JCTC Journal of Chemical Theory and Computation

Specific Reaction Parametrization of the AM1/d Hamiltonian for Phosphoryl Transfer Reactions: H, O, and P Atoms

Kwangho Nam,[†,‡] Qiang Cui,[§] Jiali Gao,[†,‡] and Darrin M. York*[,†]

Department of Chemistry and Supercomputing Institute and the Digital Technology Center, University of Minnesota, Minneapolis, Minnesota 55455−0431, and Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, 1101 University Avenue, Madison, Wisconsin 53706

Received July 30, 2006

Abstract: A semiempirical AM1/d Hamiltonian is developed to model phosphoryl transfer reactions catalyzed by enzymes and ribozymes for use in linear-scaling calculations and combined quantum mechanical/molecular mechanical simulations. The model, designated AM1/d-PhoT, is parametrized for H, O, and P atoms to reproduce high-level density-functional results from a recently constructed database of quantum calculations for RNA catalysis (http://theory.chem.umn.edu/Database/QCRNA), including geometries and relative energies of minima, transition states and reactive intermediates, dipole moments, proton affinities, and other relevant properties. The model is tested in the gas phase and in solution using a QM/MM potential. The results indicate that the method provides significantly higher accuracy than MNDO/d, AM1, and PM3 methods and, for the transphosphorylation reactions, is in close agreement with the density-functional calculations at the B3LYP/6-311++G(3df,2p) level with a reduction in computational cost of 3−4 orders of magnitude. The model is expected to have considerable impact on the application of semiempirical QM/MM methods to transphosphorylation reactions in solution, enzymes, and ribozymes and to ultimately facilitate the design of improved next-generation multiscale quantum models.

1. Introduction

Phosphorylation and dephosphorylation reactions play a central role in biochemical systems.[1] The formation and hydrolysis of ATP constitute a central mechanism for the storage and utilization of chemical energies in metabolic pathways.[2,3] Furthermore, transphosphorylations provide a key regulatory mechanism in eukaryotic cellular signaling.[4] These processes are catalyzed by enzymes such as the proton gradient-driven ATPases and protein kinases and phosphatases. Of particular interest is the study of RNA enzymes or ribozymes.[5] The understanding of the molecular mechanisms of ribozyme catalysis has been greatly facilitated by the study of small prototype self-cleaving RNA[6] such as the hammerhead,[7,8] hairpin,[9,10] and hepatitis delta virus[11−13] ribozymes.

The importance of transphosphorylation reactions has stimulated extensive theoretical and experimental investigations aimed at the identification and characterization of the underlying catalytic mechanisms.[3,14−17] Experimentalists have widely used small model compounds, such as phosphate mono- and diesters with variety of leaving groups, to carry out kinetic experiments. Many different types of experiments led to the consensus that, in solution, hydrolysis of phosphate monoesters follows a predominantly unimolecular dissociative mechanism, whereas phosphate diesters and triesters are

* To whom correspondence should be addressed. E-mail: york@chem.umn.edu.

† Department of Chemistry and Supercomputing Institute, University of Minnesota.

‡ Digital Technology Center, University of Minnesota.

§ Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin.

10.1021/ct6002466 CCC: $37.00
Published on Web 02/09/2007

hydrolyzed by a bimolecular associative mechanism via pentacovalent intermediates, transition states, or both.[3,14,16−18] Although these general conclusions have been drawn for phosphoryl-transfer reactions in solution, there is no consensus about the mechanism in enzymes and ribozymes. Unfortunately, kinetic models are not always able to discern between multiple mechanistic pathways that fit experimental data equally well,[19−21] which underscores the need for theoretical studies to help interpret experimental data.

Ab initio electronic structure methods and density functional theory (DFT) have been widely used to study phosphoryl-transfer mechanisms and to help interpret kinetic data.[19−42] However, most of these calculations have been performed on small model reactions in the gas phase or with the aid of implicit solvation models. Although these approaches provide insight into the nature of the reactions, it is of great interest to study the reactions using explicit models in aqueous solution and in the active sites of enzymes or ribozymes, where specific electrostatic interactions, hydrogen bonds, and solvent dynamics are particularly important.[43−50] Phosphoryl-transfer reactions typically involve highly charged species that undergo considerable changes in hybridization along the reaction path. This necessitates the use of a large, explicit solvation and counterion environment coupled with an accurate quantum mechanical treatment of biological phosphates that includes a d-orbital description of phosphorus. The computational cost of the calculations, however, currently precludes simulations using high-level density-functional methods if sufficient sampling is carried out.

Semiempirical quantum methods such as the modified neglect of diatomic overlap (MNDO),[51,52] Austin model 1 (AM1),[53] and parametric method 3 (PM3)[54] methods are typically 3−4 orders of magnitude faster than DFT methods but currently are not sufficiently accurate to model phosphoryl-transfer reactions. This is, in part, because the parametrization of these models does not include transition state data in the training set and because of the lack of d orbitals for phosphorus. Nonetheless, the remarkable computation efficiency of these methods allows very large scale problems to be addressed using linear-scaling electronic structure methods[55−58] and explicit solvent dynamics through molecular dynamics (MD) simulations using combined quantum mechanical/molecular mechanical (QM/MM) potentials.[48,59−62] The d orbital extension of the MNDO method (MNDO/d)[63,64] has recently been applied to QM/MM simulations of phosphoryl transfer reactions in solution.[45,46,65] Other recently developed semiempirical methods that show promise include the OMx models,[66,67] the PDDG/PM3 model,[68−70] the PM3-MAIS and PM3-PIF models,[71,72] the SCC-DFTB method,[73−76] and the very recent NO-MNDO model.[77] One problem that is prevalent in almost all semiempirical models that utilize a minimal valence basis methods is that they typically underpredict molecular polarizabilities as a function of charge. Recently a method has been proposed that greatly improves the modeling of charge-dependent response properties without increasing the atomic orbital basis and with minimal computational overhead.[78] Although there were efforts to design semiempirical models for specific reactions in phosphatases,[79,80] currently, none of these models yet provides sufficient accuracy for the phosphoryl-transfer reactions of interest in the present work. An important step toward the development of new-generation quantum models that afford significantly greater accuracy and transferability is to first identify and quantify the accuracy limits of existing models for important biological reactions. It is the hope that in this way complementary quantum models for QM/MM calculations such as semiempirical and SCC-DFTB methods can be directly compared and their advantages and disadvantages characterized and understood. These efforts will ultimately help lay the groundwork for the design of more robust and efficient quantum models for biocatalysis.

The aim of the present work is to develop a new semiempirical Hamiltonian model, designated AM1/d-PhoT, that will allow accurate calculation of a wide range of phosphoryl-transfer reactions by determination of a set of specific reaction parameters (SRP)[81] for phosphoryl transfer within the AM1 formulation with d-orbital extension. The new model has been designed to accurately reproduce high-level DFT results such as geometries, dipole moments, proton affinities, and relative energies for a large set of molecules, complexes, and chemical reactions relevant to biological phosphoryl transfer. These data were taken from a recently constructed database of quantum calculations for RNA catalysis (QCRNA).[82,83] The resulting AM1/d-PhoT Hamiltonian is tested in the gas phase and in solution using a combined QM/MM potential and is demonstrated to provide a very good agreement with high-level DFT results for transphosphorylation reactions and offers a considerable improvement over the AM1, PM3, and MNDO/d methods.

2. Theory

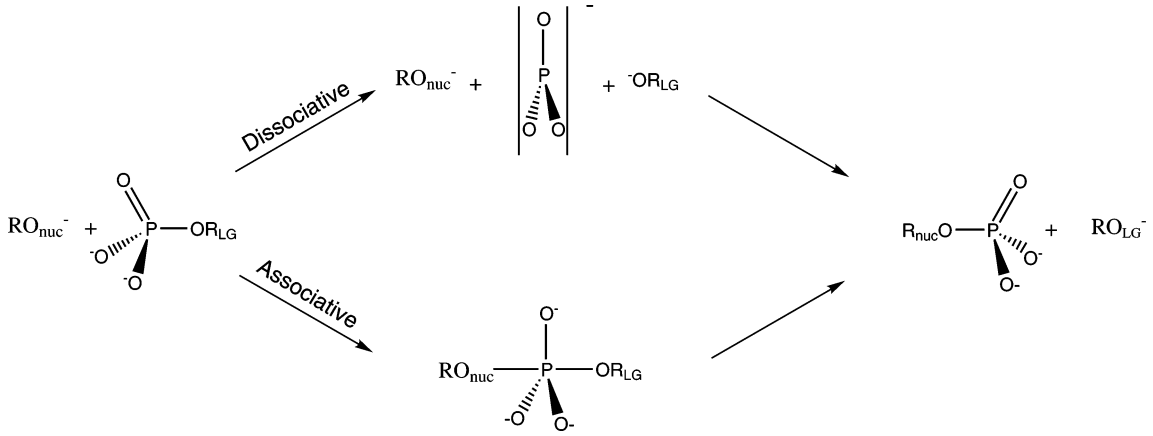
Semiempirical MNDO and AM1 types of Hamiltonians share many common features and have been discussed extensively elsewhere.[84−87] In this section, the major differences of these models are outlined, and a description of the slightly modified AM1/d-PhoT model is provided.

2.1. Semiempirical Core−Core Repulsion Term. The MNDO, AM1, PM3, and MNDO/d semiempirical models discussed here are all based on the neglect of the diatomic differential overlap (NDDO) approximation, as is the new AM1/d-PhoT model. The MNDO and MNDO/d Hamiltonians differ from the AM1/PM3 Hamiltonians in the way that core−core repulsion interactions are treated. Additionally, the MNDO/d method includes a set of d orbitals on second-row elements. In the MNDO and MNDO/d methods, the repulsion between two nuclear cores (A and B) is determined by

$$E_{\mathrm{N}}^{\mathrm{MNDO}}(\mathrm{A},\mathrm{B}) = Z_{\mathrm{A}}Z_{\mathrm{B}}\langle s_{\mathrm{A}}s_{\mathrm{B}}|s_{\mathrm{A}}s_{\mathrm{B}}\rangle(1 + e^{-\alpha_{\mathrm{A}}R_{\mathrm{AB}}} + e^{-\alpha_{\mathrm{B}}R_{\mathrm{AB}}}) \quad (1)$$

where Z_{A} and Z_{B} are the effective core charges, $\langle s_{\mathrm{A}}s_{\mathrm{B}}|s_{\mathrm{A}}s_{\mathrm{B}}\rangle$ is a Coulomb repulsion integral between two s orbitals centered on atoms A and B, and α_{A} and α_{B} are parameters that decrease screening of the nuclear charge by the electrons at small interatomic distances. For the O−H and N−H bonds, a slightly different screening form is used, a detailed

Scheme 1. Reaction Schemes of Dissociative Mechanism and Associative Mechanism for Transesterification Reaction



description for which is provided in the original paper[51] and elsewhere.[88]

In the AM1 and PM3 Hamiltonians, the core−core term includes an additional set of Gaussian terms that depend parametrically on the nature of atoms and takes the form

$$E_{\mathrm{N}}(\mathrm{A,B}) = E_{\mathrm{N}}^{\mathrm{MNDO}}(\mathrm{A,B}) + \frac{Z_{\mathrm{A}}Z_{\mathrm{B}}}{R_{\mathrm{AB}}}\left(\sum_k a_{\mathrm{A}k}\mathrm{e}^{-b_{\mathrm{A}k}(R_{\mathrm{AB}}-c_{\mathrm{A}k})^2} + \sum_k a_{\mathrm{B}k}e^{-b_{\mathrm{B}k}(R_{\mathrm{AB}}-c_{\mathrm{B}k})^2}\right) \quad (2)$$

The addition of Gaussian core−core terms leads to significant improvements in the performance of semiempirical models in energies, especially for interactions at the hydrogen-bonding distances.[54] The MNDO and MNDO/d models are known to be problematic in the description of noncovalent intermolecular interactions because of the excessive repulsion just outside bonding distances. The Gaussian core−core terms in eq 2 compensate for these undesirable repulsions. The Gaussian core−core terms are empirical adjustments to the potential devoid of any rigorous physical meaning. Previous studies showed that serious artifacts can be introduced if one only changes the empirical Gaussian core−core repulsion functions.[72,89] Nevertheless, several new core−core interaction potentials or parameters have been proposed, including the PDDG/PM3 model,[68] the PM3-MAIS and PM3-PIF models,[71,72] and the $\mathrm{PM3_{BP}}$ model.[88] These models have been designed to offer improvements in the description of proton-transfer reactions and hydrogen bonds.

2.2. AM1/d-PhoT Model Employing the Modified AM1 Formalism. It has been well-established that the study of hypervalent phosphates, such as the transition states and reactive intermediates formed in biological transphosphorylation, require an explicit *d* orbital representation.[63,64,90] In particular, transphosphorylation reactions typically involve a change of valency on phosphorus along the reaction path: it changes from tetracovalent to tricovalent in the dissociative mechanism or changes from tetracovalent to pentacovalent in the associative and concerted mechanisms (Scheme 1). The MNDO/d model[63,64] has been demonstrated to perform reliably in the study of transphosphorylation under basic conditions,[45,46] where there is no hydrogen bonding or proton transfer that occurs within the quantum region. However, the MNDO/d model was found to be of limited reliability for transphosphorylation reactions in the neutral to acidic pH range where hydrogen bonding and proton transfer play major roles.

The derivation of a new model for phosphoryl transfer reactions under more general conditions requires components of both the AM1/PM3 and MNDO/d models. The AM1 and PM3 models offer a significant improvement for hydrogen bonding relative to MNDO and MNDO/d models but have the problem that they over-stabilize hypervalent structures because of the artificially attractive core−core interactions.[64,70,90] Therefore, in the design of an AM1/d model for phosphoryl transfer reaction, it is desirable to keep the core−core interactions for hydrogen bonding but to turn these interactions off for phosphorus bonding where the d orbitals allow proper hybridization and accurate representation of hypervalent species. In this way, a balanced model may be achieved that accurately models reactive intermediates in transphosphorylation with, at the same time, improved treatment of hydrogen bonding. Toward this end, a scale factor was introduced into the Gaussian core−core terms in the present AM1/d-PhoT model as

$$E_{\mathrm{N}}(\mathrm{A,B}) = E_{\mathrm{N}}^{\mathrm{MNDO}}(\mathrm{A,B}) + \frac{Z_{\mathrm{A}}Z_{\mathrm{B}}}{R_{\mathrm{AB}}} G_{\mathrm{scale}}^{\mathrm{A}} G_{\mathrm{scale}}^{\mathrm{B}}\left(\sum_k a_{\mathrm{A}k}\mathrm{e}^{-b_{\mathrm{A}k}(R_{\mathrm{AB}}-c_{\mathrm{A}k})^2} + \sum_k a_{\mathrm{B}k}\mathrm{e}^{-b_{\mathrm{B}k}(R_{\mathrm{AB}}-c_{\mathrm{B}k})^2}\right) \quad (3)$$

where $G_{\mathrm{scale}}^{\mathrm{A}}$ and $G_{\mathrm{scale}}^{\mathrm{B}}$ are scaling parameters for atoms A and B, and in the present work, they vary from zero to one (values of 0 recover the conventional MNDO core−core model, whereas values of 1 recover the AM1 core−core model). Alternatively, the product $G_{\mathrm{scale}}^{\mathrm{A}}\,G_{\mathrm{scale}}^{\mathrm{B}}$ can be made into pairwise terms for specific atom pairs. It is worthwhile to mention here that setting Gaussian core−core parameters $a_{\mathrm{A}k}$ to be zero for a certain atom, such as P, *would not* eliminate the Gaussian core−core interactions involving P atoms since all terms from the other atoms would remain. The G_{scale} scaling parameters hence provide the flexibility to attenuate (or even shut off) Gaussian core−core interactions between certain atoms and offers a simple mechanism for interconverting between AM1-like models and MNDO-

Table 1. Optimized AM1/d-PhoT Parameters for Hydrogen, Oxygen, and Phosphorus Atoms, along with Original AM1 Parameters for Comparison[a]

	H		O		P	
parameters	AM1/d	AM1	AM1/d	AM1	AM1/d	AM1
U_{ss} (eV)	−10.934610	−11.396427	−96.760676	−97.830000	−46.250810	−42.029863
U_{pp} (eV)			−78.776203	−78.262380	−40.712918	−34.030709
ζ_s (au)	1.143846	1.188078	3.057965	3.108032	1.909168	1.981280
ζ_p (au)			2.515332	2.524039	2.008466	1.875150
β_s (eV)	−5.911108	−6.173787	−29.472306	−29.272773	−11.194791	−6.353764
β_p (eV)			−28.515785	−29.272773	−11.985621	−6.590709
α (1/Å)	2.884915	2.882324	4.404417	4.455371	1.883237	2.455322
G_{ss} (eV)	13.737453	12.848000	14.234714	15.420000	14.645747	11.560005
G_{pp} (eV)			14.454530	14.520000	11.694918	7.877589
G_{sp} (eV)			14.539451	14.480000	5.689654	5.237449
G_{p2} (eV)			12.942259	12.980000	10.328696	7.307648
H_{sp} (eV)			4.339705	3.940000	1.175115	0.779238
U_{dd} (eV)					−24.504161	
ζ_d (au)					0.840667	
β_d (eV)					−2.360095	
$\tilde{\zeta}_s$ (au)					2.085120	
$\tilde{\zeta}_p$ (au)					1.535336	
$\tilde{\zeta}_d$ (au)					1.236266	
ρ_{core} (au)					1.185130	
G_{scale}[b] (unitless)	1.000000		1.000000		0.353722	
a_1 (unitless)	0.106238	0.122796	0.288526	0.280962	−0.344529	−0.031827
b_1 (1/Å^2)	5.735290	5.000000	4.883265	5.000000	3.034933	6.000000
c_1 (Å)	1.261430	1.200000	0.850910	0.847918	1.134275	1.474323
a_2 (unitless)	0.004043	0.005090	0.061586	0.081430	−0.021847	0.018470
b_2 (1/Å^2)	7.080122	5.000000	4.435791	7.000000	1.684515	7.000000
c_2 (Å)	2.084095	1.800000	1.353681	1.445071	2.716684	1.779354
a_3 (unitless)	−0.002800	−0.018336			−0.036003	0.033290
b_3 (1/Å^2)	0.739913	2.000000			5.243357	9.000000
c_3 (Å)	3.649474	2.100000			1.924175	3.006576

[a] Standard notation for parameters taken from refs 84 and 63. [b] Scale factor of Gaussian core−core repulsion interactions.

like models. As will be shown below, adjustment of the G_{scale} parameter allows significantly improved behavior of AM1/d-PhoT model for phosphoryl transfer reactions along with the specific reaction parametrizations.

3. Parametrization Procedure

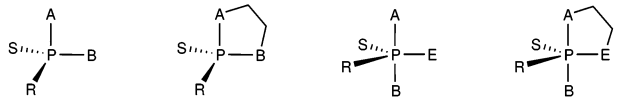
This section describes the methods and procedures employed to develop the specific reaction parameters for the AM1/d-PhoT model. The first subsection describes the DFT dataset used as the reference data. The second subsection describes the details of the parameter optimization procedure. The final parameters for the AM1/d-PhoT model are listed in Table 1.

3.1. Density Functional Calculations. All of the DFT datasets in this work, including the full structural and thermodynamic quantities, are available in the recently constructed QCRNA database[82] which is available on-line.[83] All DFT calculations were carried out using Kohn−Sham density functional theory with the three-parameter hybrid exchange functional of Becke[91,92] and the Lee, Yang, and Parr correlation functional[93] (B3LYP). Energy minimum and transition state geometry optimizations were carried out with default convergence criteria, while the stability conditions of the restricted closed-shell Kohn−Sham determinant for each final structure were verified.[94,95] Frequency calculations were performed to establish the nature of all stationary points and to evaluate thermodynamic quantities. The 6-31++G-(d,p) basis set was used in the calculations.

Electronic energies and dipole moments were refined by single-point energy calculations using the 6-311++G-(3df,2p) basis set and the B3LYP hybrid density functional at the optimized geometries. All single-point calculations were run with the convergence criteria on the SCF wave function tightened to 10^{-8} au to ensure high precision for properties sensitive to the use of diffuse basis functions.[96] The protocol applied to obtain the (refined energy)//(geometry and frequencies) is designated by the abbreviated notation B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G-(d,p). All DFT calculations were performed with the GAUSSIAN03 suite of programs.[97] This density-functional protocol has been extensively tested and applied to model biological phosphorus compounds.[23−25,47,98,99] The gas-phase proton affinities at 298.15 K for the DFT dataset were calculated by the procedure described by Range et al. without empirical correction.[100]

3.2. AM1/d-PhoT Parameter Optimization. This section presents details for the AM1/d-PhoT parametrization procedure for biological phosphorus compounds. The first section describes the construction of a set of compounds and their properties that are used as the target in the parameter

Scheme 2. Nomenclature Convention for Ligand Designations in Acyclic and Cyclic Phosphate and Phosphorane Compounds of Biological Interest[a]



P(R)(S)(A)(B) P(R)(S)(AB) P(R)(S)(E)(A)(B) P(R)(S)(EA)(B)

[a] This nomenclature is consistent with the naming convention used for similar compounds in previous work.[25,98]

optimization. The second subsection describes the construction of χ^2 merit function that was used in the parametrization procedure. The third subsection briefly outlines the parametrization strategy that was applied to arrive at the final optimized semiempirical parameters (Table 1).

3.2.1. Construction of Training Dataset. The primary focus of the current work is to develop specific reaction parameters that accurately reproduce the structures, energetics, and other properties for the transphosphorylation reactions. Taking into account of the catalytic mechanisms of several ribozymes, protein kinases, and phosphatases, which involve general acid and base catalysis, we produced a molecular set which was used to construct the training set for parametrizations and which includes the following list of properties.

Molecules. Eleven non-phosphorus-containing molecules and polyatomic ions and 16 phosphorus-containing molecules and polyatomic ions that included metaphosphates, phosphates, and phosphoranes in various protonation states, and one diphosphate ion were included in the training set. (See Scheme 2 for the nomenclature convention for phosphorus-containing compounds.)

Complexes. Three hydrogen-bonded complexes involving water with water, methanol, and acetate were included in the dataset.

Reaction Mechanisms. Three dissociative reactions and five associative reactions of various phosphates and all stationary points of these reactions were included.

Potential Energy Surfaces. One hundred fifty intermediate configurations along both one and two-dimensional potential energy surfaces (PES) for three reactions were included in the final stages of parameter refinement, one of which was chosen from the five associative reactions mentioned above, and two additional reactions were added.

All compounds and structures selected in the parameter training sets, energies, and relevant other properties are available from the on-line QCRNA database via the Internet (see ref 83). The properties considered in the parametrizations included the following.

Heat of Formation. Fourteen experimental values of heats of formation were used, including three phosphorus containing compounds. They are listed in Table 2.

Proton Affinity. Seven experimental gas-phase proton affinity (PA) values were included. In addition, 14 relative PA values for phosphate and phosphorane compounds calculated with DFT were used. These values are shown in Table 3.

Dipole Moment. The dipole moments for 11 neutral molecules and four hydrogen-bonded complexes determined from DFT calculations were used and are shown in Table 4.

Structure. Geometrical parameters including bond lengths, angles, and dihedrals were considered for all molecules in the training set. If the molecules or complexes form hydrogen bonds, the hydrogen bond distances and angles were also included in the parametrizations. The results are summarized in Table 5.

Intermolecular Interaction Energy. The intermolecular interaction energies of three hydrogen-bonded complexes were included and are listed in Table 6.

Relative Reaction Energy. The relative energies of intermediates and transition states for all reactions in the training

Table 2. Experimental (exptl) and Computational Heats of Formation for Compounds Used in Parametrizations (kcal/mol)

	exptl	AM1/d		AM1		PM3		MNDO/d	
molecule[a]	ΔH_f	ΔH_f[b]	error[c]	ΔH_f[b]	error[c]	ΔH_f[b]	error[c]	ΔH_f[b]	error[c]
H_3O^+	138.0[d]	140.1	1.2	143.5	4.6	159.1	20.2	134.2	−4.7
H_2O	−57.8[d]	−56.8	1.0	−59.2	−1.4	−53.4	4.4	−60.9	−3.1
HO^-	−33.2[e]	−26.8	6.4	−14.1	19.1	−17.5	15.7	−5.8	27.4
CH_3OH	−48.1[f]	−53.7	−5.6	−57.0	−8.9	−51.9	−3.8	−57.4	−9.3
CH_3O^-	−36.0[e]	−35.9	0.1	−38.5	−2.5	−37.9	−1.9	−39.7	−3.7
C_2H_5OH	−56.2[f]	−53.5	2.7	−62.7	−6.5	−56.9	−0.7	−63.0	−6.8
$C_2H_5O^-$	−47.5[e]	−38.1	9.4	−45.5	−2.0	−44.8	2.7	−45.3	2.2
C_6H_5OH	−23.0[f]	−19.2	3.8	−22.2	0.8	−21.7	1.3	−26.7	−3.7
$C_6H_5O^-$	−40.5[e]	−38.2	2.3	−41.0	−0.5	−44.1	−3.6	−42.2	−1.7
CH_3CO_2H	−103.3[f]	−102.5	0.8	−103.0	0.3	−102.0	1.3	−101.1	2.2
$CH_3CO_2^-$	−122.5[e]	−123.7	−1.2	−115.4	7.1	−119.7	2.8	−110.0	12.5
$P(CH_3)_3$	−24.2[g]	−16.7	7.5	−22.0	2.2	−29.8	−5.6	−28.1	−3.9
$(CH_3)_3PO$	−103.8[h]	−105.8	−2.0	−101.5	2.3	−82.7	21.1	−105.5	−1.7
PO_3^-	−225.4[d]	−206.1	19.3	−203.0	22.4	−197.2	28.2	−257.3	−31.9
MSE			3.3		2.9		5.9		−1.8
MUE			4.5		5.8		8.1		8.1

[a] Bold molecules are included in the parametrization as test sets. [b] The heat of formation is determined by $\Delta H_f = E_{el}^{mol} + E_{core}^{mol} - \sum_A E_{el}^A + \sum_A \Delta H_f^A$, where E_{el}^{mol} and E_{core}^{mol} are the molecular electronic and core−core repulsion energy, respectively, and E_{el}^A and ΔH_f^A are the electronic energy and the experimental heat of formation for isolated atom A, respectively, where the summations run over all the atoms A in the molecule. [c] Error values are computed as $\Delta H_f^{calcd} - \Delta H_f^{Exptl}$. [d] Ref 131. [e] Ref 132. [f] Ref 133. [g] Ref 134. [h] Ref 135.

Table 3. Gas-Phase Proton Affinity (PA) of Selected Molecules (kcal/mol)[a]

	ref		error[e]				
molecule (AH)[b]	exptl[c]	CBS[d]	AM1/d	AM1	PM3	MNDO/d	DFT[f]
H_3O^+	165.0	164.1	3.8	−2.0	−11.8	5.6	−1.1
H_2O	390.3	392.1	5.4	20.5	11.3	30.6	0.1
CH_3OH	381.5	382.8	2.0	2.7	−1.9	1.8	−2.2
C_2H_5OH	378.2	379.0	2.9	4.7	−0.4	5.2	−2.2
C_6H_5OH	350.1	349.5	−3.4	−3.1	−6.9	0.0	−2.4
CH_3CO_2H	347.2	347.5	−2.7	6.1	0.9	9.6	−0.8
P(O)(OH)(OH)(OH)	330.5	328.1	−3.4	7.6	15.0	−12.2	−2.4
MSE (vs exptl)[g]		0.0	0.7	5.2	0.9	5.8	−1.6
MUE (vs exptl)[g]		1.2	3.4	6.7	6.9	9.3	1.6
HPO_3		310.6	1.5	20.6	35.1	−3.6	−0.1
P(O)(O)(OH)(OH)$^-$		458.9	−1.9	16.8	24.7	−2.8	−1.1
P(O)(O)(O)(OH)2−		581.1	10.4	33.7	36.4	16.3	−1.7
P(O)(OH)(OH)(OCH_3)		329.3	0.3	7.2	14.9	−12.0	0.4
P(O)(O)(OH)(OCH_3)$^-$		454.9	0.7	16.5	22.8	−7.6	−1.4
P(O)(OH)(OCH_3)(OCH_3)		329.4	1.8	7.3	12.3	−14.1	0.7
P(O)(OH)(OCH_2CH_2O)		329.5	−0.2	7.6	11.8	−17.1	−0.1
P(OH)(OH)(OH)(OH*)(OH)		351.0	3.0	9.0	8.3	−1.3	−0.4
P(OH*)(OH)(OH)(OH)(OH)		341.0	1.8	13.6	9.0	−8.7	−1.8
P(OH)(OH)(OCH_2CH_2O)(OH*)		351.9	1.2	5.9	1.7	−11.8	−0.9
P(OH*)(OH)(OCH_2CH_2O)(OH)		343.2	−2.5	8.0	−0.5	−17.4	−1.1
P(OH*)(OCH_3)(OCH_2CH_2O)(OH)		345.2	−3.5	3.6	−2.3	−20.2	−0.7
P(OH)(OCH_3)(OCH_2CH_2O)(OH*)		352.0	2.3	5.4	−0.4	−27.0	−0.8
P(OH*)(OH)(OCH_2CH_2O)(OCH_3)		343.5	−0.7	6.2	−0.9	−19.5	−1.1
MSE (vs CBS-QB3)			1.0	11.5	12.4	−10.5	−0.7
MUE (vs CBS-QB3)			2.3	11.5	12.9	12.8	0.9

[a] PA is defined as the negative of the enthalpy change (ΔH) of the process of $A^-_{(g)} + H^+_{(g)} \rightarrow AH_{(g)}$. [b] Bold molecules and the deprotonated form (A^-) are included in parametrization as training set molecules. [c] The experimental (exptl) and calculated (CBS) reference data were taken from refs 136 and 100, respectively. In computing semiempirical PA values, the experimental heats of formation of proton (365.7 kcal/mol) is used.[137] [d] CBS is a value computed from CBS-QB3 method reported in reference.[100] [e] Error values are computed as $PA_{calcd} - PA_{ref}$, where ref is the exptl values whenever experimental values are available. Otherwise, CBS-QB3 values are used as reference values. [f] DFT proton affinity is computed as a sum of adiabatic energy contribution at B3LYP/6-311++G(3df,2p) level and the enthalpic correction computed from B3LYP/6-31++G(d,p) level. The geometries are optimized at the level of B3LYP/6-31++G(d,p) as described in section 3. [g] The MSE and MUE values are computed for molecules with known experimental values indicated with vs exptl.

set were considered in the parameter optimization. The reference data were relative "adiabatic" DFT energies. The strategy here is to have the semiempirical model reproduce adiabatic potential energy surfaces from which quantum mechanical zero-point corrections, tunneling effects, and thermodynamic corrections can be accounted for explicitly in QM/MM simulations.[101−104] The relative reaction energies are listed in Tables 8−13.

It should be noted that, in the final parameter refinement procedure, all structures (including transition states) in the training set were explicitly optimized along all unconstrained degrees of freedom for every trial semiempirical parameter set. For structures that were minima, all degrees of freedom were fully optimized; transition state structures were fully optimized (transition state search using gradient norm minimization) to a first-order saddle point (with the exception of the dianionic dissociative reactions that, for stability, had the cleaved P−O bond constrained to the DFT values), and for PESs all degrees of freedom excluding the PES variables were fully optimized (i.e., the PESs were adiabatic).

3.2.2. Construction of χ^2 Merit Function. The optimized AM1/d-PhoT parameters are determined by nonlinear optimization of the χ^2 merit function that measures the quality of properties predicted by the trial parameter set, λ, against the target data summarized above. The $\chi^2(\lambda)$ merit function is defined as

$$\chi^2(\lambda) = \sum_{i}^{\text{mol}} \sum_{\alpha}^{\text{prop}(i)} w_{i\alpha}(Y_{i\alpha}^{\text{AM1/d}}(\lambda) - Y_{i\alpha}^{\text{Ref}})^2 \qquad (4)$$

where the first summation with index i runs over molecules, complexes, or reactions (mol), and the second summation with index α runs over properties associated with the ith "molecule" (prop(i)). The term $Y_{i\alpha}^{\text{AM1/d}}(\lambda)$ is the value of the property α for molecule i calculated with a set of trial parameters, λ, $Y_{i\alpha}^{\text{Ref}}$ is the corresponding target value from experiment or from density-functional calculations, and $w_{i\alpha}$ is the associated weight in the regression. The weights $w_{i\alpha}$ are defined as the inverse square of the $\sigma_{i\alpha}$ values in eq 5

$$w_{i\alpha} = (\sigma_{i\alpha}{}^2)^{-1} \qquad (5)$$

in which the $\sigma_{i\alpha}$ values have the same units as the molecular property $Y_{i\alpha}$, and specifies the significance or importance of this particular property.

3.2.3. Nonlinear Parameter Optimizations. In the optimization of the χ^2 merit function, a suite of integrated

Table 4. Gas-Phase Dipole Moments (debye) of Selected Molecules and Hydrogen-Bonded Complexes

molecules[a]	DFT[b]	AM1/d	AM1	PM3	MNDO/d
		isolated molecules			
H_2O	1.91	2.39	1.86	1.74	1.78
CH_3OH	1.69	2.33	1.62	1.49	1.48
C_2H_5OH	1.61	2.25	1.55	1.45	1.40
C_6H_5OH	1.29	1.90	1.23	1.14	1.17
CH_3CO_2H	1.82	2.69	1.89	1.84	1.68
$P(CH_3)_3$	1.24	1.98	1.52	1.08	1.20
$(CH_3)_3PO$	4.52	5.76	5.01	3.92	3.57
P(O)(O)(OH)	3.21	3.35	2.37	1.80	3.05
P(O)(OH)(OH)(OH)	0.51	0.13	3.74	3.16	1.58
P(O)(OH)(OH)(OCH_3)	0.98	1.18	1.06	3.15	1.86
P(O)(OH)(OCH_3)(OCH_3)	1.24	1.37	1.10	2.18	1.97
P(O)(OCH_3)(OCH_3)(OCH_3)	1.09	0.47	0.84	0.43	2.23
P(O)(OH)(OCH_2CH_2O)	4.10	5.37	4.02	2.30	3.52
P(O)(OCH_3)(OCH_2CH_2O)	3.88	5.07	3.75	2.30	3.70
MSE		0.51	0.18	−0.08	0.08
MUE		0.65	0.42	0.91	0.47
		hydrogen bond complexes			
H_2O:H_2O[c]	2.89	3.05	2.53	2.48	2.71
CH_3OH:H_2O[c]	2.87	2.74	2.23	2.23	2.90
H_2O:P(O)(OH)(OH)(OCH_3)[c]	1.90	1.87	1.70	0.99	1.84
H_2O:P(O)(OH)(OCH_3)(OCH_3)[c]	3.32	2.66	1.10	2.18	1.97
H_2O:P(O)(OCH_3)(OCH_3)(OCH_3)[c]	3.57	2.98	3.54	3.33	4.75
CH_3OH:P(O)(OH)(OH)(OCH_3)[c]	1.93	1.82	1.60	1.22	1.90
CH_3OH:P(O)(OH)(OCH_3)(OCH_3)[c]	3.43	2.87	1.92	2.98	2.94
CH_3OH:P(O)(OCH_3)(OCH_3)(OCH_3)[c]	3.78	3.67	1.93	3.39	2.63
H_2O:P(O)(OH)(OCH_2CH_2O)[c]	2.85	4.32	4.42	2.34	2.17
MSE		−0.06	−0.62	−0.60	−0.30
MUE		0.42	0.97	0.60	0.57

[a] Bold molecules/complexes are included in parametrization as a training set molecules/complexes. [b] Reference DFT dipole moments are computed from B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) calculations. [c] Dipole moments of hydrogen-bonded complex.

nonlinear optimization methods were used[90] that included (1) a genetic algorithm and (2) a direction set of minimization methods. Genetic algorithms[105,106] have been demonstrated elsewhere to be useful in semiempirical parameter optimization.[81,107−109] The implementation of the genetic algorithm was based on the description by Goldberg[105] and used tournament selection and multidimensional phenotypic (parameter set) niching. The quadratically convergent direction set optimization utilized the method of Powell and has been described in detail elsewhere.[110] In the present work, we follow a stepwise approach of (1) initial coarse-grained parameter optimization, followed by (2) final parameter refinement and testing. The final optimized parameters are listed in Table 1. All properties were computed using a modified version of the MNDO97 program.[111]

3.2.3.1. Initial Coarse-Grained Parameter Optimization. On the basis of the original AM1 parameters, the hydrogen and oxygen atomic parameters were adjusted first. The optimization was carried out using a quadratically convergent direction set optimization method[110] with narrow parameter bands to avoid large changes in atomic parameters from their starting AM1 values (approximately within 10%). In the present specific reaction parameter optimization, although we restrict the computed absolute heats of formation to be close to the experimental values, it is important to obtain a good estimate of the absolute proton affinities and to balance the quality of relative energies for acid−base reactions to model general acid and base catalysis. Since the H and O atoms are directly involved in the protonation reaction in the PA calculations, we decided not to change the AM1 parameters for carbon. It was also assumed that the Gaussian core−core parameters on H and O atoms were sufficiently well balanced in the original AM1 model. Thus, the scaling parameters G_{scale}^{H} and G_{scale}^{O} were held fixed at 1.0.

After obtaining an initial optimized set of parameters for hydrogen and oxygen, we turned into the optimization of the phosphorus atomic parameters, which includes a set of functions for d orbitals. In view of the fact that the MNDO/d model was successful in transphosphorylation reactions in the absence of H-bonding and proton transfer, the initial phosphorus parameters were taken from the MNDO/d model except two one-center two-electron integral parameters (G_{sp} and H_{sp}) and the Gaussian core−core parameters, which were taken from AM1 for the consistency with the other atoms. The starting G_{scale}^{P} value was set to 1.0. Starting from this set of initial parameters, we performed several steps of direction set optimization, followed by optimization using a genetic algorithm to optimize the Gaussian core−core parameters and the G_{scale} value of P. The details of the implementation of genetic algorithms for nonlinear parameter optimizations will be described elsewhere.[112] In the present

Table 5. Comparison of Gas-Phase Geometries for the Reactions in the Training and Test Sets[a]

	AM1/d		AM1		PM3	
	MSE	MUE	MSE	MUE	MSE	MUE
			reactants and products			
bond (P−O)	0.000	0.009	−0.020	0.026	0.042	0.051
bond (P−$O_{Nuc/LG}$)[b]	−0.166	0.179	0.224	0.255	0.083	0.153
bond (H−O_{Ac})[c]	0.098	0.114	0.410	0.447	0.039	0.163
angle (O−P−O)	−0.11	1.44	−0.18	2.34	−0.16	1.81
angle (O_{Nuc}−P−O_{LG})[b]	0.70	13.77	−5.85	7.47	−8.10	9.63
angle (O_{Do}−H−O_{Ac})[c]	−21.83	21.83	−22.35	24.24	−0.60	7.28
			phosphorane intermediates			
bond (P−O_{ax})[d]	−0.026	0.026	−0.086	0.086	−0.002	0.022
bond (P−O_{eq})[d]	−0.002	0.014	−0.029	0.029	0.038	0.045
angle (O_{ax}−P−O_{ax})[d]	−7.01	7.75	−5.60	6.32	−19.76	19.76
angle (O_{ax}−P−O_{eq})[d]	0.55	2.51	0.43	2.31	1.25	7.32
angle (O_{eq}−P−O_{eq})[d]	−0.01	5.28	0.00	5.10	−0.33	7.68
			transition states			
bond (P−$O_{Nuc/LG}$)[b]	0.052	0.121	0.190	0.445	0.158	0.397
bond (P−O_{ax})[d]	−0.006	0.017	0.039	0.056	0.040	0.050
bond (P−O_{eq})[d]	−0.010	0.012	0.102	0.102	0.031	0.043
bond (H−$O_{Ac/Do}$)[e]	0.089	0.105	0.042	0.139	−0.003	0.110
angle (O_{ax}−P−$O_{Nuc/LG}$)[b,d]	−2.14	3.82	−4.62	6.79	−9.52	10.44
angle (O_{eq}−P−$O_{Nuc/LG}$)[b,d]	−0.82	3.25	−1.23	6.31	−0.06	7.19
angle (O_{eq}−P−O_{eq})[d]	−0.53	2.05	−1.37	3.69	−0.64	2.79
angle (O_{Do}−H−O_{Ac})[e]	−16.39	16.39	−9.11	15.88	−9.23	16.63

[a] All error values are computed as $Prop_{calcd} - Prop_{DFT}$, where reference geometries are from B3LYP/6-31++G(d,p) optimized geometries. the unit for the bond length is angstroms, and for the angle, it is degrees. [b] Nuc and LG are the oxygen atom in the nucleophilic and leaving group, respectively, and Nuc/LG stands for either nucleophile or leaving group. [c] Ac and Do are the hydrogen bond acceptor and donor group, respectively. [d] ax and eq are the oxygen atoms at the axial and equitorial position of phosphorane geometry, respectively. [e] Do and Ac are the donor and acceptor at the transition state for the proton transfer.

Table 6. Intermolecular Interaction Energies (kcal/mol) for Hydrogen-Bonded Complexes

complex[a]	DFT[b]	AM1/d	AM1	PM3	MNDO/d	ref
H_2O:H_2O	−4.8	−4.2	−5.4[c]	−3.5	−0.7	−5.4[d] −5.0[e]
H_2O:CH_3OH	−4.6	−3.8	−5.1	−3.2	−0.6	
H_2O:$CH_3CO_2^-$	−15.7	−16.5	−13.8	−13.5	−6.7	
H_2O:P(O)(O)$(O)^-$	−14.2	−16.4	−13.1	−9.1	−4.5	
H_2O:P(O)(O)(OH)$(OH)^-$	−16.0	−19.1	−14.1	−12.3	−6.2	
H_2O:P(O)(O)(O)$(OH)^{2-}$	−30.4	−31.2	−20.6	−24.7	−9.7	
H_2O:P(O)(O)$(OCH_2CH_2O)^-$	−15.3	−17.2	−13.2	−10.9	−4.9	
MSE[f]		−1.1	2.1	3.4	9.7	
MUE[f]		1.5	2.6	3.4	9.7	

[a] Bold complexes are included in parametrization as a training set hydrogen bond complexes. [b] B3LYP/6-311++G(3df,2p) evaluated adiabatic hydrogen bond energy. [c] AM1 water dimerization energy is for the bifurcated geometry. In the case of nonplanar C_s structure, the interaction energy is −3.3 kcal/mol. [d] Experimental hydrogen bond energy of water dimer.[138] [e] Benchmark ab initio hydrogen bond energy of water dimer.[139] [f] Errors are computed as $E_{calcd} - E_{DFT}$, where E is hydrogen bond energy.

work, genetic algorithm runs were performed using population sets of 128 members that were allowed to evolve over 100 generations. The parameter sets with the highest fitness were then refined with the quadratically convergent direction set method for 100 iterations. During the optimizations, parameters were typically allowed to vary by only around 5−10% of their initial values. This combined genetic algorithm−direction set optimization procedure was repeated, sometimes adjusting the parameter value bounds, until satisfactory results were obtained. It should be mentioned that in the process of nonlinear optimizations, multiple minima in the parameter space were identified and systematically eliminated on the basis of further testing and evaluation using the extended (testing) dataset.

3.2.3.2. Final Parameter Refinement. Up to this point, the training set only included stationary and transition state geometries. After the coarse-grained parameter optimization passed, additional geometries taken from the relaxed potential energy surfaces were included in the training set to map the potential energy surface along the reaction pathway explicitly. Because of the extreme sensitivity of the reaction energy profile and the existence of phosphorane intermediates on the protonation and alkyl substitution level, the parametrizations only with stationary point geometries do not guarantee the correct curvature and shape of potential energy surface for the reactions, except in the most simple reactions.[81] Therefore, the inclusion of geometries from potential energy surfaces for the reactions is critical to improve the accuracy of AM1/d-PhoT model with correct energy curvature near the transition states. All H, O, and P atomic parameters were refined simultaneously under narrow parameter bounds (typically 5% or less) using the direction set optimization scheme to arrive at the final optimized parameter set.

4. Results and Discussion

This section presents an analysis and discussion of the results from the semiempirical AM1/d-PhoT parametrizations for biological transphosphorylation described above. The optimized AM1/d-PhoT parameters for the H, O, and P atoms are shown in Table 1, along with the original AM1 parameters for comparison. The results are compared with experimental data, high-level DFT calculations from the QCRNA database,[82] and results from other semiempirical models including AM1, PM3, and MNDO/d. In some cases, convergence problems in transition state optimizations were encountered with the MNDO/d model, particularly for reaction steps that involved proton transfers, which disallowed comparison with the present results. In all tables and figures, the AM1/d-PhoT model is simply referred to as AM1/d.

4.1. Heats of Formation. The main focus of this paper is to develop a semiempirical model that is highly accurate in describing phosphoryl transfer reactions, including the structures and relative energies of the transition state and reactive intermediates. Nonetheless, to create a more robust model, we included a selected set of experimental heat of formation data. Table 2 shows the heats of formation (experimental, AM1/d-PhoT, AM1, PM3, and MNDO/d) obtained for the training set compounds. The overall performance of the AM1/d-PhoT model in computing heats of formation of molecules is similar with other semiempirical models. The mean signed errors (MSE) in the computed heat of formation for the AM1, PM3, MNDO/d, and AM1/d-PhoT models are 2.9, 5.9, −1.8, and 3.3 kcal/mol, respectively, whereas the corresponding values in the mean unsigned error (MUE) are 5.8, 8.1, 8.1, and 4.5 kcal/mol. In particular, the AM1, PM3, and MNDO/d have been known to overestimate the heats of formation of the hydroxide ion by 15−27 kcal/mol, but the AM1/d-PhoT model reduces the error to 6.4 kcal/mol. This offers rather significant improvement but still underscores the need to consider functional forms that allow a more balanced treatment of the stability and response properties of small anions.[78]

4.2. Gas-Phase Proton Affinity. The protonation states, pK_a values, and pH-dependent rate profiles are major issues in the identification of the underlying catalytic mechanisms of biological transphosphorylation both in enzymatic and in nonenzymatic systems. Hence, the inclusion of the absolute and relative proton affinities of molecules into the semiempirical parametrization is important in the study of biological phosphoryl transfer.[100] The calculated proton affinities (PAs) are compared with experimental and CBS-QB3[113,114] calculated values in Table 3, and Figure 1 shows the distribution of computed PA values of neutral and monoanionic species against reference CBS-QB3 calculated PAs. Previous studies have shown that the CBS-QB3 calculations have high chemical accuracy in computing absolute proton affinity.[100] The computed results from the standard AM1, PM3, and MNDO/d models, as well as DFT calculations, are also shown for comparison. The AM1/d-PhoT results exhibit considerable improvement over other semiempirical models. For the molecules with available experimental PA values, the performance of the AM1/d-PhoT, PM3, and DFT is similar in term of MSE values, which are 0.7, 0.9, and −1.6 kcal/mol, respectively, whereas the AM1 and MNDO/d models have MSE values greater than 5 kcal/mol. However, the comparisons of MUE values show that the AM1/d-PhoT model (3.4 kcal/mol) considerably outperforms the other semiempirical models (6.7 kcal/mol or greater). For molecules with phosphorus, which lack experimental PA values, the calculated proton affinities are compared against the CBS-QB3 results. The MSE and MUE values from AM1/d-PhoT calculations are only slightly worse (1.0 and 2.3 kcal/mol, respectively) than the DFT results (−0.7 and 0.9 kcal/mol, respectively), whereas other semiempirical methods shows much larger errors (over 10 kcal/mol in magnitude). When the potential inaccuracy in the evaluation of the energies of highly negatively charged molecules from semiempirical models employing minimal basis is considered, the error analysis with exclusion of the PO_4^{3-}/HPO_4^{2-} pair, which shows a maximum error in PA, further decreases the MSE and MUE values to 0.3 and 1.6 kcal/mol, respectively. The AM1/d-PhoT model is even more encouraging in the comparison of the relative PA values. For example, the proton transfer from acetic acid to the phenolate ion, which is found in several enzymatic reactions including protein tyrosine phosphatase,[79] is exothermic from the experiment, CBS-QB3, DFT, and AM1/d-PhoT calculations, while the AM1, PM3, and MNDO/d models yield endothermic reaction energies.

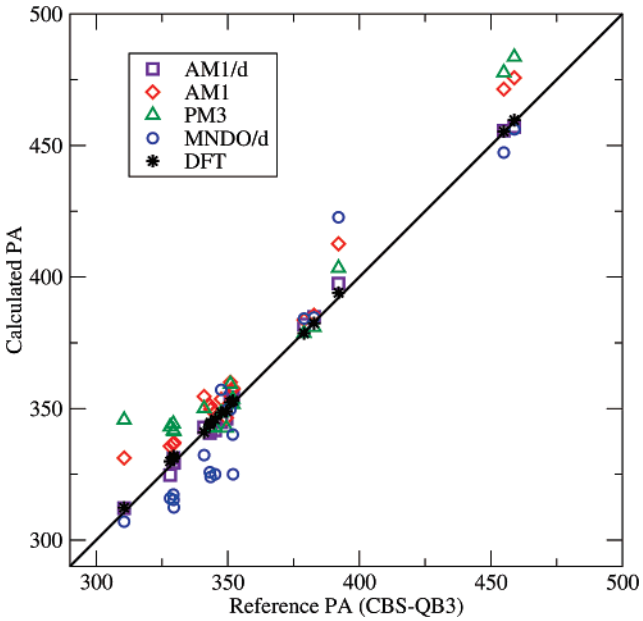


Figure 1. Gas-phase proton affinities of neutral and monoanionic species. The reference proton affinities are from CBS-QB3 calculations. Methods used in the calculations are indicated by various labels in the figure, and units are kilocalories per mole.

Table 3 also includes phosphorane compounds and the PA results at the axial and equitorial protonation sites. The phosphorane compounds are formed as intermediates in the associative reaction paths. Overall, the PAs are larger in phosphoranes than in similar phosphate compounds. The relative trends of absolute PAs between phosphorane and phosphate pairs are correctly reproduced from the AM1/

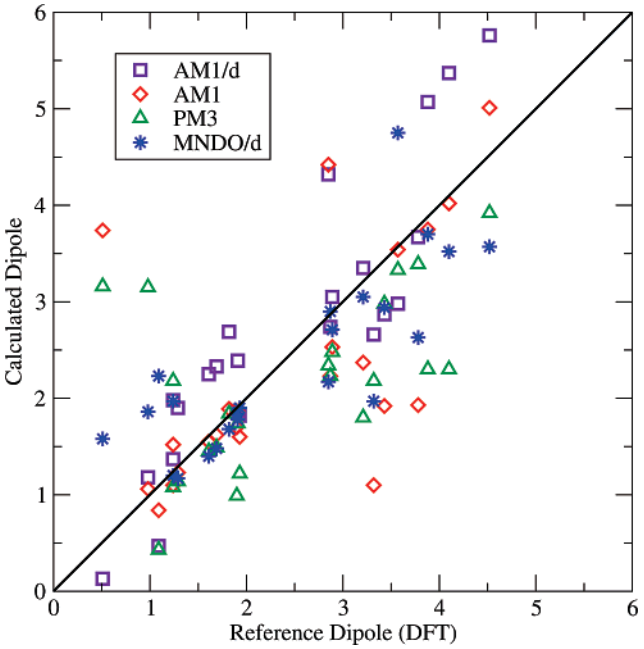


Figure 2. Scatter plot of gas-phase dipole moment. The reference dipole moments are from calculations at the B3LYP/6-311++G(3df,2p) level. The unit of dipole moments is debye.

d-PhoT model, whereas the other semiempirical methods show larger variations. The AM1 and PM3 models are reasonably reliable for the relative PAs between similar phosphorane/phosphate pairs. However, for proton transfers between dissimilar functional groups, as in enzymes, ribozymes, and solution reactions, the predictive capability of the methods is considerably reduced. It is paramount to have a quantum model that is able to predict correct absolute PA values to ensure transferability between diverse sets of proton donor−acceptor pairs. In addition, the AM1 and PM3 models show larger error with increasing negative charge on the phosphate compounds, but the AM1/d-PhoT model has smaller, more balanced errors.

4.3 Dipole Moments. The calculated dipole moments for 14 molecules and 9 hydrogen bond complexes are shown in Table 4, and Figure 2 shows the scatter plot of predicted dipole moments against reference dipole moments from the DFT calculations. In the parametrizations, the DFT dipole moments were used in the reference dataset. The error analysis shows that the MSE and MUE values for the dipole moments of isolated molecules from the AM1/d-PhoT model are 0.51 and 0.65 debye, respectively. The results are comparable to values from other semiempirical methods. The MSE and MUE values are 0.18 and 0.42 debye from AM1, −0.08 and 0.91 debye from PM3, and 0.08 and 0.47 debye from MNDO/d, respectively. The MNDO/d method is best compared to DFT dataset, while the AM1/d-PhoT model slightly overestimates the dipole moments of isolated molecules. It is sometimes the case that the deviation of the dipole moments with respect to the DFT dataset values arises from the differences in the optimized geometries at the semiempirical level. It is known that the semiempirical methods systematically underestimate the molecular polarizability because of the limitations inherent in the use of a minimal valence basis set.[78] Thus, in the absence of a quantum model that predicts accurate polarization response, it might be desirable to have the dipole moment slightly overestimated in the gas-phase such that the semiempirical models reproduce correct interaction energy in the condensed phase. The comparisons of hydrogen bond complexes support this. The computed dipole moments from the AM1/d-PhoT calculations reproduce the DFT dataset results, while other semiempirical models underestimate the interaction energy by 0.30−0.62 debye. The results suggest that the AM1/d-PhoT model can be a good method for interactions in aqueous solution. Alternately, next-generation models that allow for more accurate charge-dependent response properties[78] would be a preferable alternative. It should be noted here that proper optimizations of the QM/MM van der Waals parameters are also critical to reproduce correct relative solvation free energies from condensed-phase simulations using combined QM/MM potentials.[115−117] This aspect will be addressed in more detail in future work in the application of the AM1/d-PhoT model to study biological transphosphorylation reactions.

4.4. Geometries. In Table 5, the analysis on the structures of phosphorus compounds are presented for the reactions in training and test sets. (For details of the reactions, refer to section 4.6 and Tables 8−11.) The AM1/d-PhoT, AM1, and PM3 optimized geometries are compared with DFT optimized geometries. Since the MNDO/d model fails at the optimization of transition states for most reactions, they are not compared in this section. The geometries are compared separately at the reactant and product states, phosphorane intermediates, and transition states.

For phosphate geometries of reactants and products, the AM1/d-PhoT model performs better than the AM1 and PM3 models, but the geometries are well reproduced from all three methods tested (Table 5). The situation for the hydrogen bond geometries is somewhat different among various models at reactant and product complexes. In particular, the angle of $O_{Nuc}-P-O_{LG}$, in which the O_{Nuc} and O_{LG} are the nucleophilic and leaving oxygens in the reaction, has large deviation from the DFT values for all three models tested (Table 5). The hydrogen bond lengths in the complexes are reasonably close to the DFT geometries for the AM1/d-PhoT and PM3 models, but the errors in bond angle are large. For hydrogen bond angles, the PM3 model performs better than the AM1/d-PhoT and AM1 models. For phosphoranes, the three methods reproduce DFT geometry accurately. The AM1 model yields the smallest MSE and MUE, which is followed by the AM1/d-PhoT model, whereas the PM3 model underestimates the bond angle by as much as 20° in some cases (Table 5).

For the transition state, the most important geometrical parameters are the bond lengths of the P atom from the oxygen atoms of the nucleophile or the leaving group for nucleophilic substitution and addition−elimination reactions, and the distance from the transferring proton to donor and acceptor atoms ($H-O_{Ac/Do}$). The AM1/d-PhoT model has the smallest MSE and MUE values of 0.052 and 0.121 for the $P-O_{Nuc}$ and $P-O_{LG}$ distances and 0.089 and 0.105 for $H-O_{Ac}$ and $H-O_{Do}$ lengths, respectively. Although the PM3 model predicts quite accurate geometries for proton-transfer

Table 7. Comparison of Geometries for Hydrogen-Bonded Complexes Given in Table 6[a]

	AM1/d		AM1		PM3		MNDO/d	
	H···O[b]	O−H−O[c]	H···O[b]	O−H−O[c]	H···O[b]	O−H−O[c]	H···O[b]	O−H−O[c]
MSE	0.010	−9.40	0.838	−31.97	0.450	−12.93	1.586	−19.27
MUE	0.073	9.62	0.838	37.83	0.751	16.19	1.586	20.14

[a] All error values are computed as $Prop_{calc} - Prop_{DFT}$, where reference geometries are from B3LYP/6-31++G(d,p) optimized geometries. The unit for the bond length is angstroms, and that for the angle is degrees. [b] Hydrogen to O_{Ac} distance. [c] The angle formed by O_{Do}, hydrogen, and O_{Ac} atoms.

reactions, the errors in $P-O_{Nuc/LG}$ bond are too large. For example, the maximum error of PM3 for $P-O_{Nuc/LG}$ bond length is 1.255 Å, while the maximum error for the $P-O_{Nuc/LG}$ bond from AM1/d-PhoT model is 0.347 Å.

4.5. Hydrogen-Bonding Energy and Geometry. For bimolecular complexes, the performance of the AM1/d-PhoT model is compared with other semiempirical models. Although the semiempirical interaction energies are considered to be the enthalpies of association,[51,53,54,68] we compare the results with adiabatic interaction energies without including temperature and zero-point vibrational energies. Our purpose in the development of the present specific reaction parameters is to use a semiempirical model to accurately reproduce the potential energy surfaces for transphosphorylation reactions in condensed phase simulations. The results are presented in Tables 6 and 7. The AM1/d-PhoT model shows an improvement in the hydrogen bond geometries and interaction energies compared to the standard AM1, PM3, and MNDO/d methods. The MUE for the intermolecular interaction energy with the AM1/d-PhoT model is 1.5 kcal/mol, whereas the AM1, PM3 and MNDO/d models have corresponding MUE values of 2.6, 3.4, and 9.7 kcal/mol, respectively. Importantly, the AM1/d-PhoT model overcomes the well-documented problem in the AM1 dimer structure for water, which has bifurcated geometry. The interaction energy for a water dimer is −4.2 kcal/mol from the AM1/d-PhoT model, and the hydrogen bond distance between hydrogen bond donor and acceptor is 2.97 Å, which is comparable to the 2.89 Å value from DFT dataset. Table 6 presents six other hydrogen bond complexes including four phosphorus-containing complexes. In particular, the hydrogen bond energy of water with an ion increases with the increase of charge of ions, and the AM1/d-PhoT model performs better than other semiempirical models, in which existing semiempirical models considerably underestimate the interaction energy between H_2O and $P(O)(O)(O)(OH)^{2-}$.

The results from an error analysis of the hydrogen bond geometries are presented in Table 7. The AM1/d-PhoT model results in more accurate hydrogen bond geometries. The MSE and MUE values from the AM1/d-PhoT model are 0.010 and 0.073 Å for the hydrogen bond distance and −9.40 and 9.62° for the angle formed by hydrogen bond donor, H atom, and acceptor. The AM1 and MNDO/d models predict quite different hydrogen bond geometries from the reference geometries. However, compared with Table 5, the PM3 model performs quite reliably, although it is problematic in balancing the geometries and the interaction energies. One example is the hydrogen bond between $P(O)(O)(O)(OCH_2-CH_2O)^-$ and the water molecule. The DFT and AM1/d-PhoT model predict a doubly hydrogen-bonded complex, while the PM3 model predicts a singly hydrogen-bonded complex and causes a large error in angle.

4.6 Reaction Energetics. The reaction energies from the AM1/d-PhoT model for a series of reactions with total charge ranging from 0 to −2 are listed in Tables 8−11. The reaction dataset includes reaction energies, barrier heights, and relative energies of intermediates for over 19 associative reactions and 3 dissociative reactions. Note that although these reactions cover a fairly broad range of mechanisms in the gas phase, they do not consider more complex reaction mechanisms that may occur in the aqueous phase, such as water-assisted pathways[37,118,119] or complex water relays.[120] These mechanisms often require fairly sophisticated transition path sampling techniques[121,122] and sometimes intricate bridging water chains that are currently not amenable to efficient automated parameter optimization. The computed values from the AM1 and PM3 models are also compared with the calculated DFT values. Although the MNDO/d model was successful in the transphosphorylation reaction in the absence of hydrogen bond and proton transfer, this model fails in optimizing the geometries at the transition state involving proton transfers, which makes it difficult to compare the performance with other semiempirical methods and is not included in the comparisons. The results from error analysis are summarized in Tables 12 and 13, and Figure 3 shows a scatter plot of computed energies against the DFT reference values. The reactions are classified on the basis of the total charge and mechanism. Thus, associative (neutral, monoanionic, and dianionic) and dissociative (monoanionic) mechanisms are considered (Scheme 1). It is assumed that all associative reactions proceed via in-line reaction pathways, whenever the reaction involves a proton transfer from a nucleophile or to a leaving group. Figure 4 shows the reaction energy profiles of representative reactions for each reaction type. In this figure, the calculations using the MNDO/d model are carried out from relaxed potential energy surface scans constrained at different values of the $P-O_{Nuc/LG}$ and $H-O_{Ac/Do}$ lengths to obtain an approximate transition state, while the direct transition state optimizations are carried out for other semiempirical models. The dianionic dissociative reactions are not listed because of the failure of semiempirical methods in locating the products after the formation of the P−O bond.

The AM1/d-PhoT model performs noticeably better than the AM1 and PM3 models in the associative reactions (see Table 12). The PM3 method severely underestimates the barrier heights. In contrast, the AM1 model performs well for certain reactions and poorly for others. The computed MSE values for neutral and monoanionic reactions are 3.28 and 0.24 kcal/mol for the AM1 model and −18.76 and

Table 8. In-line Associative Neutral Reactions and Energies (kcal/mol) Used in Parametrizations and Their Tests[a]

reaction[b]	type	DFT	error ($E_{calcd} - E_{DFT}$) AM1/d	AM1	PM3	label
H_2O + P(O)(OH)(OH)(OCH_3) ⇌ CH_3OH + P(O)(OH)(OH)(OH)	R	11.39	4.54	−2.24	−9.19	$\mathbf{A}_1^0$
	$TS_{2'}$	35.01	−1.30	−1.16	−14.35	
	I	16.64	−1.59	−49.56	−31.57	
	$TS_{5'}$	36.05	−4.21	1.30	−14.79	
	PS	0.87	−1.34	−7.34	−7.30	
	P	12.94	3.83	−13.51	−18.66	
H_2O + P(O)(OH)(OCH_3)(OCH_3) ⇌ CH_3OH + P(O)(OH)(OH)(OCH_3)	R	5.65	5.87	5.05	−1.31	$\mathbf{A}_2^0$
	$TS_{2'}$	34.69	2.73	4.83	−17.28	
	I	15.19	−0.96	−43.17	−27.01	
	$TS_{5'}$	37.59	4.25	5.39	−24.17	
	PS	0.60	−3.32	−4.94	−7.88	
	P	6.92	5.22	−2.96	−10.71	
H_2O + P(O)(OCH_3)(OCH_3)(OCH_3) ⇌ CH_3OH + P(O)(OH)(OCH_3)(OCH_3)	R	6.40	3.91	−15.86	−1.44	$\mathbf{A}_3^0$
	$TS_{2'}$	40.70	−0.98	2.67	−17.93	
	I	20.36	−3.51	−45.09	−30.54	
	$TS_{5'}$	40.81	−5.41	3.75	−22.67	
	PS	1.32	−1.92	−6.04	−9.65	
	P	7.43	3.31	−23.78	−12.06	
CH_3OH + P(O)(OH)(OH)(OCH_3) ⇌	R	11.74	5.18	−3.40	−11.13	$\mathbf{A}_4^{0,c}$
	$TS_{2'}$	35.26	−2.07	9.24	−20.38	
	I	19.84	−2.17	−42.20	−26.02	
CH_3OH + P(O)(OH)(OCH_3)(OCH_3) ⇌	R	5.99	4.36	1.44	−2.24	$\mathbf{A}_5^{0,c}$
	$TS_{2'}$	35.45	−0.72	12.72	−6.47	
	I	17.66	−2.41	−38.02	−20.80	
CH_3OH + P(O)(OCH_3)(OCH_3)(OCH_3) ⇌	R	6.32	3.98	−0.21	−2.08	$\mathbf{A}_6^{0,c}$
	$TS_{2'}$	40.61	−3.55	10.54	−8.23	
	I	22.61	−3.05	−39.54	−21.81	
P(O)(OH)(OH)(OCH_2CH_2OH) ⇌ H_2O + P(O)(OH)(OCH_2CH_2O)	$TS_{2'}$	29.23	5.68	9.76	−19.66	$\mathbf{A}_7^0$
	I	10.76	2.59	−40.49	−26.04	
	$TS_{5'}$	35.99	9.17	5.44	−26.72	
	PS	−2.59	4.85	5.05	2.74	
	P	9.48	6.81	0.19	−6.91	
P(O)(OCH_3)(OCH_3)(OCH_2CH_2OH) ⇌ CH_3OH + P(O)(OCH_3)(OCH_2CH_2O)	$TS_{2'}$	31.19	−0.34	−4.10	−28.27	$\mathbf{A}_8^0$
	I	15.33	2.60	−40.21	−29.08	
	$TS_{5'}$	38.88	6.58	−15.10	−22.92	
	P	10.95	5.51	−9.57	−17.52	

[a] All the energies are relative to the reactant hydrogen-bonded complex. The reference DFT energies are computed as adiabatic energies from B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level. [b] Bold reactions are included in the parametrization as a training set reaction. [c] Symmetric reactions.

−13.01 kcal/mol from the PM3 model, respectively. In this regard, the MSE values are 0.76 and −1.12 kcal/mol from the AM1/d-PhoT model for both reaction types. The same conclusion is drawn from comparison of the MUE values. The relative energies of intermediates with respect to the reactant complexes are fairly satisfactory from the AM1/d-PhoT calculations, but they are severely underestimated using AM1 and PM3 models. These two methods produce a huge stabilization in the energy of phosphorane compounds, resulting from the use of a minimal valence basis and the unbalanced use of Gaussian core−core interactions for phosphate and phosphorane. The artificial stabilizations of the pentacoordinate phosphorus compounds are transferred into the energy at the transition state. This causes systematic underestimation of barrier heights for most reactions using the PM3 model and several reactions from the AM1 model (Tables 8, 9, and 12). This is further exacerbated in the relative energies between hydrogen-bonded reactant complexes and reaction intermediates. On the other hands, the AM1/d-PhoT model accurately reproduces the energetics from DFT calculations within several kilocalories per mole in the reaction energies, barrier heights, and relative energies of reaction intermediates. This shows the importance of the d orbitals in the handling of hypervalent phosphorus compounds and the effectiveness of scaled Gaussian core−core interactions.

For dianionic associative reactions, similar artificial stabilizations of hypervalent phosphorus compounds from the AM1 and PM3 calculations are observed (Table 10). The barrier heights from the AM1 and PM3 models are still underestimated considerably. Again, the AM1/d-PhoT model performs well in the prediction of reaction energies and barrier heights.

It is interesting to note that there are large deviations of reaction energies for the associative reactions computed from the semiempirical AM1 and PM3 models and that many

Table 9. In-line Associative Monoanionic Reactions and Energies (kcal/mol) in Parametrizations and Their Test[a]

reaction[b]	type	DFT	error ($E_{calcd} - E_{DFT}$) AM1/d	AM1	PM3	label
H_2O + P(O)(O)(OH)(OCH_3)$^-$ ⇌ CH_3OH + P(O)(O)(OH)(OH)$^-$	R	16.25	2.55	−1.85	−4.70	$\mathbf{A}_1^-$
	$TS_{2'}$	44.59	1.99	−13.36	−5.42	
	I	37.23	−7.55	−48.16	−41.91	
	$TS_{5'}$	41.83	−1.43	−10.07	−24.25	
	PS	1.31	−0.45	−1.31	−6.89	
	P	15.97	1.13	−9.73	−10.98	
H_2O + P(O)(O)(OCH_3)(OCH_3)$^-$ ⇌ CH_3OH + P(O)(O)(OH)(OCH_3)$^-$	R	17.31	1.00	−10.60	−10.93	$\mathbf{A}_2^-$
	$TS_{2'}$	39.20	1.62	1.61	−7.32	
	I	36.69	−5.02	−54.38	−44.78	
	$TS_{5'}$	40.94	−6.68	−4.38	−46.03	
	PS	−0.06	−1.88	−9.72	−11.72	
	P	15.68	0.58	−14.82	−16.03	
CH_3OH + P(O)(O)(OH)(OCH_3)$^-$ ⇌	R	15.76	1.79	−5.18	−4.65	$\mathbf{A}_3^{-,d}$
	$TS_{2'}$	43.08	−0.42	−30.73	−14.30	
	I	36.76	−6.89	−43.63	−35.20	
CH_3OH + P(O)(O)(OCH_3)(OCH_3)$^-$ ⇌	R	14.56	2.55	−9.26	−2.76	$\mathbf{A}_4^{-,d}$
	$TS_{2'}$	41.40	−4.51	−9.91	−18.12	
	I	39.62	−4.91	−20.24	−32.00	
CH_3O^- + P(O)(OCH_3)(OCH_3)(OCH_3) ⇌	R	20.04	4.89	−2.38	1.40	$\mathbf{A}_5^{-,c,d}$
	$TS_{2'}$	10.30	−5.95	28.82	12.23	
	I	1.69	−5.61	−45.48	−18.25	
P(O)(OH)(OCH_3)(OCH_2CH_2O)$^-$ ⇌ HO^- + P(O)(OCH_3)(OCH_2CH_2O)	R	5.43	7.38	4.53	−2.19	$\mathbf{A}_6^{-,c}$
	$TS_{2'}$	2.81	2.16	3.74	−5.33	
	I	−2.27	−2.19	−38.56	−30.01	
	$TS_{5'}$	23.84	2.59	12.83	6.17	
	P	37.32	12.97	19.15	5.53	
P(O)(O)(OCH_3)(OCH_2CH_2OH)$^-$ ⇌ CH_3OH + P(O)(O)(OCH_2CH_2O)$^-$	$TS_{2'}$	32.21	−1.03	17.54	−14.18	$\mathbf{A}_7^-$
	I	32.16	−1.71	−43.76	−37.36	
	$TS_{5'}$	40.80	−0.71	6.59	−26.57	
	PS	3.77	−2.34	−6.11	−21.32	
	P	18.43	−0.40	−12.43	−16.08	

[a] All the energies are relative to the reactant hydrogen-bonded complex. The reference DFT energies are computed as adiabatic energies from the B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level. [b] Bold reactions are included in the parametrization as a training set reaction. [c] No proton transfer is involved in the reaction. [d] Symmetric reactions.

Table 10. In-line Associative Dianionic Reactions and Energies (kcal/mol) in Parametrizations and Their Test[a]

reaction[b]	type	DFT	error ($E_{calcd} - E_{DFT}$) AM1/d	AM1	PM3	label
H_2O + P(O)(O)(O)(OCH_3)$^{2-}$ ⇌ CH_3OH + P(O)(O)(O)(OH)$^{2-}$	R	29.42	0.89	−8.85	−5.37	$\mathbf{A}_1^{2-,c,e}$
	$TS_{5'}$	50.43	−3.47	−48.56	−36.74	
	PS	1.99	−0.68	−5.85	−3.41	
	P	32.84	−2.82	−16.78	−10.61	
CH_3OH + P(O)(O)(O)(OCH_3)$^{2-}$ ⇌	R	29.98	−1.18	−10.14	−6.44	$\mathbf{A}_2^{2-,c,d,e}$
	$TS_{5'}$	50.56	−5.95	−33.90	−33.73	
CH_3O^- + P(O)(O)(OCH_3)(OCH_3)$^-$ ⇌	$TS_{5'}$	94.38	−4.66	−10.85	−24.12	$\mathbf{A}_3^{2-,d}$
P(O)(O)(OCH_3)(OCH_2CH_2O)$^{2-}$ ⇌ CH_3O^- + P(O)(O)(OCH_2CH_2O)$^-$	$TS_{5'}$	43.07	−0.43	−7.14	−38.92	$\mathbf{A}_4^{2-,e}$
	P	−44.10	0.84	−12.93	−21.53	

[a] The reference energy for the $\mathbf{A}_1^{2-}$ and $\mathbf{A}_2^{2-}$ reactions and the $\mathbf{A}_3^{2-}$ and $\mathbf{A}_4^{2-}$ reactions is the energy of the reactant hydrogen-bonded complex and the separated reactant state, respectively. The reference DFT energies are computed as adiabatic energies from the B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level. [b] Bold reactions are included in the parametrizations as a training set reaction. [c] Proton transfer is involved in the reaction. [d] Symmetric reactions. [e] Several points on the 1-D/2-D potential energy surfaces are included in the parametrizations to ensure the curvature of the potential energy surface near the transition states.

reactions in the reaction set consist of a water and methanol pair as nucleophile and leaving group. The errors of these two molecules in heat of formation coincide with the errors in the computed reaction energies from the AM1 and PM3 models. Even though the errors are smaller than other semiempirical methods tested, there exist similar errors from the AM1/d-PhoT model. Other defects of the AM1 and PM3 models are also evident. For example, there is a systematic increase of error in the calculated heats of formations as the charge on phosphorus containing compounds increase (Table

Table 11. Dissociative Monoanionic Reactions and Energies (kcal/mol) in Parametrizations and Their Test[a]

reaction[b]	type	DFT	error ($E_{calcd} - E_{DFT}$) AM1/d	AM1	PM3	label
H_2O + P(O)(O)(O)$^-$ ⇌ P(O)(O)(OH)(OH)$^-$	R	14.16	2.20	−1.03	−5.01	**D_1^-**[c]
	TS	24.15	10.93	5.69	−16.50	
	P	10.01	7.78	−29.59	−20.60	
CH_3OH + P(O)(O)(O)$^-$ ⇌ P(O)(O)(OH)(OCH_3)$^-$	R	13.17	0.74	−3.88	−2.19	**D_2^-**
	TS	22.52	6.92	12.85	−2.50	
	P	−10.71	7.73	−24.57	−11.49	
C_6H_5OH + P(O)(O)(O)$^-$ ⇌ P(O)(O)(OH)(OC_6H_5)$^-$	R	20.09	−2.29	−6.92	−4.17	**D_3^-**[c]
	TS	23.80	−7.17	−7.57	−23.52	
	P	−6.32	0.88	−27.39	−16.33	

[a] All the energies are relative to the reactant hydrogen-bonded complex. The reference DFT energies are computed as adiabatic energies from the B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level. [b] Bold reactions are included in the parametrizations as a training set reaction.

Table 12. Summary of Gas-Phase Reaction Energetics of Associative Reactions in the Training and Test Sets[a]

	neutral rxn			monoanionic rxn			dianionic rxn		
	AM1/d	AM1	PM3	AM1/d	AM1	PM3	AM1/d	AM1	PM3
				reaction energy[b]					
no. rxn		5			4			2	
MSE	2.07	−7.32	−10.78	0.84	−2.48	−4.94	−1.44	−9.00	−2.96
MUE	2.86	7.39	10.78	1.96	9.79	8.80	2.28	9.00	5.65
				activation energy					
no. TS		13			11			4	
MSE	0.76	3.28	−18.76	−1.12	0.24	−13.01	−3.63	−25.11	−33.38
MUE	3.61	6.62	18.76	2.64	12.69	16.36	3.63	25.11	33.38
				relative energy of intermediates					
No. int		8			7				
MSE	−1.06	−42.29	−26.61	−4.84	−42.03	−34.22			
MUE	2.36	42.29	26.61	4.84	42.03	34.22			

[a] All error values are computed as $E_{calcd} - E_{DFT}$, where the DFT values are B3LYP/6-311++G(3df,2p) single-point energies at a given B3LYP/6-31++G(d,p) optimized geometries. All energies are given in kilocalories per mole. [b] Symmetric reactions are ignored in computing reaction energies.

Table 13. Summary of Gas-Phase Reaction Energies (kcal/mol) of Three Dissociative Reactions in Trial Set and in the Training Set[a]

error of reaction	reaction energy AM1/d	AM1	PM3	activation energy AM1/d	AM1	PM3
MSE	5.25	−23.24	−12.35	3.56	3.66	−14.17
MUE	5.25	23.24	12.35	8.34	8.70	14.17

[a] All error values are computed as $E_{calcd} - E_{DFT}$, where DFT values are B3LYP/6-311++G(3df,2p) single-point energies at a given B3LYP/6-31++G(d,p) optimized geometries.

3). The errors in predicted PA values increase with the increase of charge on phosphorus compounds. This explains the relatively large error in the last reaction presented in Table 10.

The performance of the AM1/d-PhoT model on dissociative reactions is less impressive than for the associative reactions, which might be related to the slight unbalance of the heats of formation of metaphosphate and phosphate compounds. All the semiempirical methods, however, perform considerably poorly. Nonetheless, the AM1/d-PhoT model is more accurate than AM1 and PM3 by a factor of 2−3. The MSE values are 5.25 and 3.56 kcal/mol from the AM1/d-PhoT, −23.24 and 3.66 kcal/mol from the AM1, and −12.35 and −14.17 kcal/mol from the PM3 calculations for reaction energy and reaction barrier, respectively. If the dissociation reaction is the reverse reaction presented in Table 11, the reaction barriers from AM1/d-PhoT model are comparable to DFT dataset.

4.7. Free-Energy Simulations in Water. Although the AM1/d-PhoT model has been developed and tested for gas-phase processes, our goal is to study transphosphorylation reactions in aqueous solution and in enzymes or ribozymes. Thus, it is necessary to demonstrate an application of the AM1/d-PhoT model in condensed phase QM/MM simulations. To this end, molecular dynamics simulations employing umbrella sampling[123] have been carried out to compute potentials of mean force (PMF) for a model transphosphorylation reaction in water.

The transphosphorylation reaction model tested is the A_7^{-1} reaction presented in Table 9. The solute molecule, which has a total charge of −1 e, is embedded in a 40 × 40 × 40 Å^3 cubic box consisting of 2039 TIP3P water molecules.[124] In addition, we include one Na^+ ion to neutralize the charge of the system. The solutes were represented quantum-mechanically by the AM1/d-PhoT model, and the rest of the system constitutes the classical region. The geometry of water was held rigid by using the

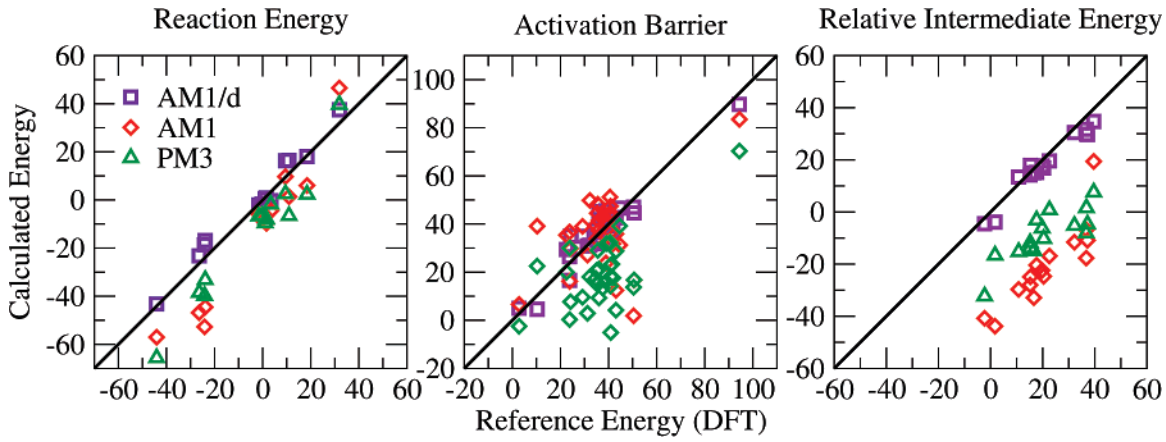


Figure 3. Scatter plots of gas-phase reaction energy (left), activation barrier (middle), and relative energy of intermediate (right). The reference reaction energies are from calculations at B3LYP/6-311++G(3df,2p) level, and symmetric reactions are not presented for reaction energy. The unit of energy is kilocalories per mole.

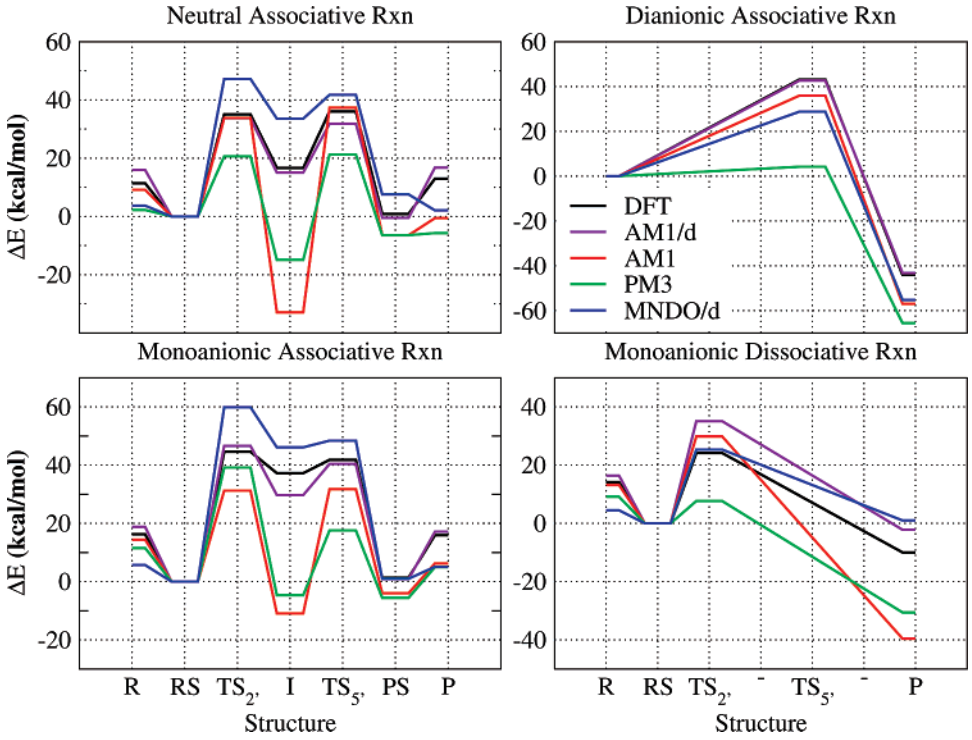


Figure 4. Comparison of performance of semiempirical methods on representative reactions. The plots are for the neutral associative reaction of A_1^0, the monoanionic associative reaction of A_1^-, the dianionic associative reaction of A_4^{2-}, and the dissociative reaction of D_1^-. The labels are shown in Table 8−11. The values for the MNDO/d model are obtained from relaxed potential energy surface as approximate energies, while the direct transition state optimizations are carried out for other semiempirical and DFT results.

SHAKE algorithm in all simulations.[125] To handle long-range electrostatic interactions, the QM/MM−Ewald scheme[65] was used along with the particle mesh Ewald (PME) method[126,127] for the electrostatic interactions between pure MM atom pairs and a 10.0 Å cutoff with potential shift in computing van der Waals interactions (refer ref 65 for further details on the simulations).

The reactions are divided into two steps: intramolecular nucleophilic attack and exocyclic cleavage. Each step involves a proton transfer, the first from the nucleophile to the nonbridging oxygen and the second from the nonbridging oxygen to the leaving group. The reaction is assumed to follow an in-line reaction pathway. Thus, each step requires a 2-dimensional umbrella sampling run to cover the entire free-energy surface. The computed results are shown in Figure 5. The contours are computed from 338 separate umbrella sampling windows, each consisting of over 30 ps averaging with at least 5 ps equilibration. Thus, the free-energy contour was constructed from a total of 12 ns of MD simulation. The computed barrier for the first step of the reaction ($TS_{2'}$) is about 32.0 kcal/mol and for the second step ($TS_{5'}$) is approximately 37.0 kcal/mol from combined QM/MM simulations (Figure 5). A phosphorane intermediate (I) has a PMF value of 29.0 kcal/mol, with barriers to collapse to reactants and products of 3.0 and 8.0 kcal/mol, respectively. The lifetime of such intermediates are im-

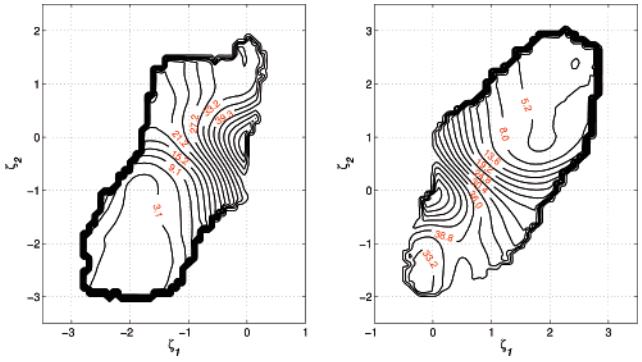


Figure 5. Computed potential of mean force (PMF) on the transphosphorylation reaction of A_7^- shown in Table 9. The reactions have two separate steps: intramolecular nucleophilic attack leading to formation of an endocyclic bond (left side) and exocyclic cleavage of the phosphorane intermediate leading to departure of the methanol leaving group (right side). The ζ_1 is defined as a distance difference in $R(O_B-P)$ and $R(P-O_A)$, where O_A and O_B are oxygens on the nucleophile and leaving group, respectively, presented in Scheme 2. The ζ_2 is the proton-transfer reaction coordinate also defined as distance differences between $R(O_A-H)$ and $R(H-O_R)$ (left plot) and between $R(O_R-H)$ and $R(H-O_B)$ (right plot), where H is the transferred proton and O_R is one of nonbridging oxygen on cyclic phosphate in Scheme 2. The unit of energy is kcal/mol, and that of distance is Å.

portant in modeling phosphoryl transfer reactions that might undergo alternate reactions after pseudorotation of the phosphorane (e.g., RNA migration).[25,47] The computed results are in reasonably good accord with the results from the B3LYP/6-311++G(3df,2p) calculations with implicit solvation modeled with the polarizable continuum model (PCM)[128,129] and a conductor-like screening model (COSMO).[130] These calculations yielded the barriers with respect to the unimolecular reactant for $TS_{2'}$ and $TS_{5'}$ of 28.2 and 38.0 kcal/mol, respectively, with PCM and 26.9 and 37.8 kcal/mol, respectively, with COSMO.[24] The DFT results identify a stable intermediate with barrier to collapse to reactants and products of 2.0 and 11.8 kcal/mol, respectively, with PCM and 2.1 and 13.0 kcal/mol, respectively, with COSMO. These values are in reasonable agreement with the QM/MM results, the main difference being that the $TS_{2'}$ transition state and I intermediate in the QM/MM calculations are shifted relative to the DFT values by approximately 3−5 kcal/mol.

5. Conclusion

A specific reaction parameter AM1/d Hamiltonian, AM1/d-PhoT, has been developed to model transphosphorylation reactions. The model is parametrized for H, O, and P atoms to reproduce high-level density-functional results from a recently constructed database of quantum calculations for RNA catalysis, including geometries and relative energies of minima, transition states and reactive intermediates, dipole moments, proton affinities, and other relevant properties. The model has been tested in the gas phase and in solution using a QM/MM potential, and demonstrated to provide overall very good accuracy with respect to the DFT results for a reduction of over 3−4 orders of magnitude in computational cost. The model offers a significant improvement over the MNDO/d, AM1, and PM3 methods for the transphosphorylation reactions. The ultimate goal of this work is to make strides toward the development of highly accurate semiempirical quantum models for biological reactions that can be used in conjunction with linear-scaling electronic structure and combined QM/MM methods to address complex problems of biocatalysis that simultaneously span large spatial domains and long time scales. The current work makes considerable progress in the development of fast, accurate quantum models for phosphoryl transfer reactions in solution and catalyzed by enzymes and ribozymes. Future work will involve critical comparison of AM1/d-PhoT with new SCC-DFTB models to assess the advantages and disadvantages of each method and facilitate the design of more robust quantum models for QM/MM calculations. It is the hope that through continued development and testing of semiempirical and SCC-DFTB models over a broad range of biological applications the next-generation of improved quantum methods for multiscale modeling of biocatalytic processes may emerge.

Acknowledgment. This work was partially supported by grants from the National Institutes of Health (GM62248 and GM46736), from the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under Cooperative Agreement DAAD19-01-2-0014, and from the Office of Naval Research (ONR) under Grant N00012-05-01-0538. Computational resources were provided by the Minnesota Supercomputing Institute and the Army High Performance Computing Research Center.

References

(1) Westheimer, F. H. *Science* **1987**, *235*, 1173−1178.

(2) Boyer, P. D.; Cross, R. L.; Momsen, W. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 2837−2839.

(3) Admiraal, S.; Herschlag, D. *Chem. Biol.* **1995**, *2*, 729−739.

(4) Ahn, N. *Chem. Rev.* **2001**, *101*, 2207−2208.

(5) Scott, W. G. *Curr. Opin. Struct. Biol.* **1998**, *8*, 720−726.

(6) Takagi, Y.; Ikeda, Y.; Taira, K. *Top. Curr. Chem.* **2004**, *232*, 213−251.

(7) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. *Science* **1996**, *274*, 2065−2069.

(8) Scott, W. G. *Q. Rev. Biophys.* **1999**, *32*, 241−294.

(9) Walter, N. G.; Burke, J. M. *Curr. Opin. Chem. Biol.* **1998**, *2*, 24−30.

(10) Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferré-D'Amaré, A. R. *Science* **2002**, *298*, 1421−1424.

(11) Shih, I.-h.; Been, M. D. *Biochemistry* **2000**, *39*, 9055−9066.

(12) Shih, I.-h.; Been, M. D. *EMBO J.* **2001**, *20*, 4884−4891.

(13) Shih, I.-h.; Been, M. D. *Annu. Rev. Biochem.* **2002**, *71*, 887−917.

(14) Perreault, D. M.; Anslyn, E. V. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 432−450.

(15) Zhou, D.-M.; Taira, K. *Chem. Rev.* **1998**, *98*, 991−1026.

(16) Oivanen, M.; Kuusela, S.; Lönnberg, H. *Chem. Rev.* **1998**, *98*, 961−990.

(17) Hengge, A. C. *Acc. Chem. Res.* **2002**, *35*, 105−112.

(18) Hengge, A. C.; Cleland, W. W. *J. Am. Chem. Soc.* **1990**, *112*, 7421−7422.

(19) Florián, J.; Åqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 11524−11525.

(20) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719−734.

(21) Åqvist, J.; Kolmodin, K.; Florian, J.; Warshel, A. *Chem. Biol.* **1999**, *6*, R71−R80.

(22) Liu, Y.; Gregersen, B. A.; Hengge, A.; York, D. M. *Biochemistry* **2006**, *45*, 10043−10053.

(23) Liu, Y.; Lopez, X.; York, D. M. *Chem. Commun.* **2005**, *31*, 3909−3911.

(24) Liu, Y.; Gregersen, B. A.; Lopez, X.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 19987−20003.

(25) López, C. S.; Faza, O. N.; de Lera, A. R.; York, D. M. *Chem.—Eur. J.* **2005**, *11*, 2081−2093.

(26) Lim, C.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 5872−5873.

(27) Uchimaru, T.; Tanabe, K.; Nishikawa, S.; Taira, K. *J. Am. Chem. Soc.* **1991**, *113*, 4351−4353.

(28) Lim, C.; Tole, P. *J. Phys. Chem.* **1992**, *96*, 5217−5219.

(29) Mercero, J. M.; Barrett, P.; Lam, C. W.; Fowler, J. E.; Ugalde, J. M.; Pedersen, L. G. *J. Comput. Chem.* **2000**, *21*, 43−51.

(30) Arantes, G. M.; Chaimovich, B. *J. Phys. Chem. A* **2005**, *109*, 5625−5635.

(31) Dejaegere, A.; Lim, C.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 4353−4355.

(32) Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **1993**, *115*, 5316−5317.

(33) Tole, P.; Lim, C. *J. Phys. Chem.* **1993**, *97*, 6212−6219.

(34) Tole, P.; Lim, C. *J. Am. Chem. Soc.* **1994**, *116*, 3922−3931.

(35) Chang, N.; Lim, C. *J. Phys. Chem. A* **1997**, *101*, 8706−8713.

(36) Florián, J.; Warshel, A. *J. Am. Chem. Soc.* **1997**, *119*, 5473−5474.

(37) Hu, C.-H.; Brinck, T. *J. Phys. Chem. A* **1999**, *103*, 5379−5386.

(38) Lopez, X.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2001**, *123*, 11755−11763.

(39) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 5010−5018.

(40) Lopez, X.; Dejaegere, A.; Leclerc, F.; York, D. M.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 11525−11539.

(41) Chen, X.; Zhan, C.-G. *J. Phys. Chem. A* **2004**, *108*, 6407−6413.

(42) Xu, D.; Guo, H.; Liu, Y.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 13827−13834.

(43) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11−27.

(44) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.

(45) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2003**, *125*, 7178−7179.

(46) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 7504−7513.

(47) López, C. S.; Faza, O. N.; Gregersen, B. A.; Lopez, X.; de Lera, A. R.; York, D. M. *Chem. Phys. Chem.* **2004**, *5*, 1045−1049.

(48) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.

(49) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *J. Mol. Biol.* **2003**, *327*, 549−560.

(50) Garcia-Viloca, M.; Poulsen, T. D.; Truhlar, D. G.; Gao, J. *Protein Sci.* **2004**, *13*, 2341−2354.

(51) Dewar, M. J.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.

(52) Dewar, M. J.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907−4917.

(53) Dewar, M. J. S.; Zoebisch, E.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(54) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.

(55) Khandogin, J.; Hu, A.; York, D. M. *J. Comput. Chem.* **2000**, *21*, 1562−1571.

(56) Khandogin, J.; York, D. M. *J. Phys. Chem. B* **2002**, *106*, 7693−7703.

(57) Khandogin, J.; Musier-Forsyth, K.; York, D. M. *J. Mol. Biol.* **2003**, *330*, 993−1004.

(58) Khandogin, J.; York, D. M. *Proteins* **2004**, *56*, 724−737.

(59) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(60) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(61) Gao, J. *Rev. Comput. Chem.* **1995**, *7*, 119−185.

(62) Gao, J. *Curr. Opin. Struc. Biol.* **2003**, *13*, 184−192.

(63) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391−404.

(64) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616−626.

(65) Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2−13.

(66) Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775−789.

(67) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495−506.

(68) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.

(69) Tubert-Brohman, I.; Guimaraes, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2003**, *25*, 138−150.

(70) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2005**, *1*, 817−823.

(71) Bernal-Uruchurtu, M.; Ruiz-López, M. *Chem. Phys. Lett.* **2000**, *330*, 118−124.

(72) Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C.; C. Millot, M. F. R.-L. *J. Comput. Chem.* **2000**, *21*, 572−581.

(73) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260−7268.

(74) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. *Phys. Status Solidi B* **2000**, *217*, 357−376.

(75) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569−585.

(76) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458−6469.

(77) Sattelmeyer, K. W.; Tubert-Brohman, I.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2006**, *2*, 413−419.

(78) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2005**, *123*, 164108.

(79) Alhambra, C.; Wu, L.; Zhang, Z.-Y.; Gao, J. *J. Am. Chem. Soc.* **1998**, *120*, 3858−3866.

(80) Arantes, G. M.; Loos, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 347−353.

(81) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231−236.

(82) Giese, T. J.; Gregersen, B. A.; Liu, Y.; Nam, K.; Mayaan, E.; Moser, A.; Range, K.; Nieto Faza, O.; Silva Lopez, C.; de Lera, A. R.; Schaftenaar, G.; Lopez, X.; Lee, T.; Karypis, G.; York, D. M. *J. Mol. Graphics Modell.* **2006**, *25*, 423−433.

(83) QCRNA, http://theory.chem.umn.edu/Database/QCRNA.

(84) Stewart, J. J. P. *Rev. Comput. Chem.* **1990**, *1*, 45−81.

(85) Thiel, W. In *New Methods in Computational Quantum Mechanics*; Advances in Chemical Physics 93; Prigogine, I., Rice, S. A., Eds.; John Wiley and Sons: New York, 1996; pp 703−757.

(86) Winget, P.; Selçuki, C.; Horn, A.; Martin, B.; Clark, T. *Theor. Chem. Acc.* **2003**, *110*, 254−266.

(87) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons: Chichester, U.K., 2004.

(88) Giese, T. J.; Sherer, E. C.; Cramer, C. J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 1275−1285.

(89) Csonka, G. I. *J. Comput. Chem.* **1993**, *14*, 895−898.

(90) Lopez, X.; York, D. M. *Theor. Chem. Acc.* **2003**, *109*, 149−159.

(91) Becke, A. D. *Phys. Rev. A.* **1988**, *38*, 3098−3100.

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

(93) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785−789.

(94) Bauernschmitt, R.; Ahlrichs, R. *J. Chem. Phys.* **1996**, *104*, 9047−9052.

(95) Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1977**, *66*, 3045−3050.

(96) Frisch, Æ.; Frisch, M. J. *Gaussian 98 User's Reference*, 2nd ed.; Gaussian, Inc.: Pittsburgh, PA, 1999.

(97) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; 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.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(98) Range, K.; McGrath, M. J.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 1654−1665.

(99) Mayaan, E.; Range, K.; York, D. M. *J. Biol. Inorg. Chem.* **2004**, *9*, 807−817.

(100) Range, K.; Riccardi, D.; Cui, Q.; Elstner, M.; York, D. M. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3070−3079.

(101) Alhambra, C.; Carochado, J.; Sánchez, M. L.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326−11340.

(102) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2001**, *114*, 9953−9958.

(103) Truhlar, D. G.; Gao, J.; Garcia-Viloca, M.; Alhambra, C.; Corchado, J.; Sanchez, M. L.; Poulsen, T. D. *Int. J. Quantum Chem.* **2004**, *100*, 1136−1152.

(104) Major, D.; York, D. M.; Gao, J. *J. Am. Chem. Soc.* **2005**, *127*, 16374−16375.

(105) Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.

(106) Coley, D. A. *An Introduction to Genetic Algorithms for Scientists and Engineers*; World Scientific: River Edge, NJ, 1999.

(107) Cundari, T. R.; Deng, J.; Fu, W. *Int. J. Quantum Chem.* **2000**, *77*, 421−432.

(108) Voityuk, A. A.; Rösch, N. *J. Phys. Chem. A* **2000**, *104*, 4089−4094.

(109) Hutter, M. C.; Reimers, J. R.; Hush, N. S. *J. Phys. Chem. B* **1998**, *102*, 8080−8090.

(110) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, W. P. *Numerical Recipes in Fortran*, 2nd ed.; Cambridge University Press, Cambridge, U.K., 1992.

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

(112) Giese, T. J.; York, D. M. In preparation.

(113) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822−2827.

(114) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532−6542.

(115) Gao, J.; Xia, X. *Science* **1992**, *258*, 631−635.

(116) Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386−395.

(117) Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467−6478.

(118) Gerratana, B.; Sowa, G. A.; Cleland, W. W. *J. Am. Chem. Soc.* **2000**, *122*, 12615−12621.

(119) Zhan, C.-G.; Landry, D. W. *J. Phys. Chem. A* **2001**, *105*, 1296−1301.

(120) König, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. A* **2006**, *110*, 548−563.

(121) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.

(122) Li, G.; Cui, Q. *J. Mol. Graphics Modell.* **2005**, *24*, 82−93.

(123) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.

(124) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(125) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(126) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(127) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Hsing, L.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(128) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.

(129) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43−54.

(130) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799−805.

(131) Hunter, E.; Lias, S. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413−457.

(132) Bartmess, J. E.; McIver, R. T. *Gas Phase Ion Chemistry*; Academic Press: New York, 1979; Vol. 2.

(133) Cox, J. O.; Pilcher, G. *Thermochemistry of Organic and Organometallic Compounds*; Academic Press: New York, 1970.

(134) Pilcher, G. In *The Chemistry of Organophosphorus Compounds*; Hartley, F. R., Ed.; Wiley: New York, 1990; Vol. 1, Chapter 5, pp 127−136.

(135) Pilcher, G.; Skinner, H. A. In *The Chemistry of the Metal−Carbon Bond*; Hartley, F. R., Patai, S., Eds.; Wiley: New York, 1982; Vol. 1, Chapter 2, pp 43−90.

(136) Linstrom, P., Mallard, W., Eds. *NIST Chemistry WebBook*; NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 2003; 20899; http://webbook.nist.gov (March 2003).

(137) Lide, D. R., Ed. *CRC Handbook of Chemistry and Physics*, 83rd ed.; CRC Press LLC: Boca Raton, FL, 2003.

(138) Dyke, T. R.; Mack, K. M.; Muenter, J. S. *J. Chem. Phys.* **1977**, *66*, 498−510.

(139) Feyereisen, M. W.; Feller, D.; Dixon, D. A. *J. Phys. Chem.* **1996**, *100*, 2993−2997.

CT6002466