Theor Chem Acc (2003) 109:149–159
DOI 10.1007/s00214-002-0422-2

Theoretical
Chemistry Accounts

Regular article

Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus

Xabier Lopez[1], **Darrin M. York**[2]

[1] Kimika Fakultatea, Euskal Herriko Unibertsitatea, P.K. 1072, 20080 Donostia, Euskadi, Spain
[2] Department of Chemistry, University of Minnesota, 207 Pleasant St. SE, Minneapolis, MN 55455-0431, USA

Received: 15 January 2002 / Accepted: 6 September 2002 / Published online: 28 March 2003

Abstract. This paper reports a new AM1/d model for phosphorus that can be used to model nucleophilic attack of phosphates relevant for biological phosphate hydrolysis reactions. The parameters were derived from a quantum dataset calculated with hybrid density-functional theory [B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p)] of phosphates and phosphoranes in various charge states, and on transitions states for nucleophilic attacks. A suite of non-linear optimization methods is outlined for semiempirical parameter development based on integrated evolutionary (genetic), Monte Carlo simulated annealing and direction set minimization algorithms. The performance of the new AM1/d model and the standard AM1 and MNDO/d models are compared with the density-functional results. The results demonstrate that the strategy of developing semiempirical parameters specific for biological reactions offers considerable promise for application to large-scale biological problems.

Keywords: Semiempirical model – AMI/d parameters for phosphorus – Nucleophilic reaction – Phosphate

1 Introduction

The reactivity and molecular properties of phosphate diesters, the chemical group that link nucleotides in DNA and RNA, play a fundamental role in biology. More specifically, hydrolysis of phosphate diesters has drawn considerable attention as the central reaction in the cleavage of the phosphate backbone in polynucleic acids [1, 2]. The reaction can be catalyzed by a variety of enzymes, and also by RNA itself (ribozymes), which has lead to the RNA world hypothesis [3], and with it, an increased effort to understand how the catalytic process works. A great deal of effort has been focused on the application of electronic structure methods to small phosphate hydrolysis model reactions [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17] (for a recent review, see that of Zhou et al. [18]). Nucleophilic attacks on model phosphates, such as ethylene phosphate and dimethyl phosphate, have been studied at different levels of electronic structure theory [18]. The computational cost associated with these theories quickly prohibits calculations of larger models that better mimic the enzyme and solvation environment.

An alternative to the computationally intensive ab initio methods is the use of semiempirical methods [19, 20, 21, 22, 23, 24]. The reduced computational cost would allow extension of theoretical studies to systems of much larger size. Moreover, semiempirical methods provide a practical avenue to incorporate dynamical effects into chemical reactions using classical molecular dynamics and hybrid quantum mechanical (QM)/molecular mechanical (MM) potentials. Currently, the major problem with semiempirical methods concerns the accuracy of the methods applied to reactions that fall outside the scope of their parameterization. In the case of phosphate ester hydrolysis, the performance of semiempirical models such as AM1 [21], PM3 [22] or MNDO/d [23, 24] is very poor, for different reasons. Part of the problems arise from the presence of hypervalent compounds (pentacovalent phosphoranes) that appear as reaction intermediates. The lack of *d* orbitals in the basis set for phosphorus in PM3 and AM1 make these methods not suitable to describe these reactions. On the other hand, the lack of Gaussian core–core functions in MNDO/d and the MNDO parameters results in a poor description when hydrogen-bonded compounds and proton transfer are involved. Finally, since almost no phosphate hydrolysis reaction data were directly considered in the parameterization of the

Contribution to the Proceedings of the Symposium on Combined QM/MM Methods at the 222nd National Meeting of the American Chemical Society, 2001

Correspondence to: D.M. York
e-mail: york@chem.umn.edu

conventional semiempirical methods, one cannot expect to obtain a high degree of accuracy. A possible way to overcome these difficulties is to reparameterize the semiempirical methods with respect to an ab initio dataset for the reactions of interest [5], specifically relevant to the problem. In this paper, efforts toward these goals are described through the use of a new AM1/d Hamiltonian, which includes *d* orbitals for phosphorus and AM1 parameters for first-row atoms. This Hamiltonian is then parameterized to accurately reproduce geometries, dipole moments and relative energies from a density-functional-based dataset for phosphate hydrolysis reactions.

2 Background/theory

2.1 MNDO, AM1 and PM3 Hamiltonians

The formalism for the electronic part of the MNDO, AM1 and PM3 Hamiltonians is based on the neglect of diatomic differential overlap approximation, and is identical for all the methods [6, 27]. The three Hamiltonians differ only in the way they treat core–core repulsions. In the MNDO method, the repulsion between two cores (A and B) is calculated as

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

where Z'_{A} and Z'_{B} are the effective nuclear charges (nuclear charge minus number of core electrons), $< s_{\mathrm{A}} s_{\mathrm{B}} | s_{\mathrm{A}} s_{\mathrm{B}} >$ is a Coulomb repulsion integral between an *s*-symmetry orbital centered on A and an *s*-symmetry orbital centered on B, and α_{A} and α_{B} are parameters in the exponential term that account for decreased screening of the nucleus by the electrons at small interatomic distances. For O–H and N–H bonds, a modified form of the screening term is used:

$$E_{\mathrm{N}}^{\mathrm{MNDO}}(\mathrm{A},\mathrm{H}) = Z'_{\mathrm{A}} Z'_{\mathrm{H}} \langle s_{\mathrm{A}} s_{\mathrm{H}} | s_{\mathrm{A}} s_{\mathrm{H}} \rangle \times \left(1 + \frac{\mathrm{e}^{-\alpha_{\mathrm{A}} \cdot R_{\mathrm{AH}}}}{R_{\mathrm{AH}}} + \mathrm{e}^{-\alpha_{\mathrm{H}} \cdot R_{\mathrm{AH}}}\right) \ . \quad (2)$$

In certain instances, such as noncovalent intermolecular interactions (e.g., hydrogen bonds), the MNDO model is problematic. The PM3 and AM1 models include a set of Gaussian core–core terms that alleviate excessive repulsion just outside bonding distances, and leads to significant improvement for intermolecular interactions. The modified core–core term takes the form

$$E_{\mathrm{N}}(\mathrm{A},\mathrm{B}) = E_{\mathrm{N}}^{\mathrm{MNDO}}(\mathrm{A},\mathrm{B}) + \frac{Z'_{\mathrm{A}} Z'_{\mathrm{B}}}{R_{\mathrm{AB}}} \left[\sum_k a_{k\mathrm{A}} \mathrm{e}^{-b_{k\mathrm{A}}(R_{\mathrm{AB}} - c_{k\mathrm{A}})^2} + \sum_k a_{k\mathrm{B}} \mathrm{e}^{-b_{k\mathrm{B}}(R_{\mathrm{AB}} - c_{k\mathrm{B}})^2} \right] . \quad (3)$$

The Gaussian core–core terms should be considered as empirical adjustments to the potential, largely devoid of rigorous physical meaning; nonetheless, their use allows critical improvements to be made for biological applications such as a much better description of hydrogen bonds and proton-transfer reactions. Alternatively one could substitute the Gaussian core–core terms with other functions that perhaps have more physical meaning [28, 29], or introduce other functional forms to the Hamiltonians [27]. This is a topic of future work.

2.2 MNDO/d and AM1/d Hamiltonians

In the case of biological phosphorus chemistry, the MNDO, AM1 and PM3 Hamiltonian models have further problems related to the fact that the atomic valence basis contains only *s*- and *p*-type orbitals. Recently, a formalism for extension of the MNDO method to include *d* orbitals has been presented (MNDO/d) [24]. Inclusion of *d* orbitals leads to an improved description of pentavalent phosphoranes, a key intermediate in the transphosphorylation and phosphate hydrolysis reactions described here. The MNDO/d method, however, retains the same problems as MNDO for modeling hydrogen bonds and proton-transfer reactions.

In this paper, an AM1/d Hamiltonian is developed that departs from the *d*-orbital formalism of MNDO/d and introduces the Gaussian core–core terms as in AM1 and PM3. Specific reaction parameters are then developed for this model that accurately reproduce geometries, dipole moments and relative energies obtained from a quantum database that include transition states as well as stable minimum-energy structures. It is the goal, then, to use this model to develop a new hybrid QM/MM potential and apply it to biological phosphate hydrolysis reactions catalyzed by enzymes and ribozymes.

3 Methods

This section describes the methods used to develop the semiempirical AM1/d model that is subsequently analyzed and tested. The first subsection describes the quantum dataset for phosphate hydrolysis reactions that was used as the reference data to fit the semiempirical AM1/d parameters for phosphorus (Table 1). The second subsection describes the details of the parameterization procedure itself.

3.1 Quantum dataset for phosphate hydrolysis reactions

The dataset consist of ten phosphate molecules and ten pentacovalent phosphoranes. The ten phosphate compounds (numbered **1–10**) are shown in Fig. 1. The phosphates are reactants and/or products of several nucleophilic attack reactions with CH_3O^-, OH_2 or CH_3OH as the nucleophile. Compounds **7–10** are all products of nucleophilic attacks on unprotonated (EP^-) and protonated (EPH) ethylene phosphate. These compounds are also the reactants of an intramolecular nucleophilic attack by a $–CH_2–OH$ group on phosphorus. These reactions have been frequently used as models for the transphosphorylation and hydrolysis reactions of RNA. Ten phosphoranes (numbered **11-20**) are shown in Fig. 2. These compounds represent key pentacovalent reaction intermediates of nucleophilic attacks on the phosphates of Fig. 1. Compound **11** is a transition state for the CH_3O^- attack on EP^-. This was the only transition state included in the dataset, for reasons specified later. An intermediate for this reaction was not considered, since dianionic phosphoranes are not stable in the gas phase

Table 1. AM1/d parameters for phosphorus. Standard notation for parameters taken from Refs. [23, 26]

a_1 (unitless)	−0.050000
b_1 (Å^{-2})	6.455034
c_1 (Å)	1.398583
a_2 (unitless)	−0.026188
b_2 (Å^{-2})	7.463832
c_2 (Å)	1.580645
a_3 (unitless)	−0.049608
b_3 (Å^{-2})	8.683773
c_3 (Å)	2.755376
U_{ss} (eV)	−42.205178
U_{pp} (eV)	−38.076209
U_{dd} (eV)	−21.320469
β_s (eV)	−9.728383
β_p (eV)	−8.687592
β_d (eV)	−1.937714
ζ_s(au)	1.444553
ζ_p(au)	1.145286
ζ_d (au)	1.191011
G_{sp}(eV)	8.176478
H_{sp}(eV)	2.217684
α(eV)	1.858232
ζ_s(eV)	2.121886
ζ_p(eV)	1.997680
ζ_d(eV)	1.200537
ρ_{core}(au)	1.340863

[30]. Compounds **12–14** are the intermediates generated in the attack of CH_3OH and OH_2 on EP^- or EPH. Compounds **15–20** are the pentacovalent intermediates that come from the OH_2 and CH_3OH attack on MP^{2-}, DMP^- and their protonated species. All these attacks are concerned with deprotonation of the nucleophile by proton transfer to one of the phosphoryl oxygens (Fig. 3), and the proton transfer will be primarily affected by oxygen and hydrogen parameters rather than those of phosphorus. Since the goal was to obtain phosphorus parameters while maintaining AM1 parameters for first-row atoms, it was decided not to include this kind of transition state in the dataset. The problem of proton transfer is a subject for future work.

All the compounds were fully optimized with Kohn–Sham density-functional-theory (DFT) methods, using the hybrid exchange functional of Becke [31] and correlation functional of Lee, Yang and Parr [32]. Geometry optimizations and Hessian evaluations were performed using a 6-31++G(d,p) basis set followed by single-point energy evaluation using a 6-311++G(3df,2p) basis set [33]. This protocol is conventionally designated as B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) and is applied throughout in this work; hereafter, it is simply referred to as B3LYP. All calculations were performed using the GAUSSIAN98 suite of programs [33]. Hessian calculations were used to verify the nature of the stationary points (minima and transition states) from the eigenvalues. In the case of **11(TS)**, only one negative eigenvalue was found with the mode corresponding to the approach of CH_3O^- to EP^- . These calculations provided a set of properties (geometry, dipole moment, ΔH_f^0, relative energies, etc.) used to construct a reference dataset to perform the semiempirical parameterization of phosphorus, while maintaining the original AM1 parameters for the first-row atoms.

In the following subsections, the specific energetic quantities that were included in the dataset are described in greater detail.

3.1.1 Heats of formation

This section concerns the generation of a set of gas-phase heats of formation based on the experimental value for dimethyl phosphate (DMP^-) and on density-functional calculations in the "gas phase". All reactions henceforth pertain to the gas phase unless it is specifically indicated otherwise.

The experimental ΔH_f^0 of DMP^- is -305.1 ± 5.1 kcal/mol [34]. To our knowledge, the experimental heats of formation are not known for any of the other phosphate compounds in this study. Consequently, a set of predicted values was constructed from the experimental value for DMP^- and on relative gas-phase enthalpies calculated with DFT (B3LYP).

Consider the reaction

$$DMP^- \overset{\Delta H_{DMP-H_2}}{\rightarrow} EP^- + H_2 \ . \tag{4}$$

The change in enthalpy, ΔH_{DMP-H_2}, for this reaction can be calculated with the B3LYP protocol described earlier. From ΔH_{DMP-H_2} and the experimental values of ΔH_f^0 (DMP^-) (-305.1 kcal/mol) and ΔH_f^0 (H_2) (0 kcal/mol), the value of ΔH_f^0 (EP^-) was estimated as

$$\Delta H_f^0(EP^-) = \Delta H_{DMP-H_2} - \Delta H_f^0(H_2) + \Delta H_f^0(DMP^-) \tag{5}$$

to give a value of -291.56 kcal/mol (Table 2). Similarly, the heat of formation of methyl phosphate, ΔH_f^0 (MPH^-), was calculated from the reaction

$$DMP^- \overset{\Delta H_{DMP-\frac{1}{2}C_2H_4}}{\rightarrow} MPH^- + \frac{1}{2}C_2H_4 \tag{6}$$

$$\Delta H_f^0(MPH^-) = \Delta H_{DMP-\frac{1}{2}C_2H_4} - \frac{1}{2}\Delta H_f^0(C_2H_4) + \Delta H_f^0(DMP^-) \tag{7}$$

with the experimental value of 12.5 kcal/mol [34] for $\Delta H_f^0(C_2H_4)$ to obtain a value of -312.5 for ΔH_f^0 (MPH^-) (Table 2).

3.1.2 Reaction energies

The primary focus of this work is to develop semiempirical parameters that reproduce the structure and energetics of phosphates and phosphoranes that model the stationary points along phosphate hydrolysis reaction paths. To do so, reaction energies of the following six nucleophilic attacks to phosphates were calculated at the B3LYP level described earlier.

$$EP^- + CH_3O^- \overset{\Delta E_1}{\leftrightarrow} \mathbf{7} \tag{8}$$

$$EP^- + CH_3OH \overset{\Delta E_2}{\leftrightarrow} \mathbf{8} \tag{9}$$

$$EP^- + OH_2 \overset{\Delta E_3}{\leftrightarrow} \mathbf{9} \tag{10}$$

$$MPH^- + CH_3OH \overset{\Delta E_4}{\leftrightarrow} DMP^- + OH_2 \tag{11}$$

$$EPH + OH_2 \overset{\Delta E_5}{\leftrightarrow} \mathbf{10} \tag{12}$$

$$MPH_2 + CH_3OH \overset{\Delta E_6}{\leftrightarrow} DMPH + OH_2 \tag{13}$$

In these reactions, compounds **7–10** correspond to the phosphates of Fig. 1.

Additionally, the relative energies of the ten pentacovalent phosphorane intermediates of Fig. 2 with respect to both reactants and products were entered into the $\chi^2(\boldsymbol{\lambda})$ function. This resulted in 16 distinct relative energies, the reactions for which are

$$EP^- + CH_3O^- \overset{\Delta E_{11-3}}{\leftarrow} \mathbf{11} \overset{\Delta E_{11-7}}{\rightarrow} \mathbf{7} \tag{14}$$

$$EP^- + CH_3OH \overset{\Delta E_{12-3}}{\leftarrow} \mathbf{12} \overset{\Delta E_{12-8}}{\rightarrow} \mathbf{8} \tag{15}$$

$$EP^- + OH_2 \overset{\Delta E_{13-3}}{\leftarrow} \mathbf{13} \overset{\Delta E_{13-9}}{\rightarrow} \mathbf{9} \tag{16}$$

$$EPH + OH_2 \overset{\Delta E_{14-4}}{\leftarrow} \mathbf{14} \overset{\Delta E_{14-10}}{\rightarrow} \mathbf{10} \tag{17}$$

$$\mathbf{15} \overset{\Delta E_{15-5}}{\rightarrow} MPH^- + OH_2 \tag{18}$$

$$\mathbf{16} \overset{\Delta E_{16-6}}{\rightarrow} MPH_2 + OH_2 \tag{19}$$

$$DMP^- + OH_2 \overset{\Delta E_{17-1}}{\leftarrow} \mathbf{17} \overset{\Delta E_{17-5}}{\rightarrow} MPH^- + CH_3OH \tag{20}$$

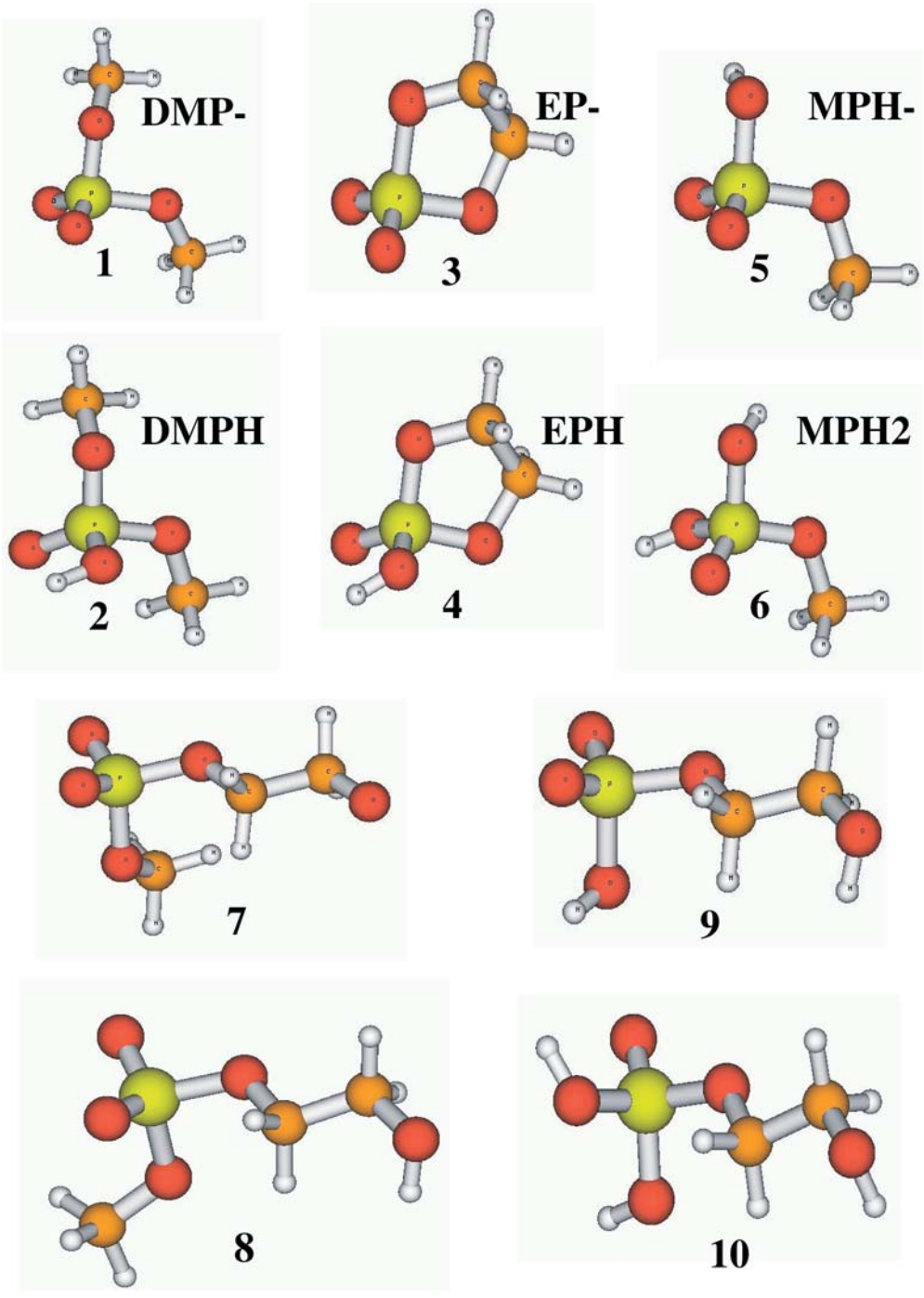

Fig. 1. B3LYP/6-31++G(d,p) structures of phosphates considered in the dataset: protonated and unprotonated ethylene phosphate (*EP*$^-$ and *EPH*) and dimethyl phosphate (*DMP*$^-$ and *DMPH*), and mono- and dianionic reaction products resulting from CH_3O^- and OH^- attack to EPH

$$\mathrm{DMPH} + \mathrm{OH_2} \overset{\Delta E_{18-2}}{\leftarrow} \mathbf{18} \overset{\Delta E_{18-6}}{\rightarrow} \mathrm{MPH_2} + \mathrm{CH_3OH} \quad (21)$$

$$\mathbf{19} \overset{\Delta E_{19-1}}{\rightarrow} \mathrm{DMP^-} + \mathrm{CH_3OH} \quad (22)$$

$$\mathbf{20} \overset{\Delta E_{20-2}}{\rightarrow} \mathrm{DMPH} + \mathrm{CH_3OH} \quad (23)$$

The B3LYP values for the relative energies corresponding to these reactions are listed in Table 3.

3.1.3 Relative proton affinities

A proper description of proton affinities is crucial to give an accurate description of the phosphate reactivity. However, the evaluation of the absolute gas-phase proton affinities by semiempirical methods requires the evaluation of the heat of formation for H^+, which can introduce significant errors and uncertainty. An alternative is to evaluate relative proton affinities with respect to an acid/base pair that can be considered as a reference state for the set of molecules of interest. In fact, one is generally only concerned with the relative protonation energies for biological reactions. For the purposes of this study, the reference state is chosen to be DMP^-, and hence relative proton affinities for phosphates and phosphoranes were calculated with respect to the DMPH/DMP^- acid/base pair according to the reaction

$$A\mathrm{H} + \mathrm{DMP^-} \overset{\Delta E}{\leftrightarrow} A^- + \mathrm{DMPH}\ , \quad (24)$$

where *A*H are the protonated phosphates DMPH, EPH, and MPH^- (Fig. 1), and the protonated phosphoranes **12**, **14**, **16**, **18** and **20** (Fig. 2).

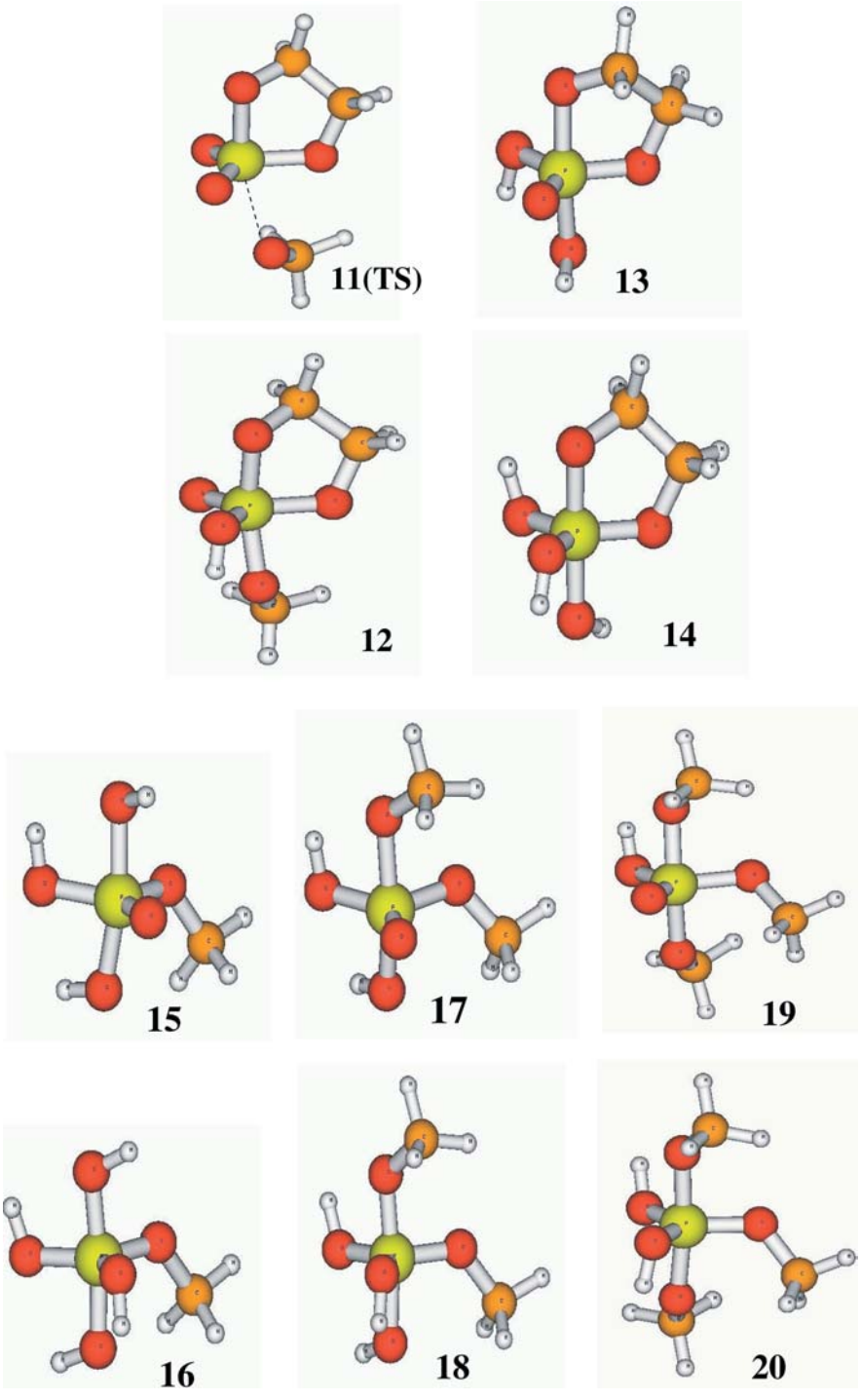

Fig. 2. B3LYP/6-31 + + G(d,p) structures of phosphoranes considered in the dataset with reference numbers used in the text

3.2 AM1/d parameterization of biological phosphorus

This section described the AM1/d parameterization procedure for biological phosphorus based on the B3LYP quantum dataset for phosphate hydrolysis calculated described in the previous section. The first step is to construct an appropriate $\chi^2(\boldsymbol{\lambda})$ merit function that measures the goodness of fit of a set of molecular properties calculated with a set of semiempirical AM1/d parameters $\boldsymbol{\lambda}$ with the corresponding reference B3LYP values. The second step is to use nonlinear optimization methods to find a suitable set of parameters by minimization of the $\chi^2(\boldsymbol{\lambda})$ merit function.

3.2.1 Construction of the $\chi^2(\boldsymbol{\lambda})$ merit function

The form of the $\chi^2(\boldsymbol{\lambda})$ merit function used in this work is given by

$$\chi^2(\boldsymbol{\lambda}) = \sum_{i}^{\text{mol}} \sum_{\alpha}^{\text{prop}} w_{i\alpha} \left[Y_{i\alpha}^{\text{AM1/d}}(\boldsymbol{\lambda}) - Y_{i\alpha}^{\text{DFT}} \right]^2 , \tag{25}$$

$$w_{i\alpha} = (\sigma_{i\alpha}^2)^{-1} / \sum_{j}^{\text{mol}} \sum_{\beta}^{\text{prop}} (\sigma_{j\beta}^2)^{-1} , \tag{26}$$

where $\boldsymbol{\lambda}$ represents a trial set of AM1/d parameters for P, $Y_{i\alpha}^{\mathrm{AM1/d}}(\boldsymbol{\lambda})$ is the value of the property α for molecule i calculated with the trial parameter set $\boldsymbol{\lambda}$, $Y_{i\alpha}^{DFT}$ is the corresponding reference value calculated with B3LYP, and $w_{i\alpha}$ is the associated least-squares weight in the fitting. The normalized weights $w_{i\alpha}$ are proportional to the inverse square of the $\sigma_{i\alpha}$ values in Eq. (25). The $\sigma_{i\alpha}$ values have the same units as the molecular property to which they are associated, and control the sensitivity of the merit function to deviations of that property from the reference value.

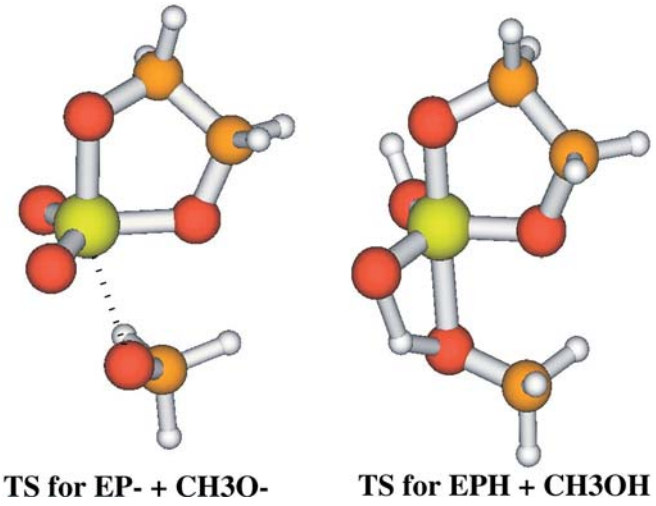

Fig. 3. Transition state B3LYP/6-31 + + G(d,p) structures for CH_3O^- attack to EP^- and CH_3OH attack to EPH

Table 2. Heats of formation (kcal/mol) for DMP^-, EP^- and MPH^-

Structure	B3LYP	AM1/d	AM1	MNDO/d
DMP^-	−305.10	−314.70	−295.45	−312.53
EP^-	−291.31	−303.90	−290.84	−305.18
MPH^-	−312.46	−317.33	−304.64	−317.53
< Err >		−9.02	5.98	−8.79
RMSD		9.56	7.18	9.54

For the semiempirical calculations a modified version of the MNDO97 [35] program was used. The properties considered include relative energies, ΔH_f^0 of DMP^- and EP^-, optimized bond lengths, angles, torsions, and dipole moments for neutral species. Each structure of the dataset was fully optimized at the semiempirical level for a given set of parameters before calculating these properties and constructing the $\chi^2(\boldsymbol{\lambda})$ function. This procedure is more time consuming, but proved to be absolutely critical for these more complicated compounds and reactions. Geometry optimization was not performed in previous work in the development of specific reaction parameter Hamiltonians, but instead single-point calculations at a stationary point were performed and the norm of the gradient in the $\chi^2(\boldsymbol{\lambda})$ was penalized. This procedure works well for very simple molecules and reactions with only small allowable variations in a few semiempirical parameters [25]. As described in more detail later, this is far from the situation in the present work. There are only two exceptions to the full geometry optimization procedure. Structure **11** was only partially optimized during the parameterization with the P–O bond frozen at 2.0 Å. This was the exact same procedure by which this structure was optimized at B3LYP. Moreover, to reduce the computational cost of trapping transition states, the P–OCH_3 bond in **TS1** was fixed to 2.486 Åand the rest of the geometrical variables were minimized. After the partial parameterizations, the gradient was calculated and was included as one of the properties to minimize in the parameterization.

3.2.2 Nonlinear optimization of the $\chi^2(\boldsymbol{\lambda})$ merit function

Semiempirical parameters were obtained by optimization of the $\chi^2(\boldsymbol{\lambda})$ merit function of Eq. (25) with respect to the set of semiempirical parameters $\boldsymbol{\lambda}$. For this purpose, a suite of integrated nonlinear optimization methods for semiempirical parameter development

Table 3. Relative energies (kcal/mol) of phosphoranes with respect to reactants and products

Phosphorane	Reaction	Ref.	B3LYP	AM1/d	AM1	MNDO/d
Dianionic phosphoranes						
11(TS)	$EP^- + CH_3O^- \leftrightarrow$ **TS**	ΔE_{TS-3}	87.17	92.40	83.56	84.20
	7 ↔ **TS**	ΔE_{TS-7}	43.07	43.54	26.55	28.87
< Err >				2.85	−10.07	−8.59
RMSD				3.71	11.96	10.26
Monoanionic phosphoranes						
12	$EP^- + CH_3OH \leftrightarrow$ **12**	ΔE_{12-3}	13.73	17.44	−17.60	34.88
	8 ↔ **12**	ΔE_{12-8}	23.01	25.32	−14.10	34.90
13	$EP^- + OH_2 \leftrightarrow$ **13**	ΔE_{13-3}	13.70	14.35	−25.92	33.02
	9 ↔ **13**	ΔE_{13-9}	23.64	23.13	−15.76	34.02
15	$MP^{2-} + OH_2 \leftrightarrow$ **15**	ΔE_{15-5}	23.72	18.01	−27.73	40.39
17	$DMP^- + OH_2 \leftrightarrow$ **17**	ΔE_{17-1}	23.45	21.07	−25.49	41.80
	$MP^{2-} + CH_3OH \leftrightarrow$ **17**	ΔE_{17-5}	23.41	21.49	−18.52	43.22
19	$DMP^- + CH_3OH \leftrightarrow$ **19**	ΔE_{19-1}	24.31	22.87	−15.62	41.73
< Err >				−0.66	−41.21	16.87
RMSD				2.82	41.65	17.25
Neutral phosphoranes						
14	$EPH + OH_2 \leftrightarrow$ **14**	ΔE_{14-4}	1.28	3.67	−39.40	22.22
	10 ↔ **14**	ΔE_{14-10}	7.28	13.09	−31.81	27.31
16	$MPH^- + OH_2 \leftrightarrow$ **16**	ΔE_{16-6}	8.36	6.24	−40.99	30.59
18	$DMPH + OH_2 \leftrightarrow$ **18**	ΔE_{18-2}	9.54	7.91	−38.73	31.41
	$MPH^- + CH_3OH \leftrightarrow$ **18**	ΔE_{18-6}	8.26	8.39	−31.98	33.07
20	$DMPH + CH_3OH \leftrightarrow$ **20**	ΔE_{20-2}	11.66	10.65	−27.79	33.35
< Err >				0.60	−42.85	21.93
RMSD				2.81	43.05	21.98
< Err >				−0.11	−35.49	14.34
Total RMSD				2.56	38.53	17.83

were used. These details of the integrated suite can be found elsewhere [42]; a brief overview of the algorithms are provided here.

The parameterization of an AM1/d Hamiltonian for phosphorus is far from a trivial modification of either the MNDO/d or AM1 models. The conventional AM1 and PM3 Hamiltonians attempt to capture the effect of *d*-orbitals in phosphorus using empirical Gaussian core–core functions. The MNDO/d method does not contain these functions, and hence parameterization of this Hamiltonian is somewhat more straight forward.[1] Inclusion of both Gaussian core–core functions and *d*-orbitals contains considerable overlap with respect to their effect on the models; however, the form of the conventional core–core functions does not allow one to simply zero out their effect by changing phosphorus parameters alone; e.g., even if the the core–core terms are set to zero on phosphorus, core–core interactions involving phosphorus are still present from other atoms (Eq. 1). Recall that the main goal of the present work is to provide an accurate semiempirical *d*-orbital phosphorus model for a specific set of biological reactions while maintaining the standard AM1 representation of first-row atoms. This was necessary to provide a significantly more accurate (although not completely satisfactory) description of critical noncovalent interactions such as hydrogen bonding. This is a first step to see what are the limits of the AM1/d Hamiltonian form, and to provide parameters that are ready to use in widely available semiempirical programs. Future work will focus on improving both the functional forms of the Hamiltonian, in addition to parameterization of more diverse biological molecules, complexes and reactions.

Three nonlinear optimization methods working in concert were applied in the present work: genetic algorithm, Monte Carlo simulated annealing, and direction set minimization methods. Genetic algorithms [36, 37] have been demonstrated elsewhere to be useful in semiempirical parameter optimization [5, 38, 39, 40]. The implementation of the genetic algorithm was based on the description by Goldberg [36], and used tournament selection and multidimensional phenotypic (parameter set) niching. Several genetic algorithm runs were performed with the number of generations ranging from 50 to 200 using a population of 64–128 members. The final population from the genetic algorithm optimization were then passed to a Monte Carlo simulated annealing procedure. The Monte Carlo procedure [41] used multidimensional simplex moves and variable exponentially decaying annealing schedules to explore the local region of parameter space around the final population provided by the genetic algorithm. The resulting parameters were then passed to a quadratically convergent direction set optimization method [41] to arrive at the final optimized parameter set. Recently, these methods have been extended and improved to make the parameterization more (although not completely) automated and robust.

4 Results and discussion

This section provides analysis and discussion of the results of the parameterization procedure described earlier. The new AM1/d parameterization for phosphorus is compared with the B3LYP reference dataset, and also to the performance of the standard AM1 and MNDO/d models. The section is divided into an analysis of geometries, dipole moments, and relative energies.

4.1 Geometries

A comparison of the AM1, MNDO/d and AM1/d P–O bond lengths and O–P–O and P–O–C bond angles for all the structures of the reference dataset except for **TS** versus the B3LYP geometries can be found in Figs. 4, 5 and 6. The mean absolute deviations for lengths and angles with respect to B3LYP values are shown in Table 4.

The P–O bond lengths can be divided in two categories, the distribution of bond lengths being bimodal. Phosphoryl oxygen bond lengths are centered around 1.5 Å, and are the shortest P–O bonds in the dataset since they have significant double-bond character.

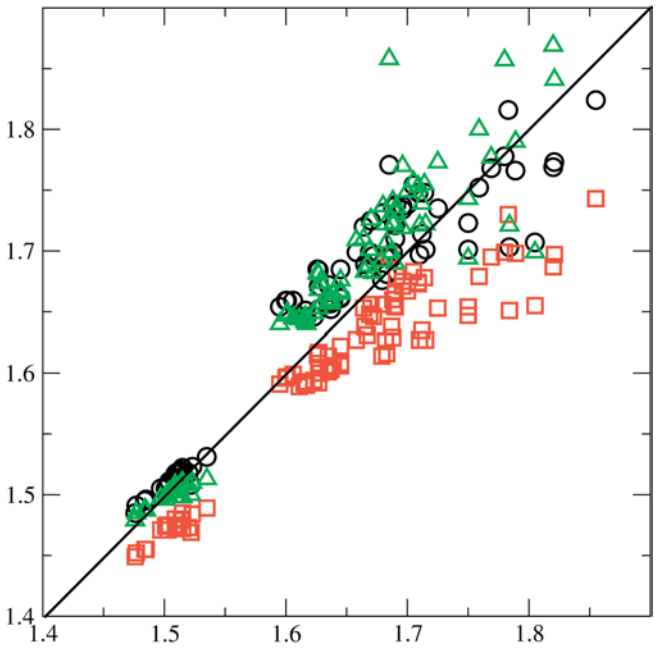

Fig. 4. Regression of AM1/d, AM1 and MNDO/d P–O bond lengths (Å) versus B3LYP/6-31 + + G(d,p) values

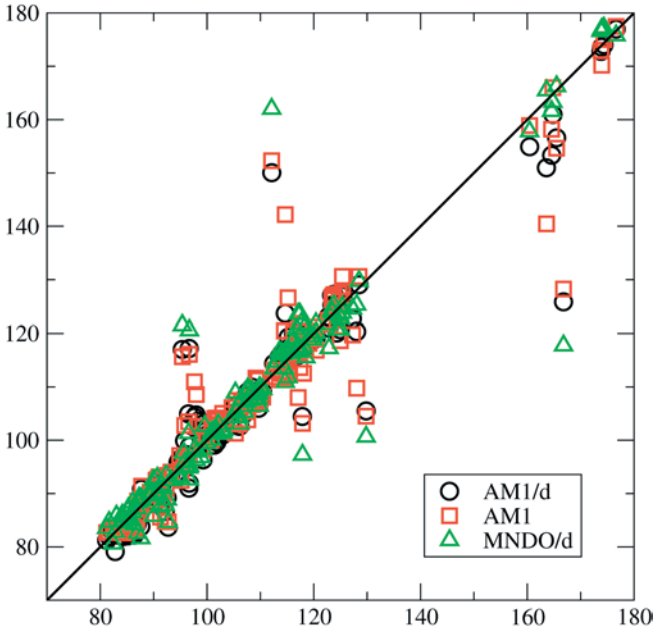

Fig. 5. Regression of AM1/d, AM1 and MNDO/d O–P–O bond angles (degrees) versus B3LYP/6-31 + + G(d,p) values

[1] Reparameterization of the MNDO/d Hamiltonian for P on the basis of the dataset in this work converges fairly quickly and performs only marginally better than the original MNDO/d (data not shown); however, this model does not work with the AM1 models for first-row atoms that describe hydrogen bonding much more accurately.

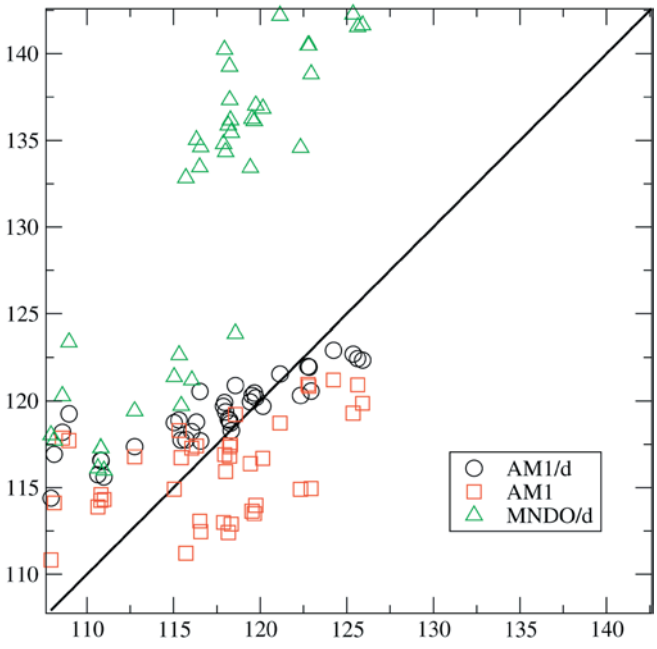

Fig. 6. Regression of AM1/d, AM1 and MNDO/d C–O–P bond angles (degrees) versus B3LYP/6-31 + + G(d,p) values

Table 4. Mean signed error (< *Err* >) and root-mean-square deviation (*RMSD*) (*italics*) of semiempirical phosphorus bond lengths and angles with respect to B3LYP

	AM1/d	AM1	MNDO/d
P–O length (Å)	0.016	−0.041	0.022
	0.034	*0.051*	*0.042*
O–P–O angle (degrees)	−0.4	−0.2	−0.1
	6.2	*7.0*	*7.2*
P–O–C angle (degrees)	1.8	−1.4	14.0
	3.8	*4.5*	*14.9*

Phosphodiester and P–OH bond lengths range from 1.6 to 1.9 Å, and consist of predominantly single P–O bonds. The variance of this category is larger than for the phosphoryl oxygens, reflecting slightly different bond lengths between tetravalent and pentavalent phosphorus single bonds. The deviations of the P–O bond lengths from the reference B3LYP values are 0.034, 0.051 and 0.042 Åfor AM1/d, AM1 and MNDO/d, respectively. The agreement between semiempirical and B3LYP values decreases as the P–O bond length increases. In general, AM1 predicts shorter P–O bond lengths than B3LYP; notice that the mean signed error (⟨Err⟩) is −0.041 Å. This tendency is corrected when *d* orbitals.

Regarding **TS**, the full optimization of the transition state, the distances between the incoming CH_3O^- nucleophile and EP^- show the largest deviation with respect to B3LYP. Thus the P–OCH_3 distance is 2.453 Å, whereas the AM1 one is 2.942 Å, MNDO/d is 2.343 Å and AM1/d is 2.279 Å.

Information about the root-mean-square deviation (RMSD) for angles can be found in Table 2 and Figs. 5 and 6. In general, there is good correlation between the B3LYP and semiempirical O–P–O angles, the RMSD being around 6–7° for the three methods. The C–O–P angle deviations can be found in Table 4 and Fig. 6. The RMSDs are similar for AM1 and AM1/d, 4.5° and 3.8°, respectively. The RMSD value with MNDO/d is considerably larger (14.9°), mainly owing to a systematic overestimate as seen by the mean signed error of 14.0°.

4.2 Dipoles

The calculated dipoles for neutral phosphate/phosphorane species are shown in Table 5. The RMSD obtained for AM1/d and AM1 is very similar, around 0.4 D, whereas it is largest for MNDO/d, 0.8 D. This is understandable since AM1 and AM1/d both share the same parameters for first-row atoms. If the dipole is computed at the B3LYP geometries instead of the semiempirical optimized geometries, the RMSD decreases to 0.1 D for AM1/d. This is suggestive that the deviation of the dipole moments with respect to the B3LYP values mainly arises from geometrical differences rather than purely electronic ones. Overall, the agreement of the AM1/d dipoles with the B3LYP results is comparable to that of the AM1 model, and represents a significant improvement over the MNDO/d model (Fig. 7).

4.3 Relative energies

This subsection discusses the relative energy results, and is divided into three parts: heats of formation, reaction energies, and relative protonation energies.

4.3.1 Heats of formation

The results for the heats of formation of EP^- , DMP^- and MP^{2-} are listed in Table 2. In general, the three methods show a similar RMSD for the heats of formation. AM1/d has an RMSD value of 9.56 kcal/mol, AM1 7.53 kcal/mol, and MNDO/d 9.54 kcal/mol. These results simply reflect that the heat of formation had a low weight in the $\chi^2(\lambda)$ merit function, since it was not the purpose of this work to get the absolute gas-phase heats of formation. Nonetheless, the results are presented for completeness. Future work might concentrate on improving these numbers, but in the end, the

Table 5. Dipole moments (debye) of the neutral species of the dataset

	B3LYP	AM1/d	AM1	MNDO/d
DMPH	1.236	0.894	1.078	1.966
EPH	4.101	3.617	4.028	3.518
MPH_2	0.984	1.112	0.901	1.755
12	2.734	2.478	2.256	2.367
14	2.658	1.675	1.821	1.016
16	1.374	1.472	1.390	1.001
18	1.091	1.301	1.104	1.025
20	1.692	1.497	1.386	0.905
< Err >		−0.228	−0.238	−0.290
RMSD		0.432	0.364	0.795

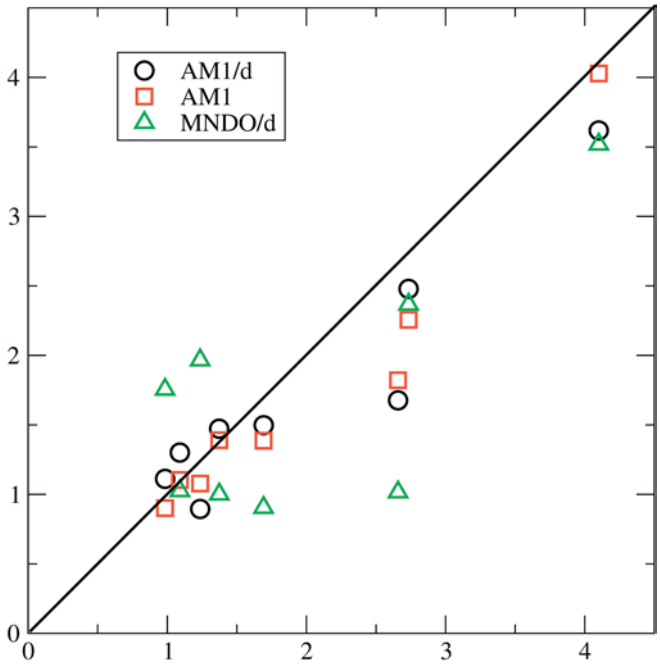

Fig. 7. Regression of AM1/d, AM1 and MNDO/d dipoles (debye) versus B3LYP/6-31 + + G(d,p) values

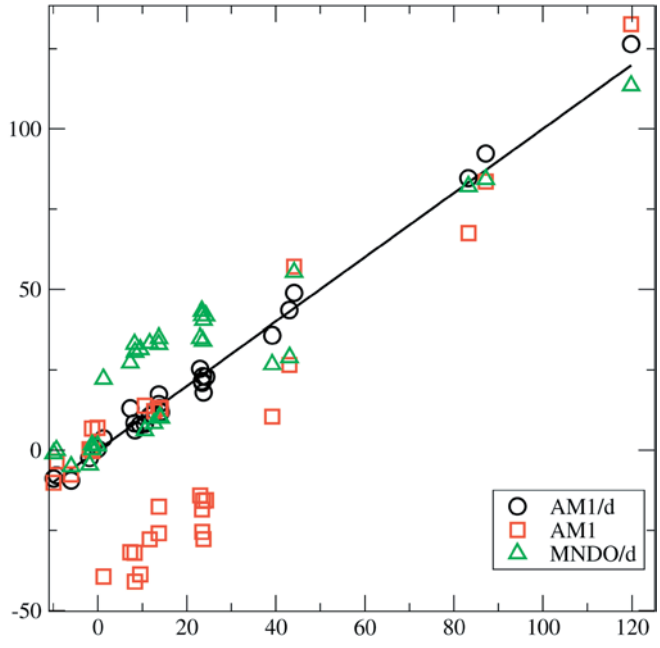

Fig. 8. Regression of AM1/d, AM1 and MNDO/d relative energies versus B3LYP values (see text). The relative energies shown are those in Tables 3, 6 and 7

most important quantities to reproduce are the relative energies most closely associated with the biological reactions. Increasing the weight for the heats of formation was observed to lead to poorer performance for important relative energies, which was the main focus of this work. Moreover, the estimated error in the heat of formation for DMP^- is ±5.1 kcal/mol [34], de-emphasizing the merit in converging the heat of formation to very high accuracy.

Table 6. Reaction energies (kcal/mol) for nucleophilic attacks to phosphates

Reaction	Ref.	B3LYP	AM1/d	AM1	MNDO/d
$EP^- + CH_3O^- \leftrightarrow$ **7**	ΔE_1	44.10	48.82	57.02	55.33
$EP^- + CH_3OH \leftrightarrow$ **8**	ΔE_2	−9.27	−7.88	−3.51	−0.02
$EP^- + OH_2 \leftrightarrow$ **9**	ΔE_3	−9.93	−8.78	−10.16	−1.00
$MP^{2-} + CH_3OH \leftrightarrow DMP^- + OH_2$	ΔE_4	−0.05	0.42	6.97	1.42
$EPH + OH_2 \leftrightarrow$ **10**	ΔE_5	−6.00	−9.42	−7.59	−5.09
$MPH^- + CH_3OH \leftrightarrow DMPH + OH_2$	ΔE_6	−1.28	0.47	6.75	1.66
< Err >			1.01	5.32	5.79
RMSD			2.60	7.26	7.10

4.3.2 Reaction energies

Energies for the six different nucleophilic reactions can be found in Table 6. The set comprises a variety of reactants and charge states. The RMSD for AM1/d (2.6 kcal/mol) is roughly a third of the RMSD for AM1 (7.3 kcal/mol) and MNDO/d (7.1 kcal/mol).

The reaction corresponding to Eq. (8) (ΔE_1 in Table 6) is an example of a dianionic mechanism, i.e., both the CH_3O^- nucleophile and the EP^- phosphate carry a net negative charge. The reaction is very endothermic, a characteristic reproduced by the three semiempirical methods, although they all overestimate the reaction energies. The error for AM1 is approximately 13 kcal/mol, for MNDO/d is approximately 10 kcal/mol, and for AM1/d is less than 5 kcal/mol. Also shown in Table 6 are examples of monoanionic reactions, where the attacking nucleophile is a neutral species (OH_2 or CH_3OH) and the phosphate is in its anionic state. They correspond to reactions 9–11 (ΔE_2, ΔE_3 and ΔE_4 in Table 6).

The first two reactions are exothermic at B3LYP by almost −10 kcal/mol, whereas the third one is almost thermoneutral. AM1 and MNDO/d perform poorly for these reactions, whereas AM1/d properly reproduces the exothermicity of ΔE_2 and ΔE_3 and the low reaction energy for ΔE_4. Finally, there are two neutral reactions, ΔE_5 and ΔE_6, which correspond to the attack of CH_3OH to MPH^- and of OH_2 to EPH. The first process is significantly more exothermic than the second, and this is well reproduced by all three semiempirical methods. However the three semiempirical methods give positive values for ΔE_6, whereas the B3LYP value is negative, namely −1.28 kcal/mol. AM1/d is the closest method of the three with a ΔE_6 of 0.47 kcal/mol.

Next, the results for the pentacovalent phosphorane structures are analyzed. The results for the pentacovalent phosphoranes considered here are summarized in Table 3. ΔE with respect to the reactants and products of the corresponding reactions is estimated and compared to the B3LYP values. The attack of CH_3O^- on EP^- is a dianionic reaction. This reaction and the closely related hydroxyl attack to EP^- have been studied pre-

viously with theoretical methods [5, 6, 7, 9, 30]. The high interest in these types of reactions resides in the fact that they have been taken as model systems to describe the hydrolysis (forward attack) and transphosphorylation (reverse attack) steps of RNA. It has long been recognized that the reactions show big barriers in the gas-phase, owing mainly to large repulsions for the approach of the two anions [5, 30]. The large negative charge of the system also precludes a stable phosphorane intermediate to be formed and is consistent with the facile cleavage of dianionic phosphoranes found experimentally [1] and the difficulty for ab initio calculations to trap a dianionic phosphorane intermediate. The B3LYP results reported here are in qualitative agreement with previous ab initio calculations, typically performed at lower basis set levels [5, 30]. The reaction has a high barrier (87.27 kcal/mol) and is quite endothermic (44.05 kcal/mol), which leads to a barrier for the reverse reaction (PROD → TS1) of 42.77 kcal/mol. Compounds **12**, **13**, **15**, **17** and **19** are monoanionic phosphoranes. The smaller net negative charge makes these species more stable than the dianionic phosphoranes. The values for ΔE are around 23 kcal/mol when the energies are calculated with respect to a reference state (reactant or product) that correspond to a separate phosphate and nucleophile (XOH; X = H, CH_3), and are around 13 kcal/mol when the reference is a phosphate isomeric with the phosphorane (i.e., same number and type of atoms as the phosphorane). Finally, compounds **14**, **16**, **18** and **20** are neutral phosphoranes with ΔE values that range between 1 and 10 kcal/mol.

An accurate semiempirical method for this set of reactions should reproduce the previously described trends for the ΔE of pentacovalent phosphoranes with respect to tetravalent phosphates. It would also be important to account for the destabilization of phosphoranes as the amount of negative charge of the system increases. The semiempirical results are depicted in Table 3. A first glance to this table shows that AM1 describes poorly the pentacovalent phosphoranes. The AM1 RMSD for ΔE is 38.27 kcal/mol. The RMSD for MNDO/d (17.44 kcal/mol) is considerably lower. The new AM1/d parameterization is able to reduce this error to 2.81 kcal/mol.

The origin of the failure of AM1 for biological phosphorus is rooted in the tendency to overstabilize the pentacovalent phosphoranes. In fact, a full optimization of a dianionic intermediate (not shown) leads to a stable pentacovalent structure with significant barriers toward bond breaking of either phosphodiester axial bond, in total disagreement with both B3LYP and experimental data. This behavior is corrected by MNDO/d and AM1/d; however, the former underestimates significantly the barrier for the transphosphorylation reaction in the dianionic mechanism; i.e., ΔE_{TS-7} in Table 3. The overstabilization of phosphoranes by AM1 is even more extreme for monoanionic and neutral phosphoranes. In these cases, AM1 gives very negative values for the corresponding ΔE, in stark contrast to the B3LYP results. The RMSD for the relative energies of the monoanionic phosphoranes is 41.65 kcal/mol, and for neutral phosphoranes is 43.05 kcal/mol. MNDO/d, on the other hand, shows significantly improved energies for the phosphorane structures and it gives a more balanced description between pentavalent and tetravalent phosphoranes/phosphates. Nonetheless, the tendency is for MNDO/d to slightly destabilize the phosphoranes relative to the B3LYP values, leading to a RMSD for the relative energies of 17.25 kcal/mol for monoanionic phosphoranes and 17.70 kcal/mol for neutral phosphoranes. One of the major successes of the new AM1/d model is to drastically reduce the RMSD of the relative energies of the phosphoranes relative to the B3LYP values to 2.82 kcal/mol for monoanionic phosphoranes and 2.18 kcal/mol for the neutral phosphoranes. This is a key step toward modeling more accurately reaction barriers and energies in hybrid QM/MM applications of phosphate hydrolysis in enzymes or ribozymes.

4.3.3 Relative proton affinities

Results for the relative proton affinities with respect to the DMPH/DMP^- acid/base pair can be found in Table 7. A positive AH energy value indicates that DMPH is a stronger "acid" in the gas phase than AH (i.e., DMP^- has a smaller gas-phase proton affinity than A^-). In the case of phosphates, the relative proton affinities with respect to DMP^- are small. However, for phosphoranes these numbers are considerably larger, of the order of 10–15 kcal/mol. The proton affinities of the phosphates and phosphoranes are key quantities to accurately reproduce for biological applications since protonation/deprotonation events are frequently steps in the catalytic mechanism of phosphate hydrolysis. AM1/d, AM1 and MNDO/d describe these quantities reasonably well, with RMSD values relative to B3LYP of 1.49, 1.76 and 3.56 kcal/mol, respectively.

5 Conclusions

The aim of this paper is to demonstrate that an AM1/d semiempirical Hamiltonian with proper parameterization can be a powerful method for QM/MM calculations on large systems where accuracy is needed at a low computational cost. Inclusion of d orbitals is essential for a balanced treatment of normal valent and hypervalent compounds. AM1 is unacceptable for the description of reactions in which pentacovalent intermediates are formed. MNDO/d provides a significant improvement in the energies of phosphoranes, but still

Table 7. Protonation energies defined with respect to the DMPH/DMP^- acid/base pair (kcal/mol)

Phosphorane	B3LYP	AM1/d	AM1	MNDO/d
EPH /EP^-	−1.85	−2.42	0.38	−4.60
MPH^-/ MP^{2-}	−1.23	0.05	−0.22	0.24
13/14	10.58	8.26	13.86	6.20
15/16	14.12	11.82	13.04	10.04
17/18	13.91	13.16	13.24	10.39
19/20	12.64	12.22	12.17	8.38
< Err >		−0.85	0.72	−2.92
RMSD		1.49	1.76	3.56

suffers from the MNDO parameterization of first-row atoms that describe very poorly hydrogen bonds and proton-transfer processes (a problem that is largely overcome by AM1). AM1/d gives overall good agreement with the B3LYP calculations for pentavalent and tetravalent phosphorane and phosphate compounds in the three charge states (dianionic, monoanionic and neutral) considered in this work, while maintaining the AM1 description of first-row atoms. The AM1/d model for phosphorus thus provides considerable improvement relative to either AM1 or MNDO/d for the specific biological reactions considered here. Further improvement of the model for biological reactions could be realized through the design of new semiempirical Hamiltonian forms and careful reparameterization for a specific set of relevant chemical reactions obtained from a quantum database. The work presented here demonstrates a first step to building extremely fast, accurate quantum models for QM/MM applications of biological reactions.

Acknowledgements. D.M.Y. is grateful for financial support provided by the National Institutes of Health (grant 1R01-GM62248-01A1) and the Donors of The Petroleum Research Fund, administered by the American Chemical Society, and the Minnesota Supercomputing Institute through a 6-month research scholar award (X.L.). Computational resources were provided by the Minnesota Supercomputing Institute.

References

1. Chang N-Y, Lim C (1998) J Am Chem Soc 120: 2156–2167
2. Blaskó A, Bruice TC (1999) Acc Chem Res 32: 475–484
3. Chang NY, Lim C (1997) J Phys Chem A 101: 8706–8713
4. Lim C, Karplus M (1990) J Am Chem Soc 112: 5872–5873
5. Thiel W, Voityuk AA (1996) J Phys Chem 100: 616–626
6. Lim C, Tole P (1992) J Phys Chem 96: 5217–5219
7. Dejaegere A, Liang XL, Karplus M (1994) J Chem Soc Faraday Trans 90: 1763–1767
8. Uchimaru T, Tanabe K, Nishikawa S, Taira K (1991) J Am Chem Soc 113: 4351–4353
9. Voityuk AA, Ro1sch N (2000) J Phys Chem A 104: 4089–4094
10. Rossi I, Truhlar DG (1995) Chem Phys Lett 233: 231–236
11. Wladkowski BD, Krauss M, Stevens WJ (1995) J Am Chem Soc 117: 10537–10545
12. Florián J, Warshel A (1997) J Am Chem Soc 119: 5473–5474
13. Florián J, Warshel A (1998) J Phys Chem B 102: 719–734
14. Yarus M (1999) Curr Opin Chem Biol 3: 260–267
15. McCarthy WJ, Smith DMA, Adamowicz L, Saint-Martin H, Ortega-Blake L (1998) J Am Chem Soc 120: 6113–6120
16. Hu C-H, Brinck T (1999) J Phys Chem A 103: 5379–5386
17. Mercero JM, Barrett P, Lam CW, Fowler JE, Ugalde JM, Pedersen LG (2000) J Comput Chem 21: 43–51
18. Zhou D-M, Taira K (1998) Chem Rev 98: 991–1026
19. Dewar MJ, Thiel W (1977) J Am Chem Soc 99: 4899–4907
20. Dewar MJ, Thiel W (1977) J Am Chem Soc 99: 4907–4917
21. Dewar MJS, Zoebisch E, Healy EF, Stewart JJP (1985) J Am Chem Soc 107: 3902–3909
22. Saint-Martin H, Ruiz-Vicent LE, Ramírez-Solís A, Ortega-Blake I (1996) J Am Chem Soc 118: 12167–12173
23. Thiel W (1998) Program MNDO97. University of Zurich
24. Thiel W, Voityuk AA (1992) Theor Chim Acta 81: 391–404
25. Press WH, Teukolsky SA, Vetterling WT, Flannery WP (1992) Numerical Recipes in Fortran, 2nd edn. Cambridge University Press, Cambridge
26. Stewart JJP (1989) J Comput Chem 10: 209–220
27. Stewart JJP (1990) Rev Comput Chem 1: 45–81
28. Bernal-Uruchurtu M, Ruiz-López M (2000) Chem Phys Lett 330: 118–124
29. Bernal-Uruchurtu MI, Martins-Costa MTC, Millot MFR-L (2000) J Comput Chem 21: 572–581
30. Dejaegere A, Lim C, Karplus M (1991) J Am Chem Soc 113: 4353–4355
31. Becke AD (1993) J Chem Phys 98: 5648–5652
32. Lee C, Yang W, Parr RG (1988) Phys Rev B 37: 785–789
33. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98, revision A.9. Gaussian, Pittsburgh, PA
34. NIST chemistry WebBook (July 2001) http://webbook.nist.gov/
35. Thiel W (1996) Adv Chem Phys 93: 703–757
36. Goldberg D (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Reading, MA
37. Coley DA (1999) An introduction to genetic algorithms for scientists and engineers. World Scientific, Singapore
38. Cundari TR, Deng J, Fu W (2000) Int J Quantum Chem 77: 421–432
39. Uebayasi M, Uchimaru T, Koguma T, Sawata S, Samayama T, Taira K (1994) J Org Chem 59: 7414–7420
40. Hutter MC, Reimers JR, Hush NS (1998) J Phys Chem B 102: 8080–8090
41. Perreault DM, Anslyn EV (1997) Angew Chem Int Ed Engl 36: 432–450
42. York DM, in preparation