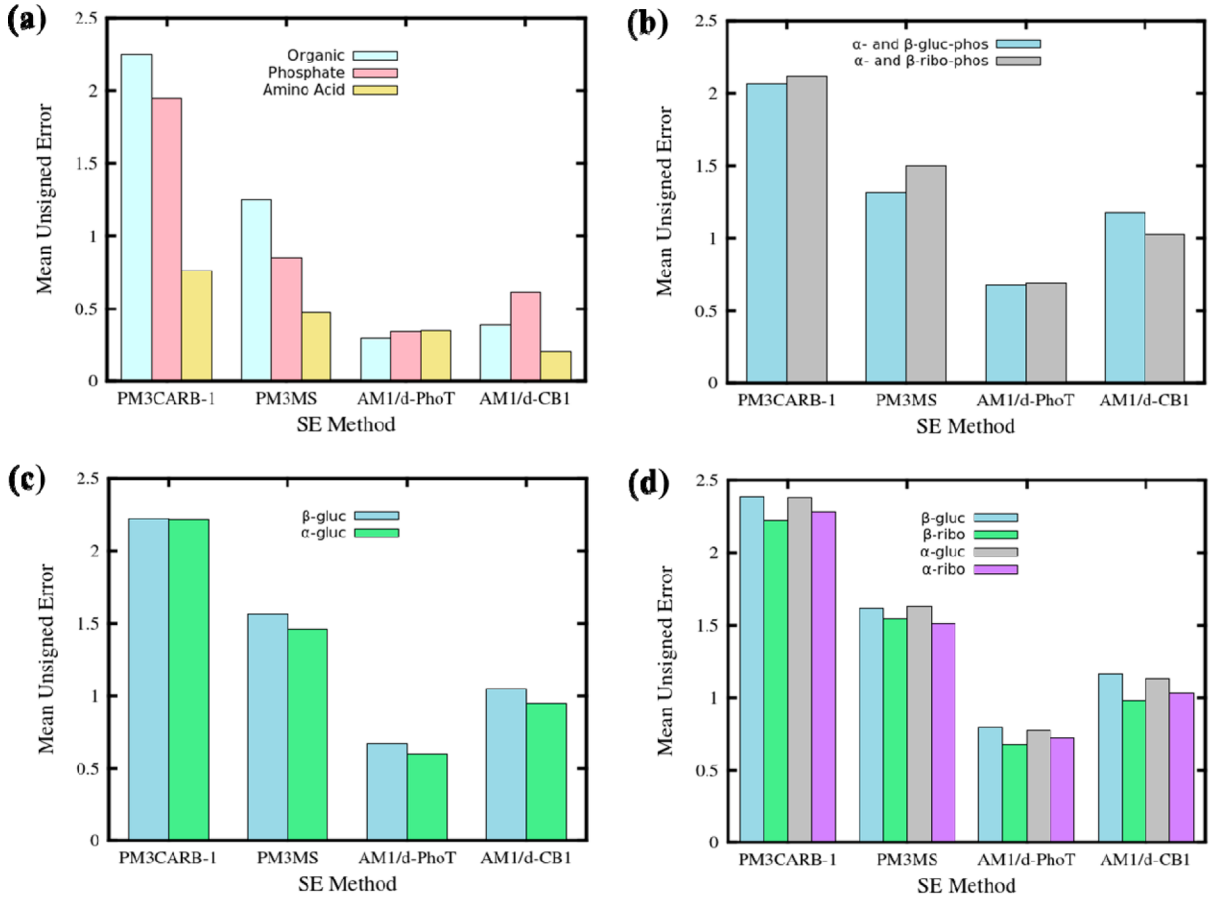


Figure 5. Mean unsigned errors (eV) of ionization potentials for (a) organic molecules, phosphates, and amino acids, (b) carbohydrate chair phosphorylated conformers, (c) carbohydrate chair conformers of glucopyranose and (d) no equilibrium puckered carbohydrate conformers of glucopyranose and ribofuranose.

Table 4. Experimental and Theoretical Interaction Energies for Molecules Used in Parameterization (kcal/mol)

	reference		error			
	exp	DFT[b]	PM3CARB-1	$PM3^{MS}$	AM1/d-PhoT	AM1/d-CB1
H_2O/H_2O	5.00[a]	−5.18	1.82	0.39	0.89	2.91
H_2O/CH_3OH		−5.17	1.96	0.67	1.83	2.72
H_2O/PO_3^-		−15.90	5.62	7.52	0.46	−0.77
$H_2O/H_2PO_4^-$		−18.20	6.75	8.83	−0.19	−2.41
H_2O/HPO_4^{2-}		−33.27	3.50	7.26	3.56	−1.62
MUE (vs DFT)			3.93	4.93	1.38	2.08
MSE (vs DFT)			3.93	4.93	1.31	0.17

[a]Experimental value obtained from Feyereisen et al.[35] [b]The DFT interaction energies where computed with M06-2X/6-31+G(df). All errors are computed as $\Delta H_{int}^{calc} - \Delta H_{int}^{ref}$.

Gaussian terms (FN). However, while this optimization step did cause an increase in the errors for the ionization potential (IP), which we discuss next.

4.4. Ionization Potential. There are two types of IPs, namely *vertical* and adiabatic. The vertical IP is produced when the energy difference between the precursor molecule M_1 and the species M_2, formed by removing an electron have the same molecular geometry. The adiabatic IP arises when M_2 has a geometry, which differs from that of M_1; that is, M_2 has its own equilibrium geometry. For the purposes of the current work, the vertical IP was used. All cationic forms of species listed in Tables S11−S13, Supporting Information, were computed as single point calculations using M06-2X/6-311++G(3df,2p) optimized neutral molecule geometries.

AM1/d-PhoT and AM1/d-CB1 give the smallest errors for all molecular subclasses (Figure 5) with PM3CARB-1 and $PM3^{MS}$ giving considerably larger MUEs. In general the AM1/d-CB1 errors are second only to AM1/d-PhoT. Ionization potentials play an important role in the modeling of nucleophilic reactions. We therefore prioritized the ionization potentials of amino acids in the VPO strategy. The result is that AM1/d-CB1 outperforms all the methods for the chemically glycobiologically significant amino acids used in the training set presented in Figure 5a with an error of 0.20 eV (Table S11, Supporting Information).

Unlike with properties mentioned above, AM1/d-CB1 produces errors that are second to AM1/d-PhoT for the minima and transition state conformers of carbohydrate rings

dx.doi.org/10.1021/ct500372s | J. Chem. Theory Comput. 2014, 10, 4694−4707

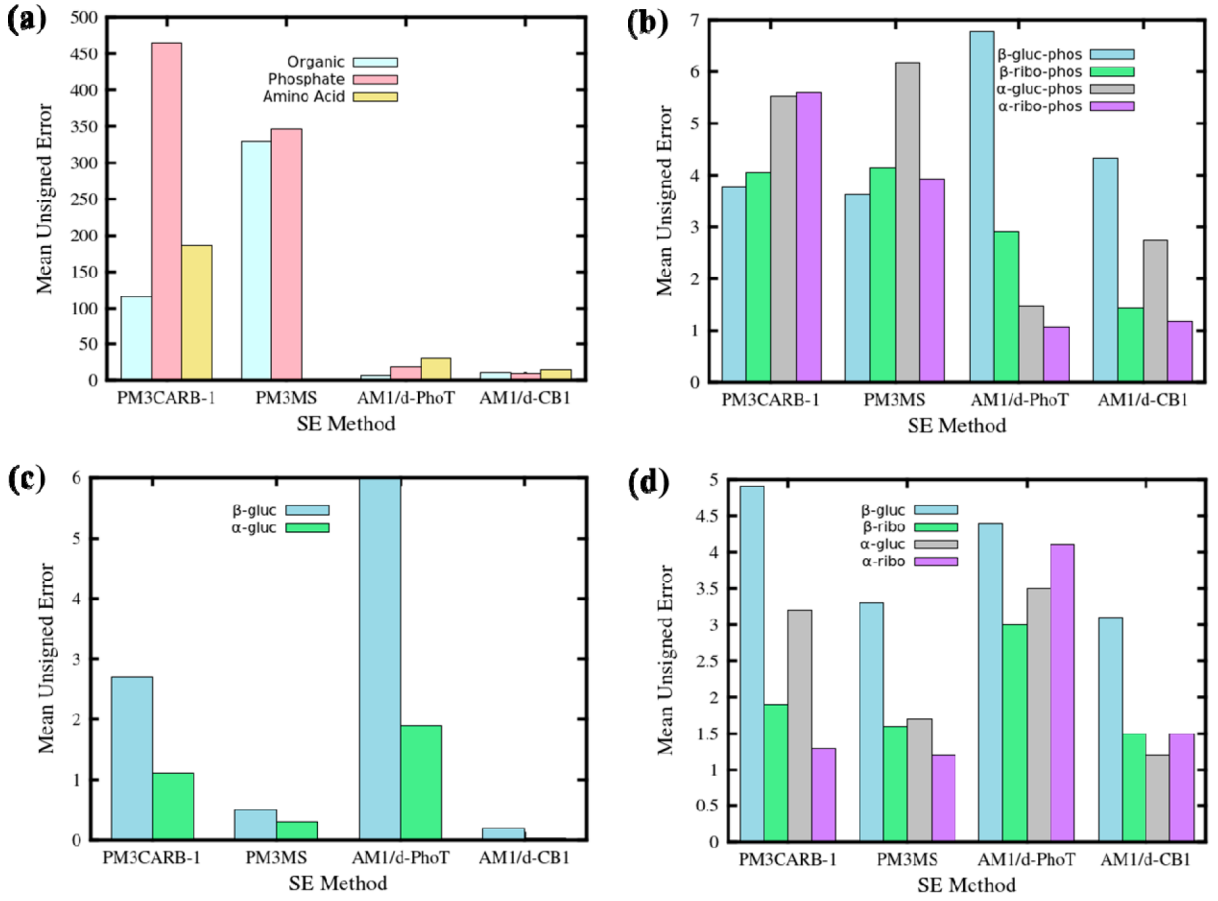


Figure 6. Mean unsigned errors (kcal/mol) for heats of formation for (a) organic molecules, phosphates, and amino acids, (b) carbohydrate chair phosphorylated conformers, (c) carbohydrate chair conformers of glucopyranose, and (d) no equilibrium puckered carbohydrate conformers of glucopyranose and ribofuranose.

(Figure 5c, d). We achieved much better errors for the ionization potential calculations prior to the final optimization step that improved the dipole moments but led to poorer ionization potential performance.

4.5. Interaction Energies. The DFT energies for bimolecular complexes were obtained using M06-2*X*/6-31+G(df). The purpose of this work is to develop a method, which will accurately model reactions important in glycobiology, and such reactions would involve hydrogen bonding with the surrounding water environment. A number of hydrogen bonded dimers were used in the parametrization of AM1/d-CB1. Results for the various hydrogen bond dimers are provided in Table 4 and Table S14 (Supporting Information).

AM1/d-PhoT gives the smallest MUE of 1.38 kcal/mol followed by AM1/d-CB1 where the MUE (2.08 kcal/mol) is approximately half that of PM3CARB-1 and $PM3^{MS}$. A closer look at the individual errors shows that $PM3^{MS}$ yields the smallest errors for the water dimer and water−methanol interaction while giving the highest MUEs for water phosphate complexes. In the case of water-phosphate interactions AM1/d-CB1 compares substantially better with the DFT calculations than does PM3CARB-1 or $PM3^{MS}$. Although some of the errors acquired with AM1/d-CB1 may appear large it should be noted that the results given in Table 4 are being compared to DFT and M06-2X functional that implicitly includes corrections for dispersion and hydrogen bonding. Interaction energies for the complexes shown in Table 4 computed using AM1, PM3, and RM1 are tabulated in Table S14 of the Supporting Information.

4.6. Heats of Formation (ΔH_f). In the optimization process, experimental heats of formation were used as the target where available, whereas DFT results from M06-2X calculations were used when experimental results are absent. In our optimization of AM1/d-CB1 parameters, we assigned the lowest weighting to the heat of formation property. We compare carbohydrate specific methods (PM3CARB-1, $PM3^{MS}$, and AM1/d-CB1) and AM1/d-PhoT that includes a *d*-orbital treatment of phosphorus. The comparative performance of these methods for general organic, phosphates, and amino acid molecular subclasses often present in chemical glycobiology events are shown in Figure 6a. Here AM1/d-PhoT and AM1/d-CB1 produce the smallest errors for organic systems with AM1/d-PhoT yielding a MUE of 7.4 kcal/mol, while AM1/d-CB1 has an error of 10.7 kcal/mol. When introducing longer aliphatic chains into the alcohols (i.e., ethanol (C_2H_5OH), propanol (C_3H_7OH), and finally butanol (C_4H_9OH)), AM1/d-CB1's heat of formation performance declines (8.5, 15.6, and 22.8 kcal/mol, respectively). This is a result of the low weighting attached to ΔH_f during the optimization. AM1/d-CB1 errors for the phosphate molecules (Figure 6a) are very similar to those of AM1/d-PhoT. The error produced for the amino acids with AM1/d-CB1 is 15.0 kcal/mol (Table S15 in Supporting Information), which is two orders in magnitude lower than that of AM1/d-PhoT (30.6 kcal/mol).

AM1/d-CB1 has mixed success for carbohydrate-phosphate systems (Figure 6b). The heats of formation results for the β-anomer of glucopyranose are better for AM1/d-CB1 (MUE

dx.doi.org/10.1021/ct500372s | J. Chem. Theory Comput. 2014, 10, 4694−4707

4.33 kcal/mol) than AM1/d-PhoT (MUE 6.78). Although PM3^{MS} and PM3CARB-1 produce better results than AM1/d-CB1 with errors of 3.64 and 3.77 kcal/mol, respectively, as we pointed out earlier, these methods do not incorporate *d*-orbitals onto the phosphorus representation making them less accurate. For β-D-ribofuranose, AM1/d-CB1 surpasses the other methods by 3 to 4 orders in magnitude, with an MUE of 1.44 kcal/mol (Table S16, Supporting Information). For the α-anomer of glucopyranose, it is AM1/d-PhoT that produces the lower errors of 1.48 kcal/mol, while the α-anomer of ribofuranose is modeled similarly for both AM1/d-PhoT and AM1/d-CB1 with MUE of 1.07 and 1.18 kcal/mol.

For the equilibrated chair conformations of glucopyranose (Figure 6c) AM1/d-CB1 produces the most accurate results with errors of 0.2 and 0.04 kcal/mol for the β- and α-anomers, respectively (Tables S17−S18, Supporting Information). The MUEs for heats of formation of carbohydrate conformers puckered away from the equilibrated chair conformations are shown in Figure 6d. In β-anomers of both glucopyranose and ribofuranose systems, AM1/d-CB1 outperforms all other methods in computing the heats of formation of possible transition state puckered rings. A minor deficiency is its performance that lags behind that of PM3CARB-1 and PM3^{MS} for the α-ribofuranose.

5. CONCLUSION

A parametrization of the AM1/d Hamiltonian initiated from RM1 and AM1/d-PhoT models has been conducted with the aid of a genetic algorithm. The H, C, N, O, and P atoms were tuned using a *variable property optimization (VPO)* parameter approach that prioritizes selective molecular classes for each property (dipole moment, heat of formation, etc.) in addition to weighting properties differently to achieve the goal deriving a parameter set that is capable of modeling a glycan as well as its immediate environment in a chemical glycobiological context. We called this model AM1/d-CB1. In optimizing the semiempirical parameters for these atoms, we prioritized the dynamic performance of ring pucker and key elements of the glycoenzymes reaction class such as proton affinity (commonly associated with acid/base catalysis) and ionization energies (commonly associated with formation of oxocarbenium ions).

Computing accurate transition state properties for glycans in enzymatic reactions has been a universal failing of many SE methods. The property prioritization of ring puckering dynamics, the dipole moments and heats of formation of nonequilibrium ring conformers as well as the proton affinities, heats of formation, and dipole moments of amino acids was central to the development of AM1/d-CB1. The AM1/d-CB1 carbohydrate pyranose and furanose rings are less flexible compared with NDDO methods presented here (except for PM3-CARB1's pyranose ring). Proton affinities of both amino acids and nonequilibrium ring conformers (i.e., transition state structures) are better reproduced with AM1/d-CB1. The dipole moments of transition state ring carbohydrates and amino acids have AM1/d-CB1 surpassing all other NDDO methods considered in this work. However, due to the improved results obtained for the dipole moments the ionization potentials of AM1/d-CB1 were sacrificed. While the errors here are not the best performers, they are nonetheless not poor, as only AM1/d-PhoT gives better results. Overall, the heats of formation for molecules used in the training set show improvement over the existing NDDO type methods. However, the AM1/d model suffers from historically poor NDDO descriptions of hydrogen bond and dispersion interactions. Corrections to these deficiencies are currently under development for AM1/d-CB1, which will further improve its description of glycan reactivity as studied in chemical glycobiology.

The performance of AM1/d-CB1 across the many role players in chemical glycobiological reactions is presented elsewhere. There an evaluation of carbohydrate free energy pucker surfaces and volumes, phosphate reactions, and base pair associations is reported.[15]

■ ASSOCIATED CONTENT

Supporting Information

RMSD figures for various atoms that make up puckering of carbohydrate rings and tables listing detailed calculated property values and errors. 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 Scientific Computing Research Unit (SCRU), Minnesota Supercomputing Institute (Minnesota, U.S.A.) and Center for High Performance Computing (Cape Town, S.A.) for the provision of generous computational resources. We also thank Kwango Nam for valuable input regarding the use of AM1/d-PhoT.

■ REFERENCES

(1) (a) Gloster, T. M.; Davies, G. J. Glycosidase inhibition: Assessing mimicry of the transition state. *Org. Biomol. Chem.* **2010**, *8* (2), 305−320. (b) 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.

(2) Berti, P. J.; McCann, J. A. B. Toward a detailed understanding of base excision repair enzymes: Transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chem. Rev.* **2006**, *106* (2), 506−555.

(3) Kapitonov, D.; Yu, R. K. Conserved domains of glycosyltransferases. *Glycobiology* **1999**, *9* (10), 961−978.

(4) Rye, C. S.; Withers, S. G. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **2000**, *4* (5), 573−580.

(5) Sharma, R. A.; Dianov, G. L. Targeting base excision repair to improve cancer therapies. *Mol. Aspects Med.* **2007**, *28* (3−4), 345−374.

(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) 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.

(8) (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.
(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) 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.
(15) Govender, K. K.; Naidoo, K. J. Evaluating AM1/d-CB1 for chemical glycobiology QM/MM simulations. *J. Chem. Theory Comput.* **2014**, DOI: 10.1021/ct500373p.
(16) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 38.[1] The MNDO Method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(17) Stewart, J. J. P. MOPAC: A semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **1990**, *4* (1), 1−105.
(18) (a) Lopez, X.; York, D. M. Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus. *Theor. Chem. Acc.* **2003**, *109*, 149−159. (b) Thiel, W.; Voityuk, A. A. Extension of MNDO to *d*-orbitals: Parameters and results for the second-row elements and for the zinc group. *J. Phys. Chem.* **1996**, *100*, 616−629. (c) Thiel, W.; Voityuk, A. A. Extension of the MNDO formalism to d orbitals: Intergral approximations and preliminary numerical results. *Theor. Chim. Acta* **1992**, *81*, 391−404. (d) Thiel, W.; Voityuk, A. A. Erratum: Extension of the MNDO formalism to *d*-orbitals: Integral approximations and preliminary numerical results. *Theor. Chim. Acta* **1996**, *93*, 315.
(19) (a) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2003**, *125*, 7178−7179. (b) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 7504−7513.
(20) 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.
(21) López-Canut, V.; Roca, M.; Bertrán, J.; Moliner, V.; Tuñón, I. Promiscuity in alkaline phosphatase superfamily. Unraveling evolution through molecular simulations. *J. Am. Chem. Soc.* **2011**, *133* (31), 12050−12062.
(22) 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.
(23) Jalbout, A. F.; Adamowicz, L.; Ziurys, L. M. Conformational topology of ribose: A computational study. *Chem. Phys.* **2006**, *328* (1−3), 1−7.
(24) (a) Giese, T. J.; Gregersen, B. A.; Liu, Y.; Nam, K.; Mayaan, E.; Moser, A.; Range, K.; Faza, O. N.; Lopez, C. S.; Lera, A. R. d.; Schaftenaar, G.; Lopez, X.; Lee, T.-S.; Karypis, G.; York, D. M. QCRNA 1.0: A database of quantum calculations for RNA catalysis. *J. Mol. Graphics Model.* **2006**, *25* (4), 423−433. (b) QCRNA, http://theory.rutgers.edu/QCRNA/ (accessed June 2013).
(25) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(26) 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.
(27) 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 A.02*, Gaussian, Inc., Wallingford CT, 2009.
(28) (a) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. Direct dynamics calculations with neglect of diatomic differential overlap molecular orbital theory with specific reaction parameters. *J. Phys. Chem.* **1991**, *95*, 4618−4627. (b) Tejero, I.; Gonzalez-Lafont, A.; Lluch, J. M. A PM3/d specific reaction parameterization for iron atom in the hydrogen abstraction catalyzed by soybean lipoxygenase-1. *J. Comput. Chem.* **2007**, *28*, 997−1005.
(29) 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.
(30) (a) Thiel, W. *MNDO97, version 5.0*, University of Zurich, Zurich, Switzerland, 1998. (b) 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.
(31) Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.
(32) Rossi, I.; Truhlar, D. G. Parameterization of NDDO wavefunctions using genetic algorithms. An evolutionary approach to parameterizing potential energy surfaces and direct dynamics calculations for organic reactions. *Chem. Phys. Lett.* **1995**, *233*, 231−236.
(33) (a) Alexeev, Y.; Windus, T. L.; Zhan, C.-G.; Dixon, D. A. Accurate heats of formation and acidities for H_3PO_4, H_2SO_4, and H_2CO_3 from ab initio electronic structure calculations. *Int. J. Quantum Chem.* **2005**, *102* (5), 775−784. (b) Almerindo, G. I.; Tondo, D. W.; Pliego, J. R. Ionization of organic acids in dimethyl sulfoxide solution: A theoretical ab initio calculation of the pK_a using a new parametrization of the polarizable continuum model. *J. Phys. Chem. A* **2003**, *108* (1), 166−171. (c) Fu, Y.; Liu, L.; Li, R.-Q.; Liu, R.; Guo, Q.-X. First-principle predictions of absolute pK_a's of organic acids in dimethyl sulfoxide solution. *J. Am. Chem. Soc.* **2003**, *126* (3), 814−822. (d) Hudáky, P.; Perczel, A. Conformation dependence of pK_a: Ab initio and DFT investigation of histidine. *J. Phys. Chem. A* **2004**, *108* (29), 6195−6205. (e) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. Theoretical evaluation of pK_a in phosphoranes: Implications for

phosphate ester hydrolysis. *J. Am. Chem. Soc.* **2002**, *124* (18), 5010−5018. (f) Magill, A. M.; Cavell, K. J.; Yates, B. F. Basicity of nucleophilic carbenes in aqueous and nonaqueous solventstheoretical predictions. *J. Am. Chem. Soc.* **2004**, *126* (28), 8717−8724. (g) 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. (h) Range, K.; López, C. S.; Moser, A.; York, D. M. Multilevel and density functional electronic structure calculations of proton affinities and gas-phase basicities involved in biological phosphoryl transfer. *J. Phys. Chem. A* **2005**, *110* (2), 791−797. (i) Range, K.; McGrath, M. J.; Lopez, X.; York, D. M. The structure and stability of biological metaphosphate, phosphate, and phosphorane compounds in the gas phase and in solution. *J. Am. Chem. Soc.* **2004**, *126* (6), 1654−1665. (j) 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.

(34) 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.

(35) Feyereisen, M. W.; Feller, D.; Dixon, D. A. Hydrogen bond energy of the water dimer. *J. Phys. Chem.* **1996**, *100* (8), 2993−2997.