Theor Chem Acc (2009) 124:197–215
DOI 10.1007/s00214-009-0600-6

REGULAR ARTICLE

Benchmark calculations on models of the phosphoryl transfer reaction catalyzed by protein kinase A

Manuel Montenegro · Mireia Garcia-Viloca · Àngels González-Lafont · José M. Lluch

Received: 3 June 2009 / Accepted: 11 June 2009 / Published online: 4 July 2009

Abstract Protein phosphorylation has been proved to be of great importance in many stages of cell life. In the last few years, its reaction mechanism has been extensively studied. In this work we present the analysis of the performances of several computational methods with different computational costs (from multilevel to semiempirical) to point out the best method to be used at each level in the study of phosphoryl transfer. Finally, we center on the semiempirical methods, and mainly on the AM1/d Hamiltonian with different sets of parameters, which will permit hybrid quantum mechanics/molecular mechanics (QM/MM) free energy calculations on big models at an acceptable computational cost. We have used quite a large set of molecules and model reactions to test the computational methods, reproducing all the chemical steps involved in the mainly accepted reaction pathways for the protein phosphorylation. In the end, we also present the results for an enlarged model, cut out from an entire biological model: we compare the 2-D PES at the B3LYP and AM1/d levels with the purpose of obtaining a correction for the semiempirical method. The AM1/d-PhoT semiempirical parameterization corrected using single-point energy calculations at the B3LYP/MG3S level seems to be suitable to carry out reliable QM/MM calculations of the complete biological system.

Keywords Semiempirical methods · Protein kinase A · Phosphoryl transfer reaction · Dual-level correction

1 Introduction

Protein phosphorylation is the main signaling mechanism to control metabolic pathways in the cell. This process consists in the transfer of the γ-phosphoryl group of ATP to a specific protein's serine, threonine or tyrosine residue, and results in the activation of the substrate protein for a specific function. This cell's life critical process is regulated by protein kinases. As a consequence, this important family of enzymes is nowadays the subject of an increasing number of studies. However, at the end of the 1990s the research in protein kinases was mainly centered on experimental work and the number of theoretical studies was limited, probably due to the hypervalency of phosphorous, which requires to be represented with d-orbitals. In spite of that, the advances in quantum theory and computational science have encouraged theoretical investigations of protein kinases [1–16], most of them centered on protein kinase A, PKA, which is the best characterized member of the protein kinase family.

Different mechanisms have been proposed for this enzyme on the basis of kinetic and structural data and different theoretical groups have studied them using models and methods of increasing complexity and accuracy. The first theoretical study used gas phase models of the enzymatic reaction and semiempirical methods [5, 17].

Electronic supplementary material The online version of this article (doi:10.1007/s00214-009-0600-6) contains supplementary material, which is available to authorized users.

M. Montenegro · M. Garcia-Viloca (✉) · À. González-Lafont · J. M. Lluch
Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
e-mail: mireia@bioinf.uab.es

M. Garcia-Viloca · À. González-Lafont · J. M. Lluch
Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Although they provided relevant information to know the chemical system under study, they were not complete microscopic models of the biological system. Furthermore, the inadequacy of semiempirical methods that lack d-orbitals to study phosphoryl transfer reactions was demonstrated by Sheppard et al. [2]. More recent calculations on the PKA system used quantum mechanics (QM) cluster models [6, 7, 9] or complete microscopic hybrid quantum mechanics/molecular mechanics (QM/MM) models [8, 14], where the QM system is represented by a density functional theory (DFT) method. Free energy calculations have been finally carried out by combining a DFT/MM reaction path with appropriate MM molecular dynamics simulations along that reaction path [14]. However, the huge computational cost of direct free energy calculation on a true DFT/MM potential energy surface does not allow to carry out simulations at the nanosecond time scale to assure the convergence of the resulting free energy profiles. On the other hand, different semiempirical Hamiltonians containing d-orbitals for P and Mg have been developed in the last few years, which, well calibrated, can provide accurate enough QM/MM potentials to explore the different mechanisms proposed for PKA [18–22]. In this regard, our strategy is to test these Hamiltonians in front of highly accurate experimental or theoretical data. The first step in the calibration is the choice of the database, that is, a set of molecules and reactions directly related to the chemical reaction in the biological system for which high-quality data, either from experiment or high-level ab initio calculations, exists. Recently, a database has been published by York et al. [23] for phosphoryl transfer reactions in general. This gave us an opportunity to complete this database to be specifically representative of the PKA system, by including molecules containing magnesium and adding three model reactions of the different mechanisms that have been proposed in the literature for the reaction catalyzed by PKA [24–26]. The set of molecules chosen as the database in this study ranges from 2 to 22 atoms, starting from the hydroxyl anion to a methylphosphate–magnesium–water complex. In addition, we have the aim of analyzing the performance of the semiempirical Hamiltonians to reproduce the ab initio or DFT structures and energetics along the three following model reactions that represent the two possible reaction mechanisms proposed for the PKA reaction. As will be explained in more detail in Sect. 3, Eq. 1 models the associative mechanism (Scheme A in Fig. 1), whereas Eqs. 2 and 3 model the dissociative mechanism (Scheme B in Fig. 1).

$$[CH_3PO_4]^{--} + H_2O \rightarrow [CH_3O]^{-} + [H_2PO_4]^{-} \quad (1)$$

$$CH_3CH_2OH + CH_3COO^{-} \rightarrow CH_3CH_2O^{-} + CH_3COOH \quad (2)$$

$$[CH_3PO_3O]^{--} + H_2O + Mg^{++}[H_2O]_4 \rightarrow [CH_3PO_3OH]^{-} + H_2O + Mg^{++}[H_2O]_3[HO]^{-} \rightarrow CH_3OH + [H_2PO_4]^{-} + Mg^{++}[H_2O]_3[HO]^{-} \quad (3)$$

The size of the largest models does not allow to compare the semiempirical Hamiltonians directly to the ab initio calculations, but it is possible to do so through the use of a medium level quantum method, such as a DFT method, whose accuracy has in turn been tested against a reference ab initio method.

In summary, the aim of this study is to respond to the following key questions:

(1) Which is the ab initio method that better reproduces the reference data for the database chosen? (2) Which is the best DFT method in comparison to the method chosen in (1)? (3) Which is the best semiempirical method in comparison to the medium level tested in (2)?

The answers to these three questions might be relevant when choosing the QM level in any future QM/MM study of PKA.

2 Methods

We considered three multilevel methods (CBS-QB3 [27, 28], G3S/3 [29, 30], and MCG3/3 [31, 30]), three hybrid density functional theory methods (B3LYP [32, 33], MPWB1K, and MPW1B95 [34]) with different basis sets, and eight semiempirical methods and parameterizations, some of them including d-orbitals.

2.1 Multilevel methods

All the multilevel methods combine the results of several electronic structure calculations at different levels and/or different one electron basis sets, usually with empirical parameters, in order to extrapolate to molecular energies of higher accuracy than the most accurate component calculation. These methods have been extensively tested and shown to be generally reliable; however, they are too computationally intensive to be applied to large biological model systems.

A series of multilevel methods, referred to as complete basis set (CBS) methods, calculate the total energy by a combination of additive energy terms [27, 28]. The central idea in the CBS methods is an extrapolation procedure to determine second-order correlation energy in the limit of a complete basis set. Several empirical corrections are also added like in the Gn procedures. In our application of the CBS-QB3 method, selected for this study, we have combined the CBS-Q energy with B3LYP/MG3S [32, 33, 35] geometries and non-scaled frequencies.

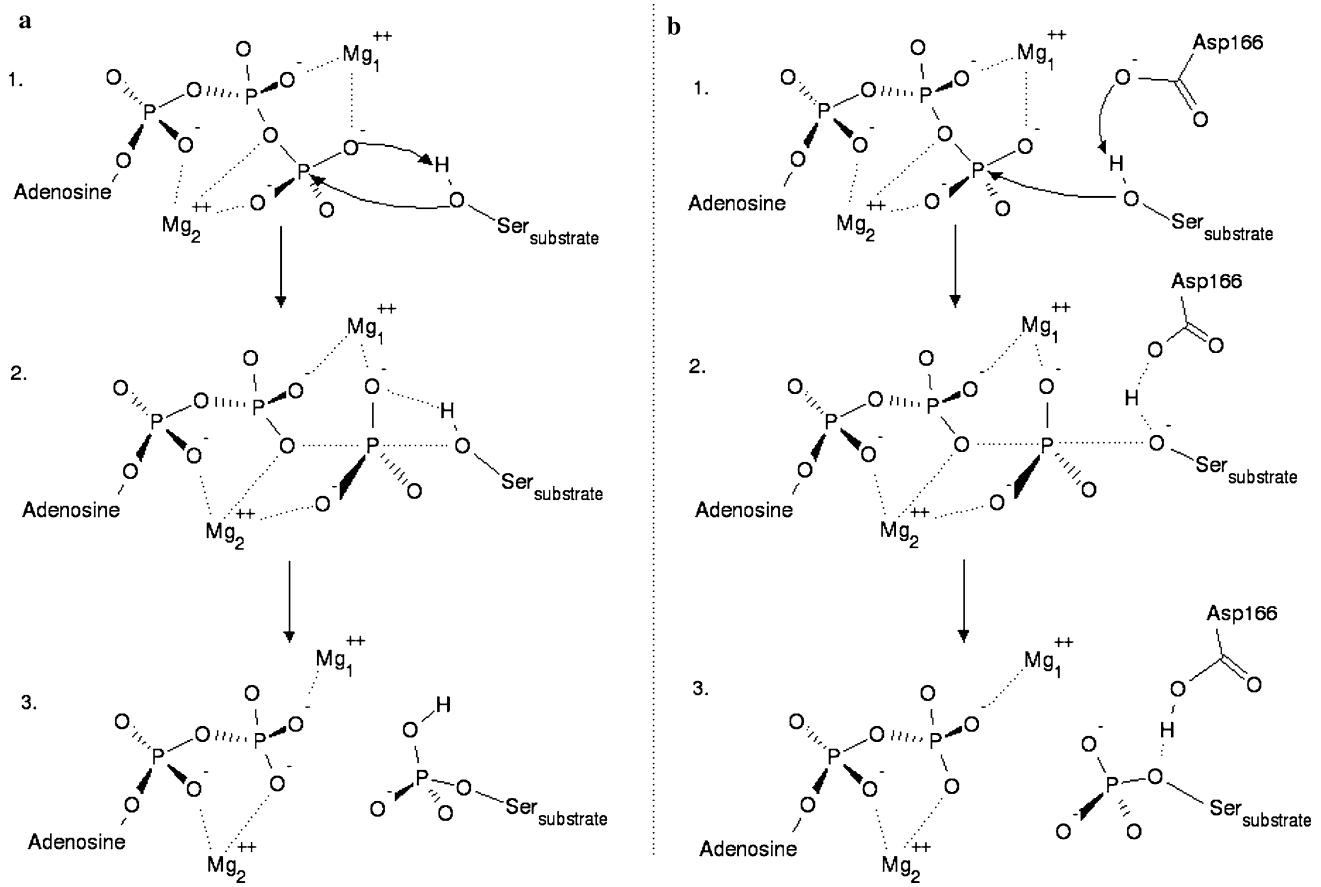

Fig. 1 Reaction scheme for the phosphoryl transfer step: **a** is the associative mechanism which consists in the phosphoryl- and proton-transfers between the substrate and the ATP molecule. **b** represents the dissociative reaction pathway; the proton of Ser passes to an oxygen of Asp166 during the nucleophilic displacement

Another approach for multilevel methods that has been proposed is a scaling of the calculated energy using multiplicative parameters determined by fitting to experimental data. The G3S/3 and MCG3/3 methods belong to that second group, also known as multicoefficient correlation methods (MCCM) [30]. These two MCCM methods have been proved to represent a very good compromise of accuracy, cost, and ease of use for practical calculations in thermochemical kinetics. This is a consequence of both approaches using the scaling procedure instead of the additive higher-level correction energy term of the G3 theory. In the case of MCG3/3 technique, the good performance is achieved with a lower computational cost by modifications in the way higher-order correlation energies are divided and in the basis set extension of the multilevel expression. Also for the MCG3/3 method and the G3S/3 one we have used the optimized structure and the non-scaled frequencies at the B3LYP/MG3S level. However, these methods may be used with any reasonable optimization level because the geometry choice is not considered as an intrinsic part of the multilevel approach, in contrast with other previous Gn procedures [36]. More technical information on these methods and on the electronic structure calculations required by them are given elsewhere [27–31].

All the geometry optimizations and single-point energy calculations with multilevel methods were carried out with the Gaussian03 package [37].

2.2 Density functional methods

At the hybrid density functional theory level, the optimization of the database molecules was carried out with the following combinations of method and basis set:

- The B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level of theory that has been extensively applied to model biological phosphorous compounds [23, 38].
- The B3LYP/MG3S//B3LYP/MG3S level, where the Pople basis sets have been substituted by the MG3S basis set from Truhlar's group. The MG3S basis set [35] is identical to 6-311++G(2df,2p) for C and O, and for H, except that on H diffuse functions have been removed. For phosphorous this basis set is very similar

to 6-311+G(3d,2f). In choosing the MG3S basis we were particularly concerned that the basis sets for second row atoms would include tight d-functions, because this seems to be a prerequisite for consistent accuracy on that row [39]. The MG3S satisfies this criterion because the exponent of the tightest d-function for P is 2.2 [40].

- The MPWB1K/basis set//MPWB1K/basis set, where basis set//basis set is either 6-311++G(3df,2p)//6-31++G(d,p) or MG3S//MG3S. The hybrid meta DFT MPWB1K method has been optimized against a database of barrier heights and reaction energies, and it gives remarkably good performance for kinetics. However, the results of the atomization energy calculations are somewhat deteriorated by the increased percentage of Hartree-Fock exchange in MPWB1K [34]. In any case, it is a recommended method for thermochemical kinetics.
- The MPW1B95/MG3S//MPW1B95/MG3S level of calculation has also been tested. The MPW1B95 is a hybrid meta DFT method recommended for thermochemistry. Even though it has been optimized to reproduce a small set of atomization energies, it has been shown to give good performance in the calculation of proton affinities and gas phase basicities [34]. However, it is not an accurate DFT method for barrier height calculations.

All the geometry optimizations and thermochemical analysis with density functional theory methods were carried out with the Gaussian03 package [37].

2.3 Semiempirical methods

The semiempirical methods derive their efficiency from explicit treatment of only valence electrons with minimal basis set, the neglect of three- and four-center integrals, and the use of parametrized expressions for two-center integrals. The parameters are usually obtained by fitting properties to a variety of very small compounds. Often these training sets are not representative of reactions in biological systems. However, the development of specific reaction parameters (SRP) [41] improves the results at the expense of losing generality. In the case of systems containing P and Mg, different standard and specific reaction semiempirical Hamiltonians have been developed in the recent years. For magnesium, there are, for instance, standard MNDO [42], PM3 [43, 44], AM1 [45], and PM6 [46] parameters, which were fitted to reproduce mainly properties of divalent magnesium compounds. For phosphorous, there are, for instance, standard MNDO, PM3, AM1 [47], PM6, and PDDG [48] parameters. However, AM1 and MNDO Hamiltonians have been extended to include d-orbitals for phosphorous and other second-row elements [18, 20]. More recently, different sets of semiempirical SRP parameters for phosphate hydrolysis reactions [19], phosphoryl transfer reactions [22], and for magnesium in metalloenzymes [21], have been developed. In this work, the performance of standard AM1, PM3, MNDO, and MNDO/d Hamiltonians in reproducing the experimental or multilevel ab initio or DFT results for the selected molecules and reactions in our database has been tested, together with the following reaction-specific Hamiltonians:

- The AM1/d parameterization for phosphorous developed by Lopez and Darrin [19]. This Hamiltonian has been optimized to model nucleophilic attack of phosphates relevant for biological phosphate hydrolysis reactions. These P parameters have been combined with the AM1/d parameters for magnesium developed by Imhof et al. [21], which have been derived specifically for oxygen-based ligands, modeling magnesium coordination spheres as the ones found in the PKA active site. The resulting model will be called AM1/d-P,Mg in this paper.
- The AM1/d-PhoT model, where SRP parameters were developed for H, O, and P atoms to reproduce high-level DFT results from a database of quantum calculations for RNA catalysis [22]. In this model, a scale factor was introduced into the Gaussian core–core terms. In the present work, for the molecules containing magnesium, the AM1/d-PhoT model which has been combined with the standard AM1 parameters for Mg will be called AM1/d-PhoT, and the one combined with the AM1/d parameters of Imhof et al. will be called AM1/d-PhoT,Mg.

All the semiempirical method calculations were performed with the MNDO97 package [49].

2.4 Calculation of proton affinities and gas phase basicities

In the present work we have calculated proton affinities (PA) and gas phase basicities (GPB) of several molecules with different methods. Referring to the general reaction equation:

$$A^-_{(g)} + H^+_{(g)} \rightarrow AH_{(g)} \tag{4}$$

the proton affinity of A^- is the negative of the enthalpy variation according to the process, and the gas phase basicity of A^- is the negative of the Gibbs free energy variation. The required thermodynamic properties were obtained from the electronic structure calculations under the harmonic and rigid-rotor approximations. The standard

state adopted in the gas phase was a mole of particles at 298.15 K and 1 atm pressure. For the proton, the enthalpy was calculated from the ideal gas equation

$$\mathrm{H}(H^+) = U + PV = \frac{5}{2}RT \tag{5}$$

where U is the internal energy, P the pressure, V the volume, R the universal gas constant, and T the absolute temperature. The entropy was obtained from the Sackur-Tetrode equation [50]

$$\mathrm{S}(H^+) = R \cdot \ln\left(\frac{e^{\frac{5}{2}} k_{\mathrm{B}} T}{P\Lambda^3}\right) \tag{6}$$

where k_{B} is the Boltzmann constant and Λ is the thermal De Broglie wavelength.

From Eqs. 5 and 6 we obtained the value 1.48 kcal mol^{-1} for the enthalpy and 26.02 kcal mol$^{-1}K^{-1}$ for the entropy, and, from the relation between these properties and the Gibbs free energy, the value of $\mathrm{G}(H^+) = -6.28$ kcal mol^{-1}.

2.5 Error analysis

To estimate the goodness of the ab initio, DFT and semi-empirical methods we have compared the theoretical values of PA and GPB with the experimental ones when the experimental data are available in the literature. Four error metrics have been used to analyze the general accuracy of the methods:

$$\mathrm{MAXE} = \max(\mathrm{error}_i)_{i=1}^{N} \tag{7}$$

$$\mathrm{RMSE} = \frac{\sqrt{\Sigma_{i=1}^{N}(\mathrm{error}_i)^2}}{N} \tag{8}$$

$$\mathrm{MUE} = \frac{\Sigma_{i=1}^{N}|\mathrm{error}_i|}{N} \tag{9}$$

$$\mathrm{MSE} = \frac{\Sigma_{i=1}^{N}\mathrm{error}_i}{N} \tag{10}$$

where "error" is the difference between the theoretical value and the experimental one. MAXE is the maximum error with sign. The mean signed error (MSE) is the mean error value, i.e. it represents the precision of the method. The mean unsigned error (MUE) is the average value of the magnitude of the errors and so it is a measure of the accuracy of the method. The root mean square error (RMSE) gives the second-order moment of the error distribution. It is clear that, if MSE and MUE have the same absolute value, it means that the method is affected by a systematic error. In such a case, if MSE has a negative value the method systematically underestimates the real energy value. Otherwise, if MSE has a positive value the energy is systematically overestimated. Then, it is possible to correct the calculated values with the MSE to obtain more precise ones.

In the case of the DFT methods, it was not possible to detect the best method looking at the errors of the two properties separately. Thus, we have used a combination of errors, that is, an average of the errors obtained for proton affinities and for gas phase basicities with the equation

$$\mathrm{AMETRIC} = \frac{\mathrm{METRIC}_{\mathrm{PA}} + \mathrm{METRIC}_{\mathrm{GPB}}}{2} \tag{11}$$

where METRIC stands for RMSE, MUE, or MSE to give ARMSE, AMUE, or AMSE, respectively.

3 Results and discussion

In all the tables in the following sections the names dmp and dmph will be used for dimethylphosphate anion and the dimethylphosphoric acid, respectively. Metphos and metphosh will be used for methylphosphate dianion and for methylhydrogenphosphate anion. The methylphosphate dianion has been chosen as a very simplified model of the ATP molecule, and it is treated as a dianion because this is the dominant form at physiological pH (7.5) [38]. Metphosmg and metphoshmg will be used for the complexes between the methylphosphate dianion or the methylhydrogenphosphate anion and a magnesium cation; and, finally, metphosmgwat and metphoshmgwat for the complexes between the methylphosphate dianion or the methylhydrogenphosphate anion, a magnesium cation, and four water molecules.

3.1 Performance of multilevel methods

In Tables 1 and 2 the PA and GPB calculated with multilevel methods are accounted.

The three methods give quite similar punctual errors that just in a really few cases are larger than 2 kcal mol^{-1}, thus the error metrics are of the same magnitude and very close to one another. It can be noticed that the behavior of MCG3/3 is not excellent with alcohols, where the other methods show a better concordance with the experimental data, but it demonstrates a good dealing with the molecules containing phosphorous (phosphoric acid and dmph). In the case of proton affinities it can be noticed that MCG3/3 has the smallest MAXE, but G3S/3 has the lowest RMSE and MUE, and CBS-QB3 the lowest MSE. In the case of gas phase basicity calculations, instead, MCG3/3 shows the best performance and all of its metrics are the smallest, with a MSE and a RMSE particularly tiny.

In this work, we have added three molecules that serve as gas phase models of ATP or ATP complexes: that is, methylphosphate dianion, methylphosphate dianion

Table 1 Proton affinities (errors in parenthesis) along with the values of the four error metrics defined in the text

Molecules	CBS-QB3	MCG3/3	G3S/3	Exp.[a]
Hydroxyl/water	392.02 (1.72)	389.72 (−0.58)	389.74 (−0.56)	390.3 (0.2)
Methanolate/methanol	378.71 (−2.79)	383.27 (1.77)	381.51 (0.01)	381.5 (1.0)
Formiate/formic	343.60 (−0.40)	345.21 (1.21)	343.60 (−0.40)	344.0 (1.6)
Ethanolate/ethanol	378.85 (0.65)	379.82 (1.62)	377.84 (−0.36)	378.2 (0.8)
Acetate/acetic	347.37 (0.17)	348.93 (1.73)	347.35 (0.15)	347.2 (1.1)
Propanolate/propanol	377.97 (1.97)	378.90 (2.90)	377.02 (1.02)	376.0 (1.1)
2-Propanolate/2-propanol	376.49 (0.79)	377.28 (1.58)	375.53 (−0.17)	375.7 (0.8)
Propanate/propanoic	346.20 (−1.20)	348.44 (1.04)	346.82 (−0.58)	347.4 (1.8)
Dihydrogenphosphate/phosphoric	326.82 (−3.68)	328.74 (−1.76)	327.04 (−3.46)	330.5 (5.0)
Dmp/dmph	329.95 (−1.65)	332.10 (0.50)	330.10 (−1.50)	331.6 (4.1)
Phenolate/phenol	349.32 (−0.78)	349.97 (−0.13)	348.67 (−1.43)	350.1 (1.1)
Methylphenolate/methylphenol	349.90 (−0.80)	351.21 (0.51)	349.79 (−0.91)	350.7 (1.3)
Metphos/metphosh	454.72 (–)	455.91 (–)	453.77 (–)	–
Metphosmg/metphoshmg	235.78 (–)	238.92 (–)	236.30 (–)	–
Metphosmgwat/metphoshmgwat	–	254.61 (–)	–	–
MAXE	−3.679	2.903	−3.464	
RMSE	0.492	0.427	0.365	
MUE	1.384	1.249	0.874	
MSE	−0.500	0.867	−0.684	

All values are given in kcal mol^{-1}

[a] Standard deviations in parenthesis

Table 2 Gas phase basicities (errors in parenthesis) along with the values of the four error metrics defined in the text

Molecules	CBS-QB3	MCG3/3	G3S	Exp.[a]
Hydroxyl/water	385.42 (1.72)	383.13 (−0.57)	383.15 (−0.55)	383.7 (0.2)
Methanolate/methanol	371.53 (−3.27)	376.79 (1.99)	375.03 (0.23)	374.8 (0.7)
Formiate/formic	336.13 (−1.77)	337.75 (−0.15)	336.13 (−1.77)	337.9 (1.2)
Ethanolate/ethanol	371.68 (0.38)	372.68 (1.38)	370.70(−0.60)	371.3 (1.0)
Acetate/acetic	339.55 (−1.85)	341.04 (−0.36)	339.45 (−1.95)	341.4 (1.2)
Propanolate/propanol	370.59 (1.19)	371.52 (2.12)	369.64 (0.24)	369.4 (1.1)
2-Propanolate/2-propanol	369.14 (0.34)	370.03 (1.23)	368.27 (−0.53)	368.8 (1.0)
Propanate/propanoic	340.00 (−0.40)	340.43 (0.03)	338.81 (−1.59)	340.4 (1.4)
Dihydrogenphosphate/phosphoric	319.87 (−3.33)	321.79 (−1.41)	320.09 (−3.11)	323.2 (4.9)
Dmp/dmph	323.97 (−0.63)	325.43 (0.83)	323.42 (−1.18)	324.6 (4.0)
Phenolate/phenol	341.83 (−1.07)	342.47 (−0.43)	341.16 (−1.74)	342.9 (1.4)
Methylphenolate/methylphenol	343.91 (0.11)	343.53 (−0.27)	342.11 (−1.69)	343.8 (1.2)
Metphos/metphosh	447.22 (–)	448.47 (–)	446.34 (–)	–
Metphosmg/metphoshmg	236.49 (–)	231.92 (–)	229.30 (–)	–
Metphosmgwat/metphoshmgwat	–	250.24 (–)	–	–
MAXE	−3.326	2.121	3.109	
RMSE	0.490	0.326	0.436	
MUE	1.338	0.865	1.310	
MSE	−0.715	0.365	−1.185	

All values are given in kcal mol^{-1}

[a] Standard deviations in parenthesis

coordinated to Mg (which forms a bidentate complex), and this last model with a octahedral coordination for magnesium that is completed by four water molecules. There are no experimental values of the PA and GPB for these molecules, which makes worthy to compare the numerical values obtained by the three multilevel methods chosen. However, this is not possible in the case of the biggest complex, because in this case we have been able to obtain converged results only for the MCG3/3 method. The results obtained for the other two molecules with the three methods indicate that they generally agree (see Tables 1 and 2).

In summary the lowest computational cost and the best overall performance is reached by the MCG3/3 method. For instance, the computational time request for a single-point energy calculation, done for methylphenolate on the same computer (AMD Operon 1.6 MHz), is ~10 h with CBS-QB3 method, ~20 h with G3S/3, and just ~6 h with MCG3/3.

The MCG3/3 method has been chosen as the reference, where no experimental data is available.

3.2 Density functional theory methods

In Tables 3 and 4 the results of the proton affinity and gas phase basicity calculations for the three DFT methods are listed. In general, these results are comparable to the values obtained with the multilevel methods. In order to permit an unambiguous selection of the best method, we have also used a further set of error metrics. It simply consists in the mathematical average between the error metrics for proton affinity and for gas phase basicity (Table 5), as defined by Eq. 11.

It can be immediately seen that MPW1B95/MG3S has the smallest error metrics and that the difference with all the other methods is fairly big. The AMSE of MPW1B95/MG3S is an order of magnitude smaller than that of B3LYP and of MPWB1K. Most of the calculated values of PA and GPB are very accurate. Anyhow, with this method it was not possible to obtain optimized structures of the two magnesium complexes of our set.

B3LYP, with the two selected basis sets, does work quite well and we have not encountered any optimization problem with any molecule of the set. Even if B3LYP/MG3S needs a long computational time it has the best global behavior and all the error metrics are smaller than the other methods except for MPW1B95/MG3S.

It has to be noticed that the B3LYP method is affected by a systematic error. With both basis sets we found very similar values for AMUE and AMSE and in both cases the calculated properties are underestimated. From Tables 3

Table 3 Proton affinities (errors in parenthesis) along with the values of the four error metrics defined in the text

Molecules	B3LYP 1	B3LYP 2	MPW1B95	MPWB1K	Exp.[a]
Hydroxyl/water	390.42 (0.12)	390.30 (0.00)	393.35 (3.05)	395.56 (5.26)	390.3 (0.2)
Methanolate/methanol	379.17 (−2.33)	380.03 (−1.47)	382.05 (0.55)	384.42 (2.92)	381.5 (1.0)
Formiate/formic	342.06 (−1.94)	342.01 (−1.99)	343.68 (−0.32)	345.28 (1.28)	344.0 (1.6)
Ethanolate/ethanol	376.02 (−2.18)	376.88 (−1.32)	379.08 (0.88)	381.79 (3.59)	378.2 (0.8)
Acetate/acetic	346.43 (−0.77)	346.46 (−0.74)	348.12 (0.92)	349.77 (2.57)	347.2 (1.1)
Propanolate/propanol	375.76 (−0.24)	376.30 (0.30)	378.22 (2.22)	380.86 (4.86)	376.0 (1.1)
2-Propanolate/2-propanol	374.37 (−1.33)	374.90 (−0.80)	377.14 (1.44)	384.90 (9.20)	375.7 (0.8)
Propanate/propanoic	346.15 (−1.25)	346.13 (−1.27)	347.59 (0.19)	349.31 (1.91)	347.4 (1.8)
Dihydrogenphosphate/phosphoric	327.21 (−3.29)	327.03 (−3.47)	328.35 (−2.15)	329.97 (−0.53)	330.5 (5.0)
Dmp/dmph	331.27 (−0.33)	330.36 (−1.24)	331.74 (0.14)	333.41 (1.81)	331.6 (4.1)
Phenolate/phenol	347.65 (−2.45)	347.81 (−2.29)	349.27 (−0.83)	351.30 (1.20)	350.1 (1.1)
Methylphenolate/methylphenol	348.89 (−1.81)	349.03 (−1.67)	350.48 (−0.22)	352.65 (1.95)	350.7 (1.3)
Metphos/metphosh	453.59 (−2.32)	453.64 (−2.27)	456.14 (0.23)	458.31 (2.40)	*455.91*
Metphosmg/metphoshmg	235.77 (−3.15)	235.44 (−3.48)	–	–	*238.92*
Metphosmgwat/metphoshmgwat	250.93 (−3.68)	251.70 (−2.91)	–	–	*254.61*
MAXE	−3.680	−3.480	3.046	9.204	
RMSE	0.546	0.508	0.375	1.043	
MUE	1.813	1.681	1.011	3.037	
MSE	−1.797	−1.641	0.469	2.955	

All values are given in kcal mol^{-1}. Due to the absence of experimental values, in the Exp. column, the values in italic characters are taken from the MCG3/3 calculations and the corresponding errors are calculated with respect to them. B3LYP 1 stands for B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p); with all the other reported methods, the basis set used for the geometry optimization as well as for the single point calculation is MG3S

[a] Standard deviations in parenthesis

Table 4 Gas phase basicity (errors in parenthesis) along with the values of the four error metrics defined in the text

Molecules	B3LYP 1	B3LYP 2	MPW1B95	MPWB1K	Exp.[a]
Hydroxyl/water	383.83 (0.13)	383.71 (0.01)	386.75 (3.05)	388.96 (5.26)	383.7 (0.2)
Methanolate/methanol	372.65 (−2.15)	373.55 (−1.25)	375.55 (0.75)	377.94 (3.14)	374.8 (0.7)
Formiate/formic	334.59 (−3.31)	334.55 (−3.35)	336.20 (−1.70)	337.81 (−0.09)	337.9 (1.2)
Ethanolate/ethanol	368.87 (−2.43)	369.74 (−1.56)	371.91 (0.61)	374.61 (3.31)	371.3 (1.0)
Acetate/acetic	338.38 (−3.02)	338.57 (−2.83)	339.16 (−2.24)	341.86 (0.46)	341.4 (1.2)
Propanolate/propanol	368.45 (−0.95)	368.91 (−0.49)	370.84 (1.44)	373.51 (4.11)	369.4 (1.1)
2-Propanolate/2-propanol	367.10 (−1.70)	367.64 (−1.16)	369.88 (1.08)	377.23 (8.43)	368.8 (1.0)
Propanate/propanoic	338.27 (−2.13)	338.12 (−2.28)	339.86 (−0.54)	341.67 (1.27)	340.4 (1.4)
Dihydrogenphosphate/phosphoric	320.08 (−3.12)	320.09 (−3.11)	321.35 (−1.85)	323.01 (−0.19)	323.2 (4.9)
Dmp/dmph	324.30 (−0.30)	323.68 (−0.92)	324.66 (0.06)	326.25 (1.65)	324.6 (4.0)
Phenolate/phenol	340.13 (−2.77)	340.31 (−2.59)	341.75 (−1.15)	343.80 (0.90)	342.9 (1.4)
Methylphenolate/methylphenol	341.35 (−2.45)	341.36 (−2.44)	342.82 (−0.98)	345.28 (1.48)	343.8 (1.2)
Metphos/metphosh	446.17 (−2.30)	446.21 (−2.26)	448.70 (0.23)	451.02 (2.55)	*448.47*
Metphosmg/metphoshmg	228.68 (−3.24)	228.44 (−3.48)	–	–	*231.92*
Metphosmgwat/metphoshmgwat	246.67 (−3.57)	247.33 (−2.91)	–	–	*250.24*
MAXE	−3.570	−3.480	3.051	8.428	
RMSE	0.636	0.592	0.403	0.944	
MUE	2.238	2.043	1.206	2.526	
MSE	−2.221	−2.041	−0.095	2.483	

All values are given in kcal mol^{-1}. Due to the absence of experimental values, in the Exp. column, the values in italic characters are taken from the MCG3/3 calculations and the corresponding errors are calculated with respect to them. B3LYP 1 stands for B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p); with all the other reported methods, the basis set used for the geometry optimization as well as for the single point calculation is MG3S

[a] Standard deviations in parenthesis

Table 5 Average errors of proton affinities and gas phase basicities of all the DFT methods tested

All values are given in kcal mol^{-1}

	ARMSE	AMUE	AMSE
B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p)	0.591	2.025	−2.009
B3LYP/MG3S//B3LYP/MG3S	0.550	1.862	−1.841
MPW1B95/MG3S//MPW1B95/MG3S	0.389	1.108	0.187
MPWB1K/6-31+G(d,p)//MPWB1K/6-31G(d)	0.625	1.922	0.858
MPWB1K/6-311++G(3df,2p)//MPWB1K/6-31++G(d,p)	0.703	2.134	2.107
MPWB1K/MG3S//MPWB1K/MG3S	0.993	2.782	2.719

and 4, it can be seen that with the exception of the hydroxyl molecule (which is almost perfectly managed) all the other systems of the set have negative errors.

In comparison with the previous DFT methods, MPWB1K does not have really good performances; see the supplementary material file for the results of these methods with smaller basis sets. First of all, with the biggest basis set it was not possible to calculate the properties of the two magnesium complexes, while with the 6-311++G(3df,2p) basis set it was not possible to optimize the structure of the magnesium complex with water molecules (the largest system considered in the present work). With this method, we have also noticed an unusual behavior: the bigger the basis set and the longer the computational time request the worse are the results. MPWB1K/6-31+G(d,p) gives errors comparable with B3LYP/MG3S (see supplementary material file), but, using the MG3S basis set we obtain larger errors than those of B3LYP or MPW1B95.

The various combinations of methods and basis sets have a wide difference in the computational costs. We have carried out the geometry optimization, the frequency calculation, and the single-point energy calculation of methylphenolate on the same machine (Athlon XP 2800 with 1 GB of RAM) and obtained that B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) requires ~8 h, B3LYP/MG3S requires ~20 h, MPW1B95/MG3S needs ~21 h, MPWB1K/6-31+G(d,p)//MPWB1K/6-31G(d) requires ~1.5 h, MPWB1K/6-311++G(3df,2p)//MPWB1K/6-31

++G(d,p) needs ~5 h, and finally MPWB1K/MG3S requires ~23 h.

3.3 Semiempirical methods

In Table 6 the obtained values of the proton affinities with some of the semiempirical methods tested are reported. In this subsection we will discuss only the results obtained with AM1 and the different parameterizations of AM1/d because they show a better global behavior with respect to PM3, MNDO, and MNDO/d; the results obtained with all the remaining methods tested can be found in the supplementary material.

It can be seen that the AM1/d-PhoT has the best performance obtaining the smallest value of three error metrics: it has the highest accuracy, the smallest dispersion, and the lowest MAXE. Just the precision turns out to be higher with the AM1/d-PhoT,Mg. Anyway, these two parameterizations show similar values for MSE, so we think it is not worthy going on with further analysis of this mixed parameterization.

It is noticeable that the parameterization by Nam et al. [22] (AM1/d-PhoT and the AM1/d-PhoT,Mg methods) gives really better results for the hydroxyl/water couple, which is one of the most problematic systems for all the other methods or parameterizations: in many cases it is responsible for the maximum error. Also we have to underline that the errors along all the set of molecules are generally smaller or, at most, comparable with those from the other semiempirical methods.

3.4 Geometries

In this subsection we will briefly discuss the geometries obtained for the three model systems with methylphosphate dianion at the B3LYP/MG3S level and using the AM1/d-PhoT and AM1/d-P,Mg parameterizations. We have chosen only these three structures because they are the most representative of the biological system we want to study.

In Figs. 2, 3, and 4 the different geometries of the mentioned systems are presented. In all of them the B3LYP values are presented with normal fonts, the AM1/d-PhoT values with bold fonts, and the AM1/d-P,Mg ones with italic fonts.

In the first case (Fig. 2) it can be observed that, in the absence of Mg ions, the three methods give almost the same results, with bond distances slightly larger for the two semiempirical methods.

In the other two cases (Figs. 3 and 4), in which a Mg ion is present, some difference in the behavior must be noticed.

Table 6 Semiempirical proton affinities (errors in parenthesis) along with the values of the four error metrics defined in the text

Molecules	AM1/d-PhoT	AM1/d-P,Mg	AM1/d-PhoT,Mg	AM1	Exp.[a]
Hydroxyl/water	395.69 (5.39)	410.83 (20.53)	395.69 (5.39)	410.83 (20.53)	390.3 (0.2)
Methanolate/methanol	383.51 (2.01)	384.23 (2.73)	383.51 (2.01)	384.23 (2.73)	381.5 (1.0)
Formiate/formic	343.33 (−0.67)	353.64 (9.64)	343.33 (−0.67)	353.64 (9.64)	344.0 (1.6)
Ethanolate/ethanol	381.11 (2.91)	382.87 (4.67)	381.11 (2.91)	382.87 (4.67)	378.2 (0.8)
Acetate/acetic	344.51 (−2.69)	353.29 (6.09)	344.51 (−2.69)	353.29 (6.09)	347.2 (1.1)
Propanolate/propanol	380.19 (4.19)	381.52 (5.52)	380.19 (4.19)	381.52 (5.52)	376.0 (1.1)
2-Propanolate/2-propanol	383.55 (7.85)	383.02 (7.32)	383.55 (7.85)	383.02 (7.32)	375.7 (0.8)
Propanate/propanoic	344.72 (−2.68)	352.64 (5.24)	344.72 (−2.68)	352.64 (5.24)	347.4 (1.8)
Dihydrogenphosphate/phosphoric	327.54 (−2.96)	327.40 (−3.10)	327.54 (−2.96)	337.85 (7.35)	330.5 (5.0)
Dmp/dmph	331.18 (−0.42)	325.42 (−6.18)	331.18 (−0.42)	337.70 (6.10)	331.6 (4.1)
Phenolate/phenol	346.70 (−3.40)	346.97 (−3.13)	346.70 (−3.40)	346.97 (−3.13)	350.1 (1.1)
Methylphenolate/methylphenol	347.68 (−3.02)	346.28 (−4.42)	347.68 (−3.02)	346.28 (−4.42)	350.7 (1.3)
Metphos/metphosh	455.24 (−0.67)	459.66 (3.75)	455.24 (−0.67)	471.20 (15.29)	*455.91*
Metphosmg/metphoshmg	226.61 (−12.31)	219.12 (−19.80)	223.61 (−15.31)	234.49 (−4.43)	*238.92*
Metphosmgwat/metphoshmgwat	251.29 (−3.32)	248.54 (−6.07)	268.05 (13.44)	263.81 (9.2)	*254.61*
MAXE	−12.310	20.525	−15.310	20.525	
RMSE	0.906	2.321	1.607	2.262	
MUE	2.590	7.213	4.507	7.444	
MSE	0.388	1.519	0.265	5.847	

All values are given in kcal mol^{-1}. Due to the absence of experimental values, in the Exp. column, the values in italic characters are taken from the MCG3/3 calculations and the corresponding errors are calculated with respect to them

[a] Standard deviations in parenthesis

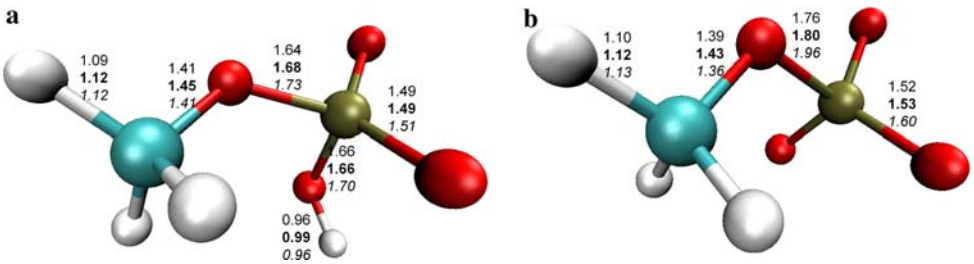

Fig. 2 Structures of **a** metphosh and **b** metphos. Normal characters B3LYP/MG3S, bold characters AM1/d-PhoT, italic characters AM1/d-P,Mg. Distances are given in Å

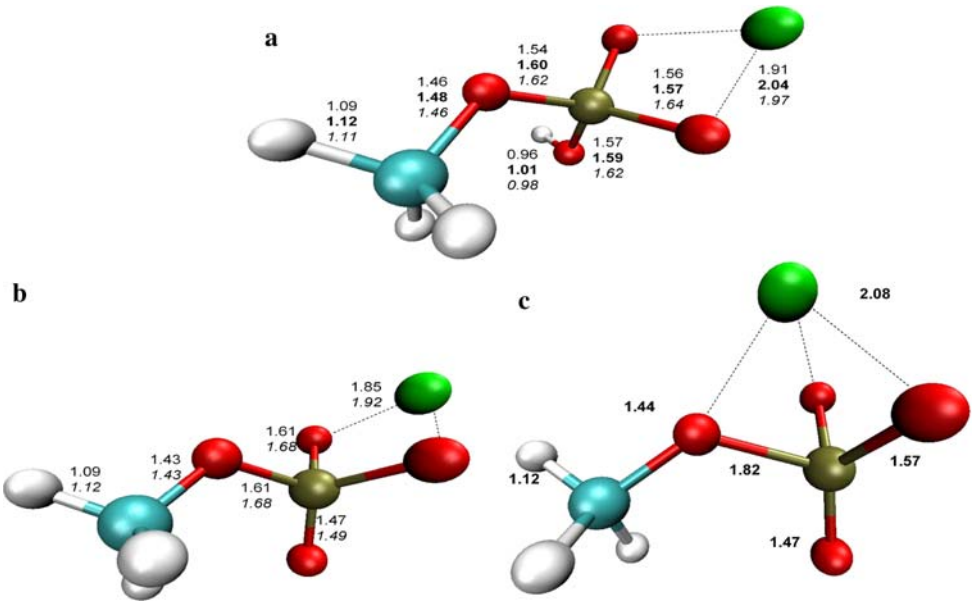

Fig. 3 Structures of metphoshmg (**a**) and metphosmg (**b** and **c**). b: B3LYP and AM1/d-P,Mg; c: AM1/d-PhoT. Normal characters B3LYP/MG3S, bold characters AM1/d-PhoT, italic characters AM1/d-P,Mg. Distances are given in Å

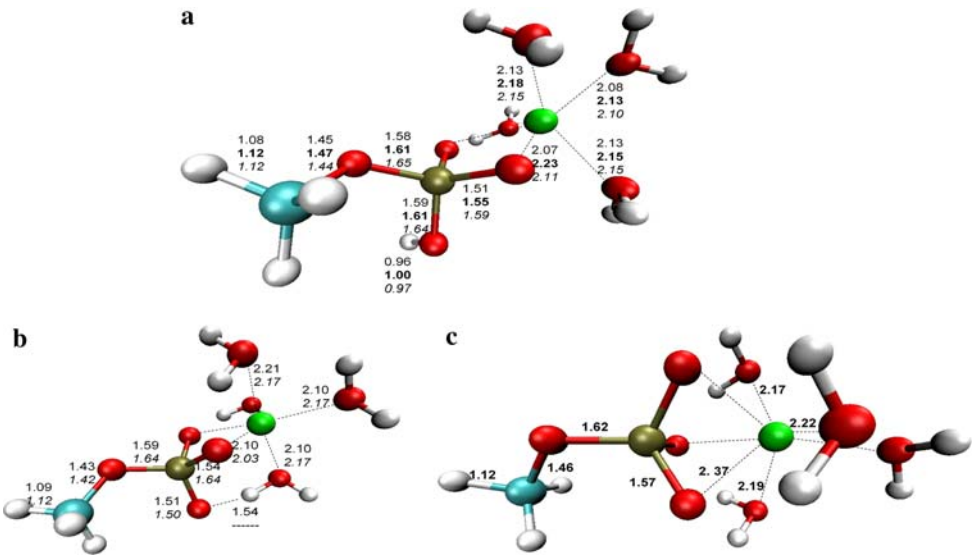

Fig. 4 Structures of metphoshmgwat (**a**) and metphosmgwat (**b** and **c**). b: B3LYP and AM1/d-P,Mg; c: AM1/d-PhoT. Normal characters B3LYP/MG3S, bold characters AM1/d-PhoT, italic characters AM1/d-P,Mg. Distances are given in Å

In both cases the DFT method and the AM1/d-P,Mg method give quite similar results, but always the semi-empirical one seems to overestimate the bond distances.

When AM1/d-PhoT is used, bigger geometric differences turn out to be present in both models with Mg when they are deprotonated, thus, when a bigger negative charge has to be treated. In these cases, this semiempirical method can find only one minimum energy configuration for each, in which the Mg ion has three phosphate oxygens coordinating. These minimum energy configurations can be found also by the other methods discussed in this subsection, but never as the lowest minima.

As a consequence, depending on the aspect which one is interested in, the choice of the parameterization turns out to be really critical. Neither one or the other set of parameters seems to allow an accurate determination of both the energy and the geometry of a system in the presence of a Mg ion, and when many negative charges are involved, at least in gas phase models.

3.5 Reactions

In this subsection we will discuss three model reactions meant to better define the goodness of the best methods previously analyzed and of the AM1/d-PhoT and the AM1/d-P,Mg parameterizations in the study of the phosphoryl transfer reaction.

3.5.1 Reaction 1: the associative mechanism

The associative mechanism is depicted in Fig. 1 (Scheme A). It consists of the phosphoryl- and proton-transfers between the substrate and the ATP molecule. We have used reaction 1 to model this mechanism in a really simple way: in our gas phase system the methylphosphate dianion is a model of an ATP molecule and the water molecule represents the serine residue of the substrate. Table 7 shows the energetic results for this reaction, while in Fig. 5 the corresponding approximated energy profiles are depicted. It can be observed that almost all the methods presented in Table 7 and Fig. 5 give quite similar values for ΔE_{RC}, surrounding the -30 kcal mol^{-1}, and that all the multilevel methods give practically the same values for all the given structures. Note that B3LYP and the three multilevel methods give nearly the same transition state barrier height. Only the AM1/d-P,Mg gives a ΔE_{RC} rather different, nearly 12 kcal mol^{-1} higher than the others. The difference between this method and the multilevel methods is maintained also in the transition state energy value even though lowered to 5.5 kcal mol^{-1}. Thus, the transition state energy turns out to be overestimated at the AM1/d-P,Mg level (using as a reference the energy of the separated reactants).

Table 7 Energy variations of the reaction $[CH_3PO_4]^{--} + H_2O \rightarrow [CH_3O]^- + [H_2PO_4]^-$ in kcal mol^{-1} obtained with selected methods

	ΔE_{RC}	ΔE_{TS}	ΔE_{reac}
AM1/d-P,Mg	−18.78	+27.83	−74.61
AM1/d-PhoT	−30.31	+16.54	−74.79
B3LYP/MG3S	−29.73	+22.14	−72.05
CBS-QB3	−31.82	+22.25	−68.46
MCG3/3	−30.68	+22.33	−68.81
G3S/3	−31.54	+21.85	−68.20

ΔE_{RC}, ΔE_{TS}, and ΔE_{reac} are the energies of the reactant complex, the transition state structure, and the separated products, respectively, taking the separated reactants as the origin of energies

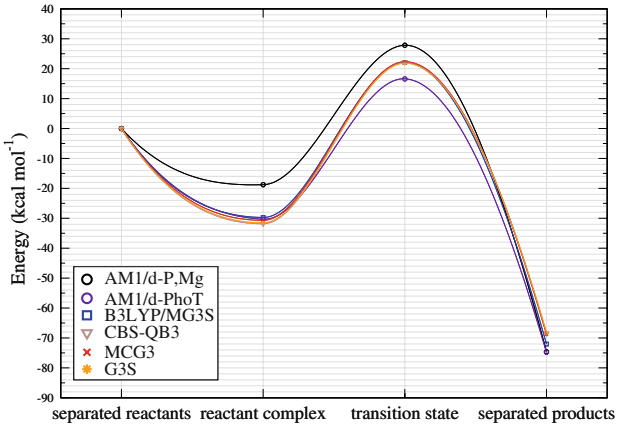

Fig. 5 Graphic representation of the energy profiles for reaction 1 obtained with different methods; *solid lines* are just suggestive of the real energy trend

AM1/d-PhoT shows better performances in the energy evaluation, although the transition state energy is underestimated by 5.6 kcal mol^{-1} with respect to B3LYP and the multilevel methods.

Figures 6, 7, and 8 show the geometric results for the reactant complex and transition state structures found for this reaction with B3LYP, AM1/d-PhoT, and AM1/d-P,Mg, respectively. In all the three cases we have found quite similar results for the reactant complex, although we have confirmed that the covalent bonds and the interaction distances with the semiempirical methods are somewhat overestimated. A big difference, instead, has been found in the transition state: with B3LYP we observe an almost in-line substitution as we were expecting for a S_N2 reaction, which results in an inversion of configuration. However, with both semiempirical methods we have obtained a transition state in which the nucleophilic substitution is not linear (Figs. 7b and 8b). As a consequence, going through this path a product without the inversion of configuration will be obtained. Obviously, just because in this case the atoms surrounding the P atoms are perfectly equivalent, the products are the same with the three methods, but this fact has to be taken into account with other non-symmetric systems.

3.5.2 Reaction 2 and reaction 3: the dissociative mechanism

The dissociative mechanism is depicted in Fig. 1 (Scheme B) and involves the conserved residue Asp166 as general base that accepts the serine proton during the nucleophilic displacement. We have separated the entire mechanism into two simple steps, thus, we have used reaction 2 as a model of the proton transfer and reaction 3 as model of the nucleophilic step.

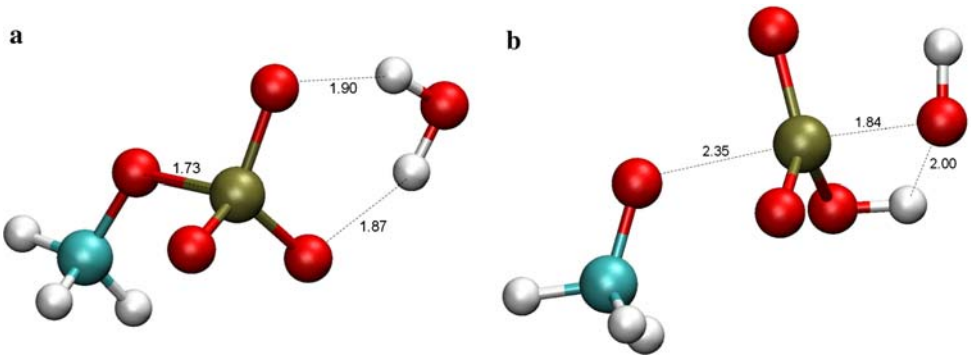

Fig. 6 Structures of the model reaction 1 obtained with B3LYP/MG3S: **a** Reactant complex; **b** transition state. Distances are given in Å

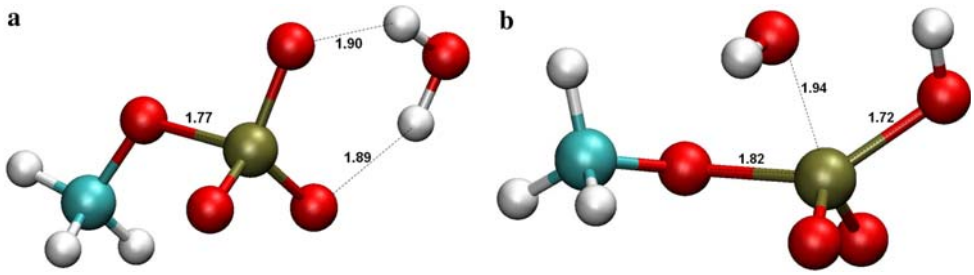

Fig. 7 Structures of the model reaction 1 obtained with AM1/d-PhoT. **a** Reactant complex; **b** transition state. Distances are given in Å

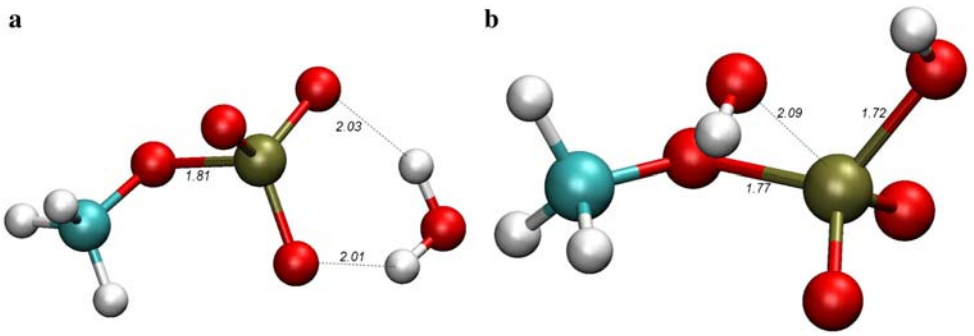

Fig. 8 Structures of the model reaction 1 obtained with AM1/d-P,Mg. **a** Reactant complex; **b** transition state. Distances are given in Å

The two reactions have been, respectively, selected to mimic a proton transfer (reaction 2) between a serine side chain (an alcohol), modeled with an ethanol molecule, and an aspartate side chain (a conjugated base of an organic acid), modeled with an acetate molecule; and a γ-phosphoryl transfer (reaction 3), in which the model of the serine is a molecule of water and the ATP is a methylphosphate dianion. It is worthy to clarify that the reason for the large size of the model of the nucleophilic step (reaction 3) is the failure of smaller models. We have found that at least a magnesium ion is required to make this chemical step possible, by stabilizing the transition state and product complex, while the presence of mobile protons (as the protons of polarized water are) are needed to allow the formation of the products. Thus, we have decided to complete the coordination sphere of magnesium with four water molecules, in order to also mimic the environment of magnesium in the active site of proteins.

In Fig. 9 the results corresponding to reaction 2 are presented. In this reaction there is no transition state, the profile just shows a continuous energy rise from the reactants (on the right) to the products (on the left). The vertical dashed line represents the equilibrium bond distance between the hydroxyl oxygen and the acid proton in the acetic acid.

Given the small size of this model we have evaluated all the DFT methods, not only B3LYP. It is easy to notice that the three multilevel methods, as we have already seen in Sect. 3.1, also in this case give almost the same results along the scan: the differences amongst the three profiles are always smaller than 1 kcal mol^{-1}.

Taking the multilevel methods as a reference, it can be noticed that, within the tested DFT, B3LYP and MPWB1K have really good performances giving results in perfect agreement with the multilevel methods. The MPW1B95, instead, gives good results just in the region around the

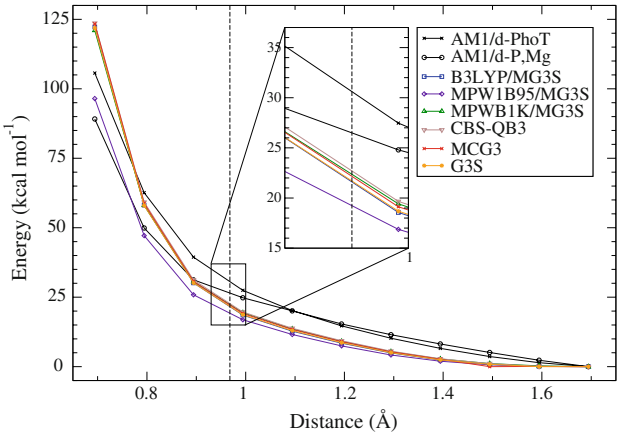

Fig. 9 Energy profiles of the reaction $CH_3CH_2OH + CH_3COO^- \rightarrow CH_3CH_2O^- + CH_3COOH$ versus the Ethanol−H···O−Acetate distance. The *solid lines* represent the energy variation, for any method, with respect to the energy of the separated reactants. The vertical dashed *black line* represents the standard O−H distance in an acetic acid molecule

reactants and gives increasingly worse results approaching the products region, underestimating the energy values.

The two parameterizations of AM1/d give quite similar results in the reactants region but they have a slightly different behavior approaching the less stable region: AM1/d-PhoT gives higher values than the reference methods along almost all the scan, just the last point is lower in energy. In the region around the vertical dashed line it overestimates the energy by $\sim$9 kcal mol^{-1} with respect to the multilevel approaches.

AM1/d-P,Mg at the beginning gives values slightly higher than AM1/d-PhoT, but its profile presents a lower slope, thus, approaching the products region it gives better results, with an error at the dashed line amounting to $\sim$4 kcal mol^{-1}: less than half the AM1/d-PhoT error.

Finally we have analyzed the phosphoryl-transfer. In this case, due to the complexity of the model, we just consider the B3LYP method and the two parameterizations of AM1/d.

In Table 8 the energy variations related to the second part of reaction 3 are shown. As in reaction 1, considering the results from the B3LYP/MG3S calculation as the correct ones, AM1/d-P,Mg overestimates the energy of the transition state, whereas the AM1/d-PhoT underestimates it, the error of AM1/d-P,Mg being larger. On the other hand, with AM1/d-P,Mg, the reaction comes out to be endoergic while following the other methods it would have to be exoergic. With AM1/d-PhoT both the $\Delta E'_{TS}$ and ΔE_{PC} values are $\sim$1.8 kcal mol^{-1} higher than the B3LYP ones.

In Figs. 10, 11, and 12 the structures of the reactant complex, transition state, and product complex, obtained at the B3LYP, AM1/d-PhoT, and AM1/d-P,Mg levels, respectively, are depicted.

Table 8 Energy variations of the reaction $[CH_3PO_3OH]^- + H_2O + Mg^{++}[H_2O]_3[OH]^- \rightarrow CH_3OH + [H_2PO_4]^- + Mg^{++}[H_2O]_3[OH]^-$ in kcal mol^{-1} obtained with selected methods

	$\Delta E'_{TS}$	ΔE_{PC}
AM1/d-P,Mg	+51.54	+4.74
AM1/d-PhoT	+37.67	−4.09
B3LYP/MG3S	+39.44	−5.80

$\Delta E'_{TS}$ and ΔE_{PC} are the energies of the transition state structure and the product complex, respectively, taking the reactant complex as the origin of energies

With B3LYP we have obtained three structures in which the Mg ion maintains an octahedral coordination slightly distorted, due to the two oxygens of the phosphoryl group, along all the reaction path. The transition state is linear and, in the product complex, the methanol is protonated and one of the Mg ion ligands is a hydroxyl molecule.

With AM1/d-PhoT we have obtained structures quite different: first of all, in both the reactant and the product complex the Mg ion is coordinating seven ligands instead of six. The transition state is linear but the octahedral coordination is highly distorted. Finally, in the product complex, the methanol molecule turns out to be negatively charged (actually a methanolate molecule) and coordinates the Mg ion.

Unexpectedly, the situation gets worse with AM1/d-P,Mg that up to now has shown better performances in the optimization of geometries. In the reactant complex, the Mg ion has an octahedral coordination sphere, but it is highly distorted, and the attacking water molecule is not in the right position: instead of being in line with the leaving group and the P atom, it is almost in line with the P atom and one of its oxygens, practically opposite to the Mg ion. The transition state is almost linear, but the methylphosphate fragment turns out to be neutral instead of negatively charged, and it does not correspond to the planar umbrella configuration. Finally, the product complex is really similar to the AM1/d-PhoT one, with the methanol negatively charged and coordinating the Mg ion.

3.6 Comparison between AM1/d-PhoT and B3LYP/MG3S 2-D PES for the associative and the dissociative mechanisms

The results described in the previous sections indicate that the AM1/d-PhoT parameterization is the best choice among the semiempirical methods studied in this work. However, as already said, it is not accurate enough in both the geometries and energies of the model reactions, which prevents us from using it without further corrections. On the contrary, we propose the following strategy to obtain an accurate QM/MM potential energy surface for the study of

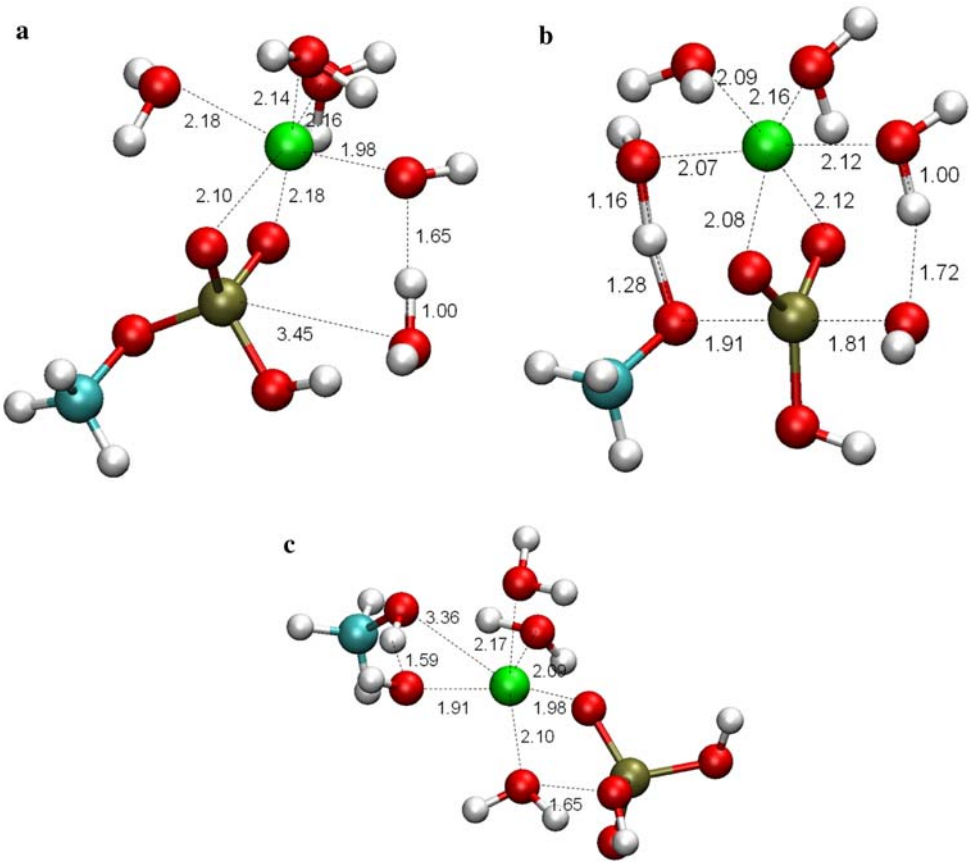

Fig. 10 Structures of the model reaction 3 obtained with B3LYP/MG3S. **a** Reactant complex; **b** transition state; **c** Product complex. Distances are given in Å

the PKA reaction mechanism. On one hand, the structural problems are solved when one uses a complete microscopic model of the biological system (including the protein, the ligands, and the solvent), and on the other hand, we propose the use of a correction term to improve the energetic results. This methodology has already been used for the study of several enzyme reaction mechanisms [51–54]. The correction term is obtained as the difference between the energy provided by the high-level method (HL) and a low-level one (LL) for an ensemble of configurations. The set of configurations used to obtain the correction term should be as representative as possible of the configurations to be explored during the simulations. A natural choice is to select the configurations from a simulation of the reaction under study obtained in the presence of the environment [52].

Thus, the strategy used in this work has been to use a more representative gas phase model of the system containing 35 atoms of the active site. Specifically, we have cut the Cartesian coordinates of a subset of QM atoms corresponding to optimized structures on a QM/MM potential energy surface obtained for a complete microscopic model of the biological system [55]. The chosen subset of atoms includes the side chains of Asp166 and the serine substrate, and the triphosphate arm of ATP, summing up to 35 atoms. We have considered the inclusion of the magnesium ions, but this will require to include their ligands too, which will result in a too large, non-computational affordable, gas phase model. We will not describe here the results obtained for the QM/MM model because this is not the purpose of this paper, but we briefly describe the model and the reaction coordinates used. We have built a model starting from the crystallographic structure of PKA complexed with an ATP analog and the PKI (5–24) inhibitor, by replacing these two last molecules by ATP and the kemptide substrate, respectively. The resulting system has a total of 38,134 atoms and has been equilibrated by carrying out 6 ns of molecular dynamics simulation [13]. Subsequently, the system has been divided into a QM subsystem, containing part of the substrate, the metal cofactors, the side-chain of their ligands, and the side-chain of the enzymatic residues implied in the biochemical reaction; and the rest of the atoms, which are represented by molecular mechanics (MM).

For the two proposed mechanisms 2D potential energy surfaces (2-D PES) have been explored as a function of two antisymmetric reaction coordinates, one for each of the bond-breaking/bond-forming processes implied in each mechanism. That is, the reaction coordinates for the associative mechanisms are: First, the antisymmetric

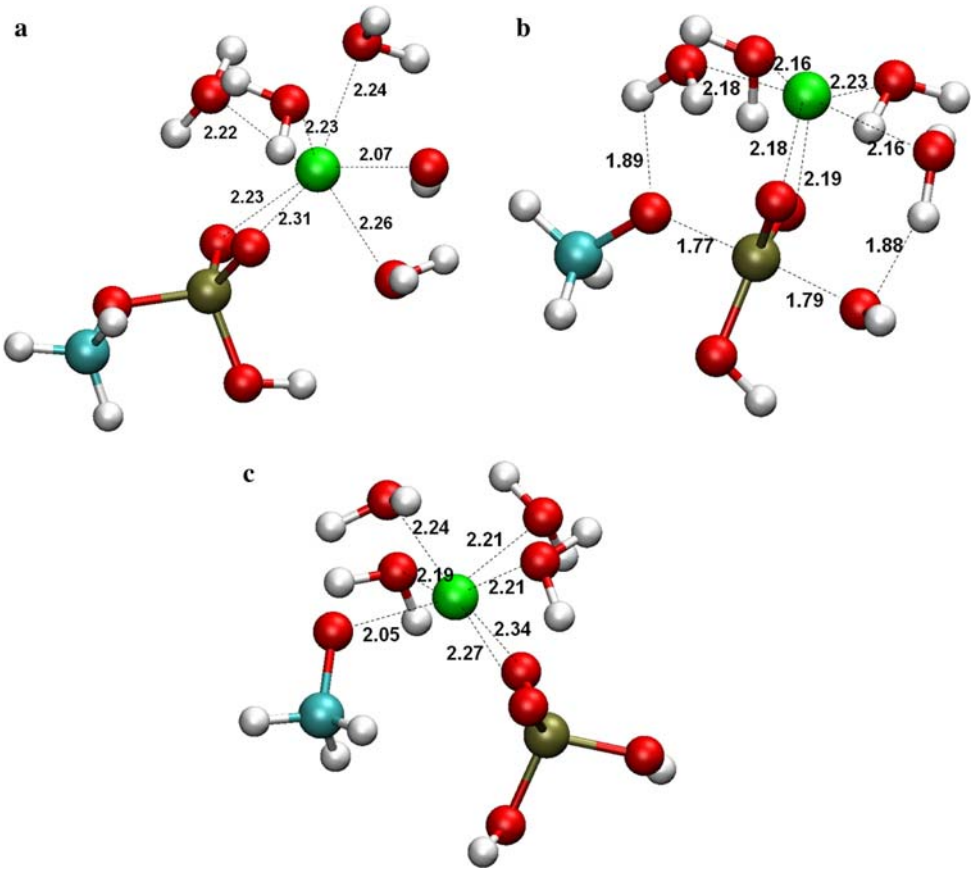

Fig. 11 Structures of the model reaction 3 obtained with AM1/d-PhoT. **a** Reactant complex; **b** transition state; **c** product complex. Distances are given in Å

combination of distances involving the gamma phosphorus of ATP (ATPPγ), the oxygen bonded to it (ATPOβ) and the oxygen of the substrate Ser residue (SerO). This reaction coordinate is labeled RC_{NUC} because it is representative of the nucleophilic attack (NUC) represented in Fig. 1 (Scheme A):

$$RC_{NUC} = 0.5(r_{ATPP\gamma-ATPO\beta}) - 1.0(r_{SerO-ATPP\gamma})$$

And second, for the proton transfer (PT) involved in the associative mechanism we have used the following distinguished reaction coordinate:

$$RC_{PT} = 0.5(r_{SerH-SerO}) - 0.5(r_{SerH-ATPO\gamma})$$

where the participants are the hydroxyl group of the serine substrate and one of the gamma oxygens of ATP (ATPOγ).

Analogously, for the dissociative mechanism (Fig. 1, Scheme B) we have used the following reaction coordinates:

$$RC_{NUC} = 0.5(r_{ATPP\gamma-ATPO\beta}) - 1.0(r_{SerO-ATPP\gamma})$$

$$RC_{PT} = 0.5(r_{SerO-SerH}) - 0.5(r_{SerH-AspO})$$

where in this case, the proton acceptor is an oxygen of the carboxylate group of the active site aspartate (AspO).

In the QM/MM optimizations the AM1/d-PhoT semi-empirical parameterization has been used for the QM subsystem, while the CHARMM force field [56] has been employed for the MM part of the system.

Figure 13 compares the gas phase 2-D PESs obtained at the B3LYP/MG3S and AM1/d-PhoT levels by carrying out single point calculations for the 35 atoms subset of the QM system on structures picked up from the AM1/d-PhoT/MM PES of the associative mechanism. The figure indicates that for all the structures of the 2-D PES the AM1/d-PhoT method underestimates the relative energy with respect to the reactant complex in comparison with B3LYP level of theory, which is the same trend observed in the results of the model reaction 1 shown in Table 7. Taking into account all the chemically relevant structures of the 2-D PES, the average correction is 23.3 kcal mol^{-1}. This result indicates that the difference between the HL and LL methods has increased in going from the smaller reaction model (12 atoms) to the larger model (35 atoms). However, it is worthy to note that the model size is not the only difference between the two models. Another difference comes from the fact that in the larger model the geometries have not been optimized in gas phase, but they have been cut from the optimized QM/MM geometries. In Fig. 13 it can be seen that the two surfaces depicted are similar just

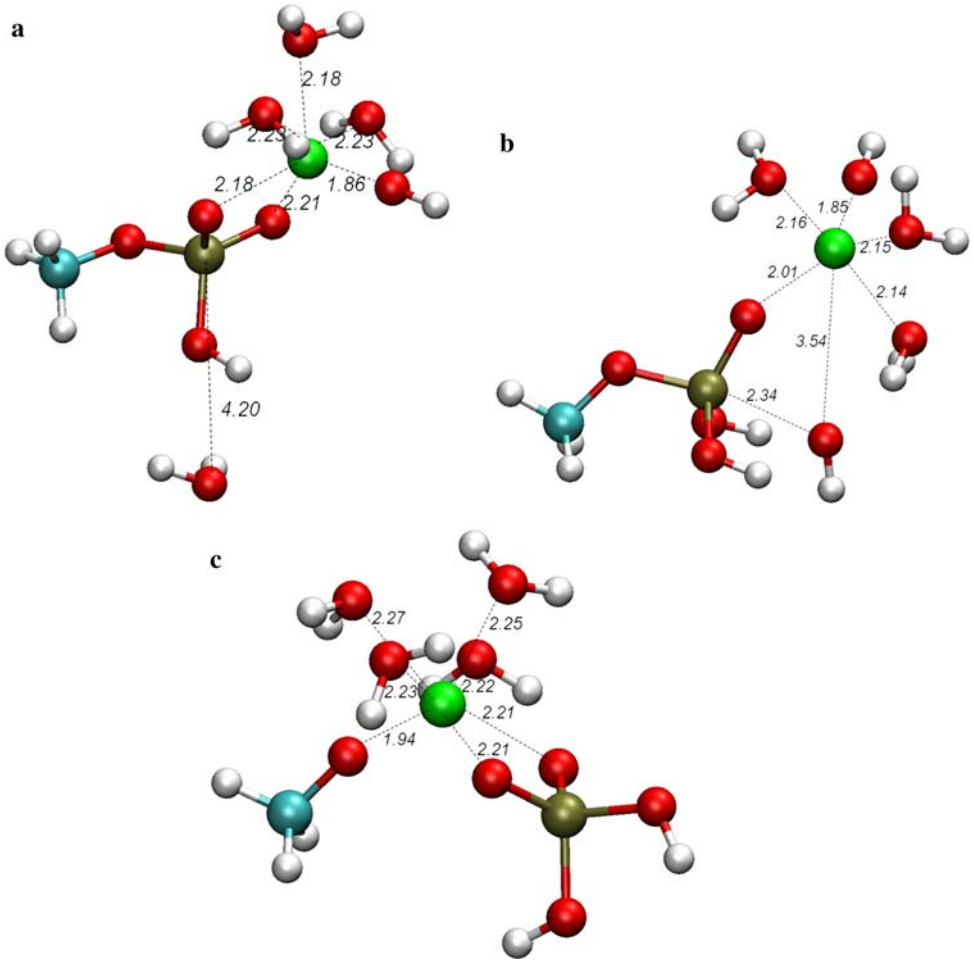

Fig. 12 Structures of the model reaction 3 obtained with AM1/d-P,Mg. **a** Reactant complex; **b** transition state; **c** product complex. Distances are given in Å

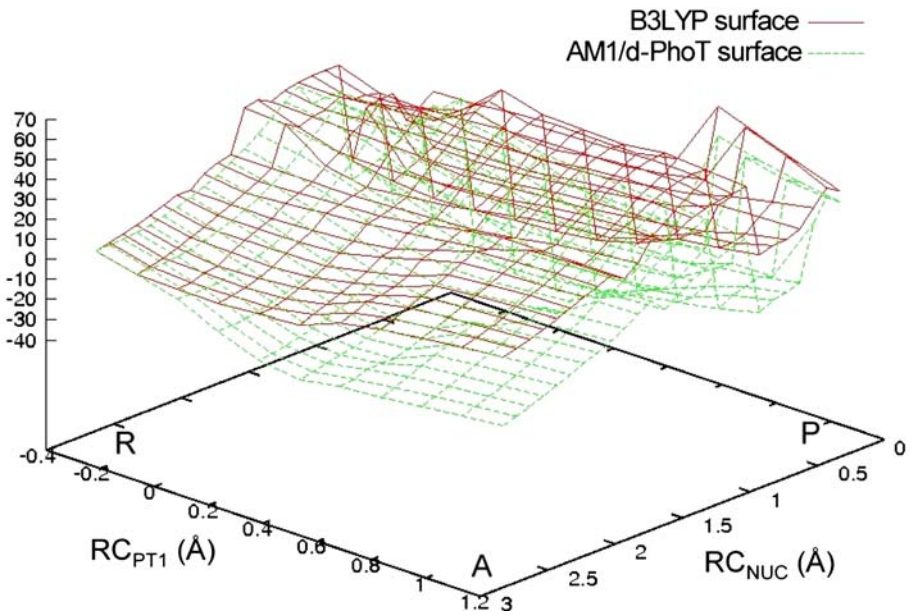

Fig. 13 2D potential energy surfaces corresponding to the gas-phase model of the associative mechanism at the B3LYP/MG3S level of theory (*solid red*) and at the AM1/d-PhoT level (*dashed green*). The R letter corresponds to the reactant complex region. The P letter corresponds to the product complex region. The A letter corresponds to one of the edges in which the bonds are stressed and the structures loose the chemical meaning

in the area surrounding the reactant complex (on the very left), but, moving away from this region, the two surfaces start to differ evidently, and the difference becomes even more important when approaching the edge marked with an A and the opposite one, where the bonds are highly stressed.

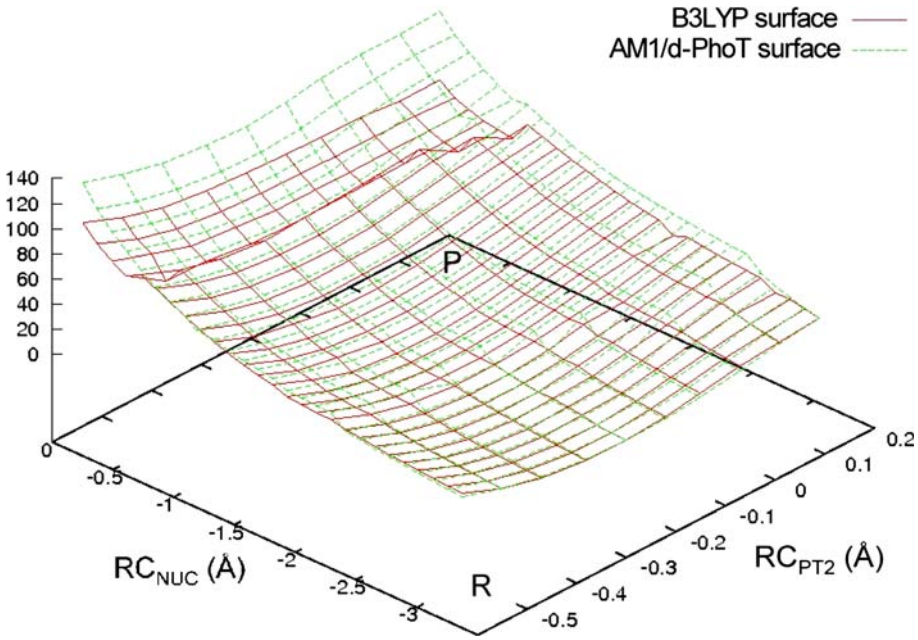

Fig. 14 2D potential energy surfaces of the gas-phase model of the dissociative model at the B3LYP/MG3S level of theory (*solid red*) and at the AM1/d-PhoT level (*dashed green*). The R letter corresponds to the reactant complex region. The P letter corresponds to the product complex region

On the other hand, Fig. 14 shows the results obtained in gas-phase at the two levels of theory for the 35 atoms model on structures picked up from the AM1/d-PhoT/MM PES of the dissociative mechanism. In this case, in agreement with the results obtained by adding the error calculated for the model reactions 2 and 3 (Fig. 9 and Table 8), the AM1/d-PhoT parameterization gives higher relative energies with respect to the overall reactant complex, 2-D PES. In this case, the average correction for the chemically relevant structures of the 2-D PES is -6.3 kcal mol^{-1}. In Fig. 14 it can be seen that the two surfaces are in quite good agreement in the entire area, and that just when we reach the outermost regions (excluding the reactant complex area), where the bonds are stressed and the configurations have no longer chemical sense, the differences become quite big.

It has to be noticed that, in both cases, the transition state is in a region rather affected by the correction: it is in the central area in Fig. 13 and near the left corner in Fig. 14.

As already said above, a correction 2-D surface can be obtained from the difference between the HL and the LL 2-D PESs, by fitting the values through a two-dimensional cubic spline function. Then, new QM/MM potential energy surfaces for both reaction mechanisms can be obtained with the incorporation of the corresponding correction 2-D surface. The corrected potential will have the form:

$$V(\mathrm{RC_{NUC}}, \mathrm{RC_{PT}}) = V(\mathrm{QM_{AM1/d}/MM}) + (E_{\mathrm{DFT}} - E_{\mathrm{AM1/d}}) \quad (12)$$

where the last term in the equation is the result of fitting the difference between the DFT and the AM1/d surfaces to a cubic spline function of the $\mathrm{RC_{NUC}}$ and the $\mathrm{RC_{PT}}$ coordinates. This work has been initiated in our laboratory to undertake a complete dynamics study of the PKA reaction, which includes the calculation of free energy profiles.

In summary, this study provides relevant information for any future QM/MM study of the phosphoryl transfer reaction catalyzed by PKA. It highlights the most accurate methods to be used when modeling it with models of different size among several levels of theory. In addition, we have proposed a way to correct the energy provided by the most accurate semiempirical parameterization with respect to the most accurate DFT method, which will allow us to carry out reliable QM/MM calculations of the complete biological system.

Acknowledgments We are grateful for financial support from the Spanish "Ministerio de Ciencia e Innovación" through Project CTQ2008-02403/BQU and "Generalitat de Catalunya" (2005SGR00400).

References

1. Hart JC, Sheppard DW, Hillier IH, Burton NA (1999) What is the mechanism of phosphoryl transfer in protein kinases? A hybrid quantum mechanical/molecular mechanical study. Chem Comm: 79
2. Sheppard DW, Burton N, Hillier IH (2000) Ab initio quantum mechanical/molecular mechanical studies of the mechanism of the enzymes protein kinase and thymidine phosphorylase. J Mol Struct 506:35
3. Hirano Y, Hata M, Hoshino T, Tsuda M (2002) Quantum chemical study on the catalytic mechanism of the C-subunit of cAMP-dependent protein kinase. J Phys Chem B 106:5788
4. Cavalli A, DeVivo M, Recanatini M (2003) Density functional study of the enzymatic reaction catalyzed by a cyclin-dependent kinase. Chem Comm: 1308

5. Hutter MC, Helms V (2003) Mechanism of phosphoryl transfer in kinases investigated by semiempirical calculations. Int J Quant Chem 95:479
6. Valiev M, Kawai R, Adams JA, Weare JH (2003) The role of the putative catalytic base in the phosphoryl transfer reaction in a protein kinase: first-principles calculations. J Am Chem Soc 125:9926
7. Diaz N, Field MJ (2004) Insights into the phosphoryl-transfer mechanism of cAMP-dependent protein kinase from quantum chemical calculations and molecular dynamics simulations. J Am Chem Soc 126:529
8. Cheng Y, Zhang Y, McCammon JA (2004) How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: an ab initio QM/MM study. J Am Chem Soc 13:206
9. Henkelman G, LaBute MX, Tung CS, Fenimore PW, McMahon BH (2005) Conformational dependence of a protein kinase phosphate transfer reaction. Proc Natl Acad Sci 102:15347
10. Cheng Y, Zhang Y, McCammon JA (2006) How does activation loop phosphorylation modulate catalytic activity in the cAMP-dependent protein kinase: a theoretical study. Prot Sci 15:672
11. Klähn M, Rosta E, Warshel A (2006) On the mechanism of hydrolysis of phosphate monoesters dianions in solutions and proteins. J Am Chem Soc 128:15310
12. DeVivo M, Cavalli A, Carloni P, Recanatini M (2007) Computational study of the phosphoryl transfer catalyzed by a cyclin-dependent kinase. Chem Eur J 13:8437
13. Montenegro M, Garcia-Viloca M, González-Lafont À, Lluch JM (2007) Comparative study of the prereactive protein kinase A Michaelis complex with kemptide substrate. J Comput Aided Mol Des 21:603
14. Valiev M, Yang J, Adams JA, Taylor SS, Weare JH (2007) Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanichs simulations. J Phys Chem B 111:13455
15. Jia R, Yang LJ, Yang SY (2008) Binding energy contributions of the conserved bridging water molecules in CDK2-inhibitor complexes: a combined QM/MM study. Chem Phys Lett 460:300
16. Szarek P, Dyguda-Kazimierowicz E, Tachibana A, Sokalski WA (2008) Physical nature of intermolecular interactions within cAMP-dependent protein kinase active site: differential transition state stabilization in phosphoryl transfer reaction. J Phys Chem B 112:11819
17. Hutter MC, Helms V (1999) Influence of the key residues on the reaction mechanism of the cAMP-dependent protein kinase. Prot Sci 8:2728
18. Thiel W, Vojtyuk AA (1996) Extension of MNDO to d orbitals: Parameters and results for the second-row elements and for the zinc group. J Phys Chem 100:616
19. Lopez X, York DM (2003) Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus. Theor Chem Acc 109:149
20. Winget P, Horn AHC, Selçuki C, Martin B, Clark T (2003) AM1* parameters for phosphorus, sulfur and chlorine. J Mol Model 9:408
21. Imhof P, Noé F, Fischer S, Smith JC (2006) Am1/d parameters for magnesium in metalloenzymes. J Chem Theory Comput 2:1050
22. Nam K, Cui Q, Gao J, York DM (2007) Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. J Chem Theory Comput 3:486
23. Range K, Riccardi D, Cui Q, Elstner M, York DM (2005) Benchmark calculations of proton affinities and gas-phase basicities of molecules important in the study of biological phosphoryl transfer. Phys Chem Chem Phys 7:3070
24. Zhou J, Adams JA (1997) Is there a catalytic base in the active site of cAMP-dependent protein kinase? Biochemistry 36:2977
25. Madhusudan, Akamine P, Xuong NH, Taylor SS (2002) Crystal structure of a transition state mimic of the catalytic subunit of cAMP-dependent protein kinase. Nat Struct Biol 9:273
26. Cole PA, Courtney AD, Shen K, Zhang ZS, Qiao YF, Lu W, Williams DM (2003) Chemical approaches to reversible protein phosphorylation. Acc Chem Res 36:444
27. Montgomery JA Jr, Frisch MJ, Ochterski JW, Petersson GA (1999) A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. J Chem Phys 110:2822
28. Montgomery JA Jr, Frisch MJ, Ochterski JW, Petersson GA (2000) A complete basis set model chemistry. VII. Use of the minimum population localization method. J Chem Phys 112:6532
29. Curtiss LA, Raghavachari K, Redfern PC, Rassolov V, Pople JA (1998) Gaussian-3 (G3) theory for molecules containing first and second-row atoms. J Chem Phys 109:7764
30. Lynch BJ, Thrular DG (2003) Robust and affordable multicoefficient methods for thermochemistry and thermochemical kinetics: the MCCM/3 suite and SAC/3. J Phys Chem A 107:3898
31. Lynch BJ, Thrular DG (2002) What are the best affordable multicoefficient strategies for calculating transition state geometries and barrier heights? J Phys Chem A 106:842
32. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785
33. Becke AD (1993) Density-functional thermochemistry. 3. The role of exact exchange. J Chem Phys 98:5648
34. Zhao Y, Truhlar DG (2004) Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and noncovalent interactions: The MPW1B95 and MPWB1K models and comparative assessments for hydrogen bonding and van der Waals interactions. J Phys Chem 108:6908
35. Lynch BJ, Thrular DG (2003) Effectiveness of diffuse basis functions for calculating relative energies by density functional theory. J Phys Chem A 107:1384
36. González-Garcia N, González-Lafont À, Lluch JM (2004) Electronic structure study of the initiation routes of the dimethyl sulfide oxidation by OH. J Comput Chem 26:569
37. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, 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 JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, Revision C.02 Gaussian, Inc., Wallingford, CT
38. Xu D, Guo H, Liu Y, York DM (2005) Theoretical studies of dissociative phosphoryl transfer in interconversion of phosphoenolpyruvate to phosphonopyruvate: solvent effects, thio effects, and implications for enzymatic reactions. J Phys Chem B 109:13827
39. Martin JML, deOliveira G (1999) Towards standard methods for benchmark quality ab initio thermochemistry—W1 and W2 theory. J Chem Phys 111:1843
40. Fast PL, Thrular DG (2000) MC-QCISD: multi-coefficient correlation method based on quadratic configuration interaction with single and double excitations. J Phys Chem A 104:6111

41. González-Lafont À, Truong TN, Truhlar DG (1991) Direct dynamics calculations with neglect of diatomic differential overlap molecular-orbital theory with specific reaction parameters. J Phys Chem 95:4618
42. Dewar MJS, Thiel W (1977) J Am Chem Soc 99:4899
43. Stewart JJP (1989) Optimization of parameters for semiempirical methods. 1. Method. J Comput Chem 10:209
44. Stewart JJP (1989) Optimization of parameters for semiempirical methods. 2. Applications. J Comput Chem 10:221
45. Dewar MJS, Zoebisch EG, Healy E, Stewart JJP (1985) J Am Chem Soc 107:3902
46. Stewart JJP (2007) Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. J Mol Model 13:1173
47. Hutter MC, Reimers JR, Hush NS (1998) Modeling the bacterial photosynthetic reaction center. 1. Magnesium parameters for the semiempirical AM1 method developed using a genetic algorithm. J Phys Chem B 102:8080
48. Tuber-Brohman I, Guimaraes CRW, Jorgensen W (2005) Extension of the PDDG/PM3 semiempirical molecular orbital method to sulfur, silicon, and phosphorus. J Chem Theory Comput 1:817
49. Thiel W (1998) MNDO97, version 5.0. University of Zürich, Zürich, Switzerland
50. McQuarrie DA (1996) Statistical mechanics. Harper Collins Publishers
51. Proust-DeMartin F, Dumas R, Field MJ (2000) A hybrid-potential free-energy study of the isomerization step of the acetohydroxy acid isomeroreductase reaction. J Am Chem Soc 122:7688
52. Ruiz-Pernía JJ, Silla E, Tuñon I, Martí S, Moliner V (2004) Hybrid QM/MM potentials of mean force with interpolated corrections. J Phys Chem B 108:8427
53. Ferrer S, Tuñon I, Martí S, Moliner V, Garcia-Viloca M, González-Lafont À, Lluch JM (2006) A theoretical analysis of rate constants and kinetic isotope effects corresponding to different reactant valleys in lactate dehydrogenase. J Am Chem Soc 128:16851
54. Wong KY, Gao J (2007) The reaction mechanism of paraoxon hydrolysis by phosphotriesterase from combined QM/MM simulations. Biochemistry 46:13352
55. Montenegro M, Garcia-Viloca M, González-Lafont À, Lluch JM (in preparation)
56. MacKerell AD Jr, Bashford D, Bellott M, Dumbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-MacCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B 102:3586