Specific parametrisation of a hybrid potential to simulate reactions in phosphatases

Guilherme Menegon Arantes*† and Michel Loos

Received 19th August 2005, Accepted 26th October 2005
First published as an Advance Article on the web 17th November 2005
DOI: 10.1039/b511805k

Phosphatases are key biomolecules because they regulate many cellular processes. These enzymes have been studied for many years, but there are still doubts about the catalytic mechanism. Computer simulations can be used to shed light on these questions. Here we develop a new and specific parametrisation, and present extensive tests of a hybrid potential that can be used to reliably simulate reactions catalysed by phosphatases. High level *ab initio* data for phosphate ester thiolysis and alcoholysis is used in the training set. The parametrised quantum mechanical Hamiltonian reproduces *ab initio* energies with a root mean-squared deviation of 3 kcal mol^{-1} for species along the pathway of various phosphate ester reactions. Preliminary results for simulation with the calibrated hybrid potential of catalysis by the phosphatase VHR indicate the calculated reaction barriers are in very good agreement with experiment.

1. Introduction

Protein tyrosine phosphatases (PTPs) catalyse the hydrolysis of phosphotyrosine from other proteins and, hence, regulate important biochemical processes, including cellular growth and differentiation.[1] The first step of catalysis is the nucleophilic attack from a PTP cysteine side chain toward the phosphate ester substrate, with a possible H^+ transfer from a general acid to the leaving group. A thiophosphorylated PTP intermediate is formed and the substrate is dephosphorylated (eqn (1)). The PTP intermediate is hydrolysed and the free enzyme is regenerated in the subsequent step (eqn (2)).[1]

$$E\cdots ROPO_3^{2-} \xrightarrow{k_2} E-PO_3^{2-} + ROH \qquad (1)$$

$$E-PO_3^{2-} + H_2O \xrightarrow{k_3} E + HOPO_3^{2-} \qquad (2)$$

There are some open questions about the first reaction step (eqn (1)) regarding, for example, the protonation state of the species involved in catalysis[2] and the reason for inactivity of certain mutants.[3] These two questions are difficult to access experimentally, so computer simulations can be very helpful.

Simulation of reactions in proteins can be carried out with hybrid potentials of quantum mechanics (QM) and molecular mechanics (MM).[4,5] Free-energy profiles, or potentials of mean force (PMF), can be easily obtained for the simulated reaction if a reasonably fast method is used to describe the QM part of the hybrid potential. A common choice has been a semiempirical method based on the neglect of diatomic differential overlap (NDDO) approximation,[6] and many enzymatic reactions have been studied with this approach.[7,8] However, standard parametrisations of NDDO semiempirical methods are rather inaccurate for calculations involving phosphorus and sulfur atoms.[9]

In this paper we present a new and specific set of NDDO parameters that can be used to accurately simulate thiolysis and alcoholysis of phosphate esters, as in reaction (1). Experimental properties involving phosphate esters reactions would be necessary for the parametrisation, however they are very scarce.[10] We followed the approach proposed by Rossi and Truhlar,[11] later applied in several works,[12–14] in which *ab initio* calculated properties are used as references in the training or data set for the parametrisation.

Hybrid QM/MM potentials are based in other two crucial approximations regarding the non-bonding interactions between QM and MM regions, and the treatment of the covalent bond in the frontier of the two regions.[15] We tested both by comparing the hybrid potential to fully QM *ab initio* energies for bimolecular complexes and reactions which are models of the protein–substrate complex in PTPs.

The remainder of the paper is organised as follows. In section 2.1 we present the data set and error function used to obtain the new specific NDDO parameters. In section 2.2 we give details about the optimisation algorithm used in the fitting procedure, and in section 3 we present the new parameters and their performance in the calculation of properties relevant to the chemistry of phosphatases. In section 3.1 details of the reference set used to test the accuracy of the hybrid QM/MM potential are given. Results for comparisons of non-bonding interactions are presented in section 3.2 and of the covalent interface are shown in section 3.3. A brief application of the calibrated hybrid potential is presented in section 4 and we end with conclusions in section 5.

2. Semiempirical Hamiltonian

2.1. Data set, parameters and error function

The training set included all stationary points of the eight phosphate ester reactions shown in Fig. 1. Both the associative

Instituto de Química, Universidade de São Paulo, Av. Lineu Prestes 748, São Paulo, SP, 05508-900, Brasil.
E-mail: gma@dinamicas.art.br
† Current address: Department of Chemistry, University of Warwick, Coventry, UK CV4 7AL.

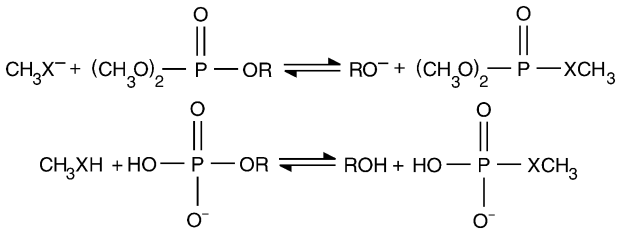

Fig. 1 Scheme of the reactions of thiolysis (X = S) and alcoholysis (X = O) of aryl (R = phenyl, Ph) and alkyl (R = CH_3) phosphate esters.

and dissociative mechanisms of monoester reactions were included.[10] Details about the species, energy profiles and comparisons to experiment were described elsewhere.[10,16] Three additional sets were included: reactions related to H^+ transfer in phosphatases (Table 1); reactions of dianionic phosphate esters (see below); and species which geometry was incorrectly calculated by the semiempirical Hamiltonian obtained during the fitting (see below).

All reference energies (E^{ref}) were obtained for species in the gas-phase with the restricted MP2 level of theory and the 6-311+G(2df,2p) basis set.[10] All semiempirical energies (E^{NDDO}) were obtained in the *ab initio* geometry, excluding E_{corr}^{NDDO} and E_{opt}^{NDDO} (see below). Zero-point energies, and thermal or entropic corrections were not added to E^{ref}, E^{NDDO} or to the energies calculated with the QM/MM potential. Reference geometries were fully optimised in the MP2/6-31+G(d) level. All *ab initio* computations were performed with the GAUSSIAN98[23] package. All semiempirical computations were carried with the GEOMOP program[24] in the restricted closed-shell formalism. For semiempirical geometry optimisations, a quasi-Newton algorithm[25] with analytic gradients was used, and up to 20 optimisation steps were allowed for minima and 10 steps for TS.

The optimised error function was defined as:

$$
\begin{aligned}
P = {} & W_{ene}\sum_{i}^{N}(\Delta E_i^{NDDO} - \Delta E_i^{ref})^2 \\
& + W_{norm}\sum_{i}^{N} \| g_i^{NDDO} \|^2 + W_{IP}\sum_{i}^{N}(IP_i^{NDDO} - IP_i^{ref})^2 \\
& + W_{opt}\sum_{i}^{N_{opt}}(\Delta E_{opt,i}^{NDDO} - \Delta E_i^{ref})^2 \\
& + W_{pt}\sum_{i}^{N_{pt}}(\Delta E_{pt,i}^{NDDO} - \Delta E_{pt,i}^{ref})^2 \\
& + W_{dian}\sum_{i}^{N_{dian}}(\Delta E_{dian,i}^{NDDO} - \Delta E_{dian,i}^{ref})^2 \\
& + W_{corr}\sum_{i}^{N_{corr}}(\Delta E_{corr,i}^{NDDO} - \Delta E_{corr,i}^{ref})^2
\end{aligned}
\qquad (3)
$$

ΔE is the relative energy for $N = 70$ stationary points of the phosphate ester reactions in the data set (Fig. 1). g^{NDDO} is the semiempirical gradient and IP is the highest occupied molecular orbital (HOMO) energy. ΔE_{pt} is the reaction energy for $N_{pt} = 12$ proton transfer reactions (Table 1). ΔE_{dian} is the relative energy for $N_{dian} = 18$ dianionic species along the pathway for breakage of the P–X bond (X = O, S) in $PhOPO_3^{2-}$, $CH_3OPO_3^{2-}$ and $CH_3SPO_3^{2-}$, producing PO_3^- and phenoxyde, CH_3O^- and CH_3S^-, respectively. Six points with different constrained P–X bond lengths were used for each reaction. ΔE_{opt}^{NDDO} is the relative energy for $N_{opt} = 43$ geometries *reoptimised* with the semiempirical Hamiltonian (only reactants, products, TS and intermediates of the reactions shown in Fig. 1 were reoptimised). Finally, ΔE_{corr} is the relative energy of the following $N_{corr} = 8$ species in the *semiempirical* geometry: $CH_3OPO_3H^-$, $PhOPO_3H^-$, one intermediate and three TS for triester reactions (B:I, C:TS1, D:TS1 and D:TS2 in ref. 10), and two monoester TSs (E:TS2a and G:TS1a in ref. 10). Weights in $(mol\ kcal^{-1})^2$ were: $W_{ene} = 1$, $W_{norm} = 3$, $W_{IP} = 4$, $W_{opt} = 10$, $W_{pt} = 4$, $W_{dian} = 3$, $W_{corr} = 3$.

The parameters $U_{\mu\mu}$, ζ_μ, β_μ, α, $g_{\mu\nu}$ and $h_{\mu\nu}$, were optimised for all elements (CHOPS), except carbon which had only $U_{\mu\mu}$, $g_{\mu\nu}$ and $h_{\mu\nu}$ optimised (μ, ν = s, p orbitals). Parameters could vary from the original values[26–28] up to ±30% for hydrogen and carbon, 40% for oxygen and phosphorus and 50% for sulfur. The MNDO parameters were used instead of AM1 [29] or PM3 [30] because only molecules with sulfur in the −2 oxidation state were included in the data set of the original parametrisation.[28] We consider that data sets with the same

Table 1 H^+ transfer reactions in gas-phase

Reaction	ΔE^{ref}	ΔE^{exp}
$CH_3COOH + CH_3O^- \rightleftharpoons CH_3OH + CH_3COO^-$	−36.9	−33.4 ± 3.2 [17,18]
$CH_3COOH + PhO^- \rightleftharpoons PhOH + CH_3COO^-$	−1.7	−1.8 ± 2.8 [17,19]
$CH_3COOH + CH_3OPO_3H^- \rightleftharpoons CH_3OPO_3H_2 + CH_3COO^-$	19.3	16.8 ± 6.9 [17,20]
$CH_3COOH + PhOPO_3H^- \rightleftharpoons PhOPO_3H_2 + CH_3COO^-$	28.0	29 ± 9 [17,21]
$CH_3SH + PhOPO_3H^- \rightleftharpoons PhOPO_3H_2 + CH_3S^-$	36.7	38 ± 8 [22,21]
$CH_3SH + CH_3O^- \rightleftharpoons CH_3S^- + CH_3OH$	−28.1	−24.2 ± 3.0 [22,18]
$CYS + SER^- \rightleftharpoons CYS^- + SER$[a]	−23.9	
$2CH_3OPO_3H^- \rightleftharpoons CH_3OPO_3H_2 + CH_3OPO_3^{2-}$	125.4	
$CH_3S^- + PhOPO_3H^- \rightleftharpoons CH_3SH + PhOPO_3^{2-}$	78.8	
$CH_3S^- + CH_3OPO_3H^- \rightleftharpoons CH_3SH + CH_3OPO_3^{2-}$	97.4	
$CH_3O^- + PhOPO_3H^- \rightleftharpoons CH_3OH + PhOPO_3^{2-}$	69.3	
$CH_3O^- + CH_3OPO_3H^- \rightleftharpoons CH_3OH + CH_3OPO_3^{2-}$	50.8	

[a] Aminoacids were modeled in the neutral form.

atom in different oxidation states deteriorate results obtained with NDDO semiempirical methods.[31]

Three Gaussian functions were added to the core–core repulsion function for each element, except carbon.[29] The Gaussian parameters were allowed to vary between ±0.02 eV Å for K_i, 0 Å^{-2} and the respective AM1 values for L_i, and from 0 to 6.0 Å for M_i. Derived parameters D_1, D_2 and ρ_a (a = 0, 1 and 2) were recalculated for each optimised set.[6,31] A total of 84 parameters was simultaneously optimised.

2.2. Optimisation algorithm: GA + simplex

A sequential combination of genetic algorithm (GA)[32] and simplex[33] was used to minimise the error function (eqn (3)). This sequential algorithm was already described elsewhere.[34] The GA source code written by Carroll[35] and the Nelder and Mead simplex in ref. 33 section 10.4 were used.

The fitting procedure was initiated with a randomly generated population, with 150 individuals. Each chromosome was represented by 8 to 14 binary digits. Normal and creep mutation probabilities were both 0.08. Niching and elitism were activated.[32] Single point crossover with a 0.8 probability was used to generate descendents. After evolution during a fixed number of generations, the 85 fittest individuals from the last generation were used as initial vertices in the simplex algorithm. The fittest individual generated during all the GA evolution was included because elitism was activated. The total number of GA generations was 1.5×10^4.

The number of simplex cycles performed was the smallest between a fixed number or until the difference between two successive error function evaluations was smaller than 10^{-5}. The simplex algorithm options were: α = 1.0, β = 0.5 and γ = 2.0. The total number of simplex steps was 4.3×10^4.

2.3. Performance of the semiempirical parameters

The system specific parameter set obtained after the optimisation procedure was denominated MNDO + G/CHOPS and is shown in Table 2.

Deviations observed in all properties calculated with the MNDO + G/CHOPS set (Table 3) are significantly smaller than the deviations of original semiempirical sets and the set used by Gao *et al.* (AM1 parameters for H, C and O, and MNDO parameters to P and S).[36]

A comparison between the reference potential energy surface (PES) and the semiempirical PES is best given by ΔE_{opt}. The root mean-squared deviation (RMSD) for this property is 1.75 kcal mol^{-1} (Table 3) for phosphate monoester reactions, indicating that a high accuracy can be expected for calculations with the MNDO + G/CHOPS set. The calibration was more efficient for monoesters than for triesters because the data set used in the optimisation procedure was mostly composed of monoesters and their reactions.

The RMSD for the bond lengths, angles and dihedrals was calculated for N_{opt} = 43 molecules. Table 3 shows only the medium ($\overline{RMSD}$) and the maximum ($RMSD_{max}$) deviations. Although no geometric information was explicitly included in

Table 2 Specific semiempirical MNDO + G/CHOPS parameters for H, C, O, P and S

	MNDO + G/CHOPS				
Parameters and unities	H	C	O	P	S
U_{ss}/eV	−13.320 770	−52.772 930	−81.496 830	−29.926 330	−107.17 470
U_{pp}/eV		−44.135 150	−70.963 590	−46.743 540	−56.396 210
ζ_s/bohr	1.126 840	1.787 537	3.334 590	1.792 970	3.570 580
ζ_p/bohr		1.787 537	2.467 500	1.715 640	1.859 890
β_s/eV	−6.758 660	−18.985 044	−43.021 230	−6.393 650	−16.671 710
β_p/eV		−7.934 122	−25.674 230	−4.632 620	−9.115 900
α/Å^{-1}	2.958 707	2.546 388	3.237 480	2.385 950	2.492 397
g_{ss}/eV	16.020 432	11.495 681	13.464 052	11.188 208	12.390 890
g_{sp}/eV		11.949 816	9.546 084	6.267 279	11.013 247
g_{pp}/eV		12.851 185	16.368 546	10.974 892	9.760 673
g_{p2}/eV		11.683615	13.129 728	9.718 182	8.751 465
h_{sp}/eV		2.334 164	3.557 707	2.602 274	2.379 333
D_1/bohr		0.807 466	0.470 220	1.149 910	0.516 220
D_2/bohr		0.685 158	0.496 350	0.975 330	0.899 690
ρ_0/bohr	0.849 220	1.183 490	1.010 470	1.216 010	1.097 980
ρ_1/bohr	0.849 220	0.827 950	0.508 370	0.949 580	0.637 340
ρ_2/bohr	0.849 220	0.762 300	0.436 100	0.942 970	0.961 230
K_1/eV Å	−0.015 119	0.000 000	−0.003 600	−0.000 288	−0.017 855
K_2/eV Å	0.000 432	0.000 000	−0.001 584	0.000 720	0.003 168
K_3/eV Å	0.003 456	0.000 000	0.011 375	0.001 008	0.003 600
K_4/eV Å	0.000 000	0.000 000	0.000 000	0.000 000	0.000 000
L_1/Å^{-2}	1.915 663		4.772 553	0.367 927	2.545 350
L_2/Å^{-2}	0.157 448		0.286 400	0.185 017	4.576 466
L_3/Å^{-2}	3.049 013		2.407 364	0.170 983	2.453 788
M_1/Å	2.175 055		5.076 263	0.132 665	1.083 146
M_2/Å	0.054 415		5.277 657	0.310 188	2.670 814
M_3/Å	3.059 781		1.533 687	0.145 587	1.762 779

The ρ_a parameters can be transformed to the AX parameters used in semiempirical programs based in the MOPAC code[6] using the relation: $AX = 0.5/\rho_a$ (a = 0, 1 or 2 and X = M, D or Q, respectively). All other parameters presented are in the unities directly used by the MOPAC code.

Table 3 Modulus of maximum deviation (max) and root mean-squared deviation (RMSD) for properties calculated using different NDDO semiempirical parametrisations in relation to MP2/6-311 + G(2df,2p)//MP2/6-31 + G(d) reference values

			Semiempirical parametrisation				
Property	Deviation	N^a	MNDO	AM1	PM3	Gao[b]	MNDO + G/CHOPS
ΔE	RMSD	70	27.06	17.06	12.11	14.94	3.97
IP^c		70	0.70	0.62	0.90	0.72	0.42
ΔE_{dian}		18	8.17	4.15	10.41	3.34	2.17
ΔE_{pt}	RMSD	12	11.09	12.54	16.76	9.63	2.71
	Max		20.97	20.04	29.12	14.27	5.00
	Num[d]		11	09	09	09	3
ΔE_{opt}	RMSD[e]	43	15.80	16.22	10.38	14.73	2.79
	Tri	16	12.82	16.11	8.88	17.61	4.09
	Mono	27	17.85	16.87	11.53	13.34	1.75
	Max		41.48	28.71	27.15	27.72	7.31
	Num[d]		34	42	36	40	7
Bonds[f]	RMSD	668	0.16	0.05	0.11	0.05	0.03
	RMSD_{max}		0.63	0.24	0.53	0.21	0.08
Angles	$\overline{\text{RMSD}}$	629	9.33	5.56	5.46	4.04	3.41
	RMSD_{max}		21.30	28.40	18.98	25.98	8.84
Dihedrals	$\overline{\text{RMSD}}$	590	27.30	20.47	21.79	23.06	17.30
	RMSD_{max}		80.58	53.00	58.49	83.14	35.93

Definitions of properties given in Section 2.1. RMSD and max values in kcal mol^{-1}.[a] Number of species or reactions considered. [b] AM1 parameters for C, H and O, and MNDO parameters for P and S.[36] [c] Values in eV. [d] Number of species with deviation greater than 3 kcal mol^{-1}. [e] Total RMSD and RMSD between the phosphate triesters (tri) and monoesters (mono) reactions. [f] Mean RMSD ($\overline{\text{RMSD}}$) and maximum RMSD (RMSD_{max}) for the 43 geometries optimized by the semiempirical methods (see text for details). Values given in Å for bonds, and in ° for angles.

the error function (eqn (3)), it is noticeable that the optimised MNDO + G/CHOPS set calculates internal coordinates with significantly smaller deviations.

3. Hybrid QM/MM contributions

3.1. Data set and computational details

Hydrogen bonding bimolecular complexes were studied. Models of the interaction between the substrate phosphate group and the protein main chain, and its active site ARG side-chain[1] are shown in Fig. 2 and 3, respectively. Models of solvent water interaction with the ASP general acid and the substrate leaving group are shown in Fig. 4. Equilibrium geometries were obtained by optimisations at the MP2/6-31 + G* level. Orientations similar to those observed in PTP complexes were chosen and preserved by fixing the angles explicitly shown in Figs. 2–4. Reference energies for complex formation ($\Delta E^{\text{ref}}_{\text{comp}}$) were calculated in the MP2/6-311 + G(2df,2p) level for different hydrogen bond distances (R, shown in dotted line in the figures). Basis set superposition error (BSSE) was corrected following the counter-poise method.[37] The GAUSSIAN98 program system was used in all *ab initio* computations.

Hybrid QM/MM energies for complex formation were calculated in the same geometries according to:

$$\Delta E^{\text{hyb}}_{\text{comp}}(R) = E^{\text{hyb}}_{\text{AB}}(R) - E^{\text{QM}}_{\text{A}} - E^{\text{MM}}_{\text{B}} \qquad (4)$$

where molecule A was in the QM region, molecule B is in the MM region and $E^{\text{hyb}}_{\text{AB}}$ is the hybrid potential energy for the AB complex. There was no covalent bond between the two regions. In Fig. 2 and 3, PO_3^- was in the QM region and formamide or guanidinium were in the MM region. In Fig. 4, water was in the MM region and phenol or acetic acid were in the QM region.

The link-atom approach[15,38] was tested by calculating the reaction energy of three isolated model reactions (Table 4). Amino acids were modelled in the neutral form. The CH_3OH, PO_3^- and the amino acid side-chains were treated in the QM region. The rest of each amino acid was in the MM region. Hence, the frontier between the two regions was localised in the C_α–C_β bond.

The QM region was described by the MNDO + G/CHOPS calibrated Hamiltonian (section 2.3). The OPLS-AA[39] force field and the water TIP3P[40] potential were used for the MM

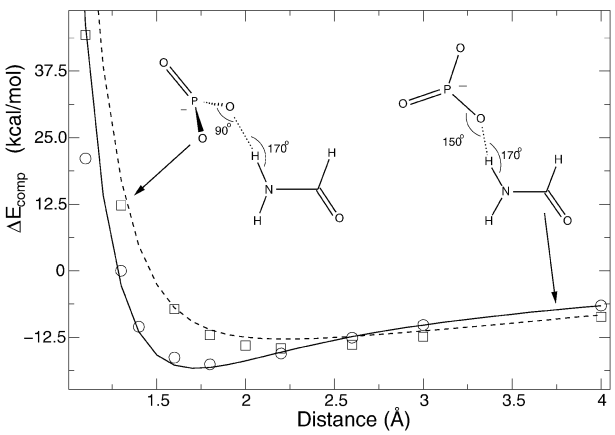

Fig. 2 Complexation energies between formamide and metaphosphate in the gasphase. Symbols indicate *ab initio* energies $\Delta(E^{\text{ref}}_{\text{comp}}$ and lines indicate hybrid potential energies ($\Delta E^{\text{hyb}}_{\text{comp}}$). Two different orientations are shown. The scanned hydrogen bond distance (R) is shown as a dotted line.

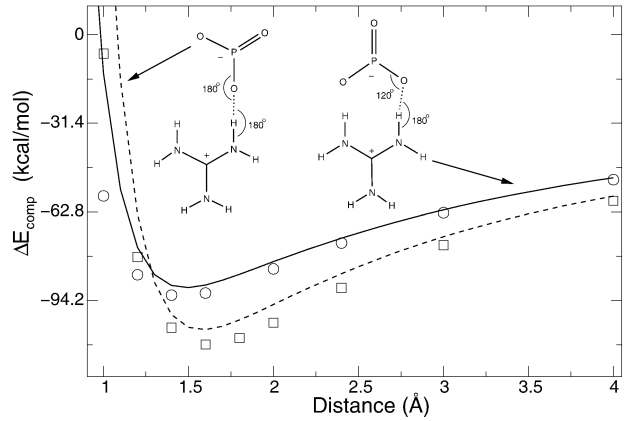

Fig. 3 Complexation energies between the guanidinium cation and metaphosphate in the gas phase. Two different orientations are shown and the legend is the same as in Fig. 2.

region. LJ parameters for the QM atoms were taken from the OPLS-AA force field and parameters for PO_3^- phosphorus and oxygen were taken from the AMBER force field.[41] All hybrid potential calculations were performed with the DYNAMO library.[38]

3.2. Bimolecular complexes

For the anionic complexes shown in Fig. 2, both shape and energy values obtained with the hybrid potential are in good agreement with the reference, especially near equilibrium distances, where deviations were *ca.* 1 kcal mol^{-1}. For zwitterionic complexes shown in Fig. 3, deviations of the hybrid potential near minima are *ca.* 5 kcal mol^{-1}. However, the shape of the hybrid potential is correct and relative deviations are smaller than 10%. The hybrid potential systematically underestimates the interaction energy for phosphate complexes.

Interaction energies are considerably smaller for the neutral complexes shown in Fig. 4. The shape and the values of the hybrid potential are satisfactory. A small deviation is seen in

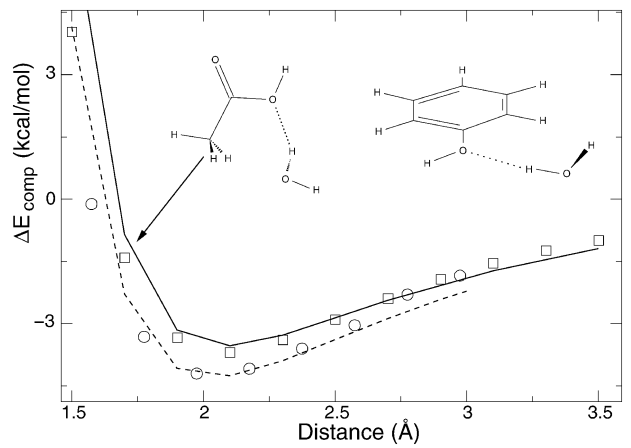

Fig. 4 Complexation energies between phenol and water (circles and dashed line), and between acetic acid and water (squares and solid line) in the gas phase. The orientation of both complexes is also shown and the legend is the same as in Fig. 2.

Table 4 Reaction energies calculated for models with a covalent bond between the QM and MM regions

	Hamiltonian[a]	
Reaction	Ref	QM/MM
$CYS–PO_3^{2-} \rightleftharpoons CYS^- + PO_3^-$	−59.2	−60.0
$SER–PO_3^{2-} \rightleftharpoons SER^- + PO_3^-$	−43.0	−45.8
$ASP + CH_3O^- \rightleftharpoons ASP^- + CH_3OH$	−39.1	−40.5

Only the amino-acid side-chain participate in the indicated reactions.[a] Ref is the reference MP2/6-31+G*//MP2/6-311+G (2df,2p) energy and QM/MM is the hybrid potential energy.

larger distances, when the hybrid potential overestimates the interactions.

For a total of 53 reference points (R) of various bimolecular complexes, the RMSD of the interaction energy obtained with the hybrid potential is 6.1 kcal mol^{-1}. The highest contribution to this deviation comes from the points in the repulsive part of the interaction ($R < 1.2$ Å). Deviations are considerably smaller near equilibrium distances (see above).

A specific parametrisation of the LJ parameters[12] was carried out without success (results not shown). Any significant improvement in minima region resulted in degradation of the repulsive region of the potential and *vice versa*. This confirms suggestions from previous works that a LJ function is not ideal for calculation of van der Waals interactions in hybrid potentials.[42]

Considering these limitations, the original OPLS-AA and AMBER parameters can be used for QM atoms with satisfactory results.

3.3. Covalent frontier

The first two reactions in Table 4 are models of side-chain phosphorylation in CYS and SER. The third reaction corresponds to a H^+ transfer between the ASP side-chain and methoxyde and is a model of the general acid catalysis. Deviations in the hybrid potential energy are smaller than 3 kcal mol^{-1} for the three model reactions. Hence, the link-atom approach is sound and does not introduce significative errors in the enzymatic simulations.

4. Phosphatase test simulation

The first step (eqn (1)) in the dephosphorylation of phenyl phosphate catalysed by the dual-specificity phosphatase VHR1 was simulated with the calibrated hybrid potential. A brief description of the simulation method and the results are given below.

Coordinates of a VHR protein–phenyl phosphate complex were generated by Bashford *et al.* from the VHR PDB structure.[43] All the ARG, LYS and HIS residues were protonated as well as the general acid (ASP92). The phenyl phosphate substrate was modelled as a dianion and the nucleophile cysteine (CYS124) in the ionised form (Fig. 5).

This complex was superposed into an equilibrated cubic water box (55.92 Å side, with 5832 water molecules) represented by the TIP3P potential. The OPLS-AA potential was used for the protein. The fully solvated system was relaxed and

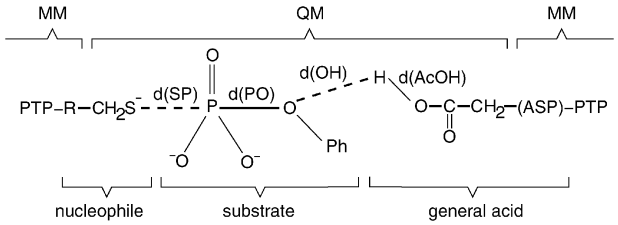

Fig. 5 Scheme of the QM and MM atoms and regions used in the enzymatic simulation.

equilibrated to 300 K by fully classical molecular dynamics runs following standard procedures.[44–46]

Positions of all atoms (protein and solvent) located more than 14.0 Å away from the phenyl phosphate phosphorus were frozen. However, interactions between frozen and moveable atoms were normally included. The system was divided into a QM region containing the full substrate, $C_\beta H_2$–S^- from CYS124, and $C_\beta H_2$–COOH from ASP92 (Fig. 5). The MM region contained the rest of the protein and all solvent. The frontier between regions was treated with the link-atom approach (section 3.1). The QM region was represented by the calibrated MNDO + G/CHOPS potential, OPLS-AA was used for the protein MM region and TIP3P was used for water, as described in section 3.1. A force-switching truncation scheme was employed with r_{on} = 10 Å and r_{off} = 13 Å for non-bonding interactions.[38]

Molecular dynamics with a velocity Verlet–Langevin algorithm was used to sample configuration space (ref. 47 section 9.3). A time interval τ = 1 fs and a friction coefficient γ = 10 ps^{-1} were used.[44,45] Potentials of mean force (PMF) were calculated for the reaction using an umbrella potential (k_{umb} = 1500 kJ mol^{-1} $Å^{-2}$)[48] to enhance sampling of the reaction coordinate $\xi = d(PO) - d(PS)$, where d(PO) and d(SP) are the distances from phosphorus to the leaving group oxygen and to the nucleophile sulfur, respectively (Fig. 5). Each window had reference ξ_i equally spaced by 0.10 Å covering the relevant range of reaction coordinate.[44,45] Equilibration time was 4 ps and data was collected during 10 ps. The WHAM[48] method was used to obtain the final PMF from the ξ occurrence values. All simulations and PMF were obtained with the DYNAMO library.

The reaction starts from the Michaelis complex ($\xi \sim -1.9$ Å, Fig. 6) and proceeds to a dissociative TS ($\xi \sim -0.3$ Å) with the P–O bond largely broken. H^+ is transferred from ASP92 to the leaving group oxygen in the TS, but almost no P–S bond is formed. The reaction products ($\xi \sim 2.3$ Å) are the thiophosphorylated VHR with phenol coordinated outside the active site. The experimental activation energy for this reaction is 15.5 kcal mol^{-1}, calculated by the transition state theory[49] from the rate constant (k_2) measured for reaction of *p*-nitrophenyl phosphate catalysed by VHR.[50] We note the rates (k_{cat}) for the catalysed reactions of *p*-nitrophenyl phosphate and of phenyl phosphate are about the same.[1,50] This value is remarkably similar to the barrier obtained from the PMF (Fig. 6), 16.4 kcal mol^{-1}, indicating that the hybrid potential parametrisation is accurate and can be used to rigorously study catalysis by phosphatases.

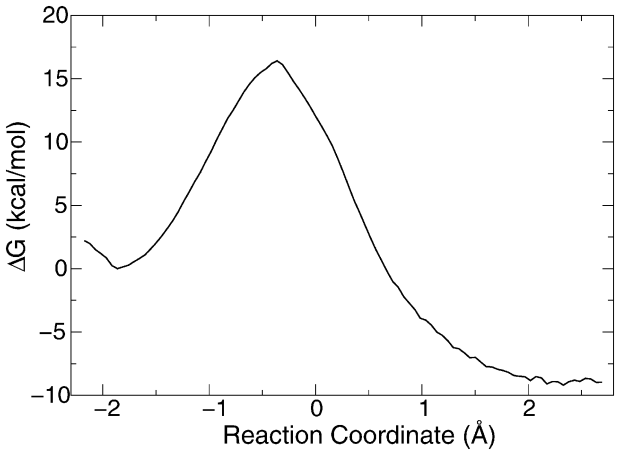

Fig. 6 Potential of mean force for the reaction of phenyl phosphate dianion catalised by VHR wild-type.

5. Conclusions

A new parametrisation of an NDDO semiempirical Hamiltonian (MNDO + G/CHOPS) that is specific for phosphate ester thiolysis and alcoholysis was presented. Because phosphate mono- and triesters were included in the training set, the parameters are useful for calculations with phosphates in any esterification state. Proton transfer reactions were also included in the data set. Hence, the parameters can be valuable in the determination of phosphate ester proton affinities. The accuracy of the new parametrisation is close to MP2/6-311 + G(2df,2p), which was the level of theory used to generate the training set. For thiolysis and alcoholysis of monoesters the energy RMSD of species in the reaction pathway is smaller than 2 kcal mol^{-1}, in comparison to the *ab initio* reference. We expect that the QM part of the hybrid potential used in section 4 will have the same accuracy.

The QM/MM interaction part of the hybrid potential was also tested. Energies near equilibrium geometries are reasonably well described. However it is difficult to evaluate the deviations introduced in the enzymatic simulation by this part of the potential because many of the modelled interactions occur simultaneously in the PTP active site. The difference of less than 1 kcal mol^{-1} between the experimental and the calculated barrier for the reaction in eqn (1) (section 4) indicates that the hybrid potential correctly describes catalysis in phosphatases.

Because semiempirical calculations are fast, the calibrated Hamiltonian can also be used in computer-intensive applications. For example, quantum mechanical effects, such as zero-point energy or tunnelling,[38] could be calculated with satisfactory quality and speed for large phosphate ester molecules. The tests and parametrisation presented here should be transferable to other phosphatases and their mutants. For example, alkaline phosphatases employ a serine nucleophile to attack phosphorus centres.[51] This reaction could be studied reliably with the MNDO + G/CHOPS Hamiltonian.

The hybrid potential described and tested here has been used to simulate the reaction in eqn (1) catalysed by two

phosphatases, VHR and CDC25B.[1] Different substrates and protonation states, as well as the VHR CYS124 → SER mutant were studied. Full details and results will be reported shortly.

Acknowledgements

GMA would like to acknowledge Prof. Martin J. Field (Grenoble, France) for freely distributing the DYNAMO library and for answering some questions about its usage, Prof. Hernan Chaimovich for initial support, Dr Luis Gustavo Dias for helpful discussions, Mr David A. Haggart for reading the manuscript, and a fellowship from FAPESP (Fundação de Amparo a Pesquisa do Estado de Sãao Paulo, project 99/7688-9).

References

1 M. D. Jackson and J. M. Denu, *Chem. Rev*., 2001, **101**(8), 2313–2340.
2 K. Kolmodin and J. Åqvist, *FEBS Lett*., 2001, **498**, 208–213.
3 G. Scapin, S. Patel, V. Patel, B. K. B and E. Asante-Appiah, *Protein Sci*., 2001, **10**, 1596–1605.
4 J. Gao, in *Reviews in Computational Chemistry*, ed. K. Lipkovitz and D. Boyd, VHC, New York, 1995, vol. 7, pp. 119–185.
5 P. Sherwood, in *Modern Methods and Algorithms of Quantum Chemistry*, ed. J. Grotendorst, John von Neumann Institute for Computing, Jülich, 2001, vol. 3, pp. 257–277.
6 J. J. P. Stewart, *J. Comput. Aided Mol. Des*., 1990, **4**, 1–45.
7 P. Amara and M. J. Field, in *Encyclopedia of Computational Chemistry*, ed. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, I. Schaefer, H. Fa and P. R. Schreiner, Wiley, Chichester, 1st edn, 1998, vol. 1, pp. 431–436.
8 M. J. Field, *J. Comput. Chem*., 2002, **23**, 48–58.
9 W. Thiel and A. A. Voityuk, *J. Phys. Chem*., 1996, **100**, 616–626.
10 G. M. Arantes and H. Chaimovich, *J. Phys. Chem. A*, 2005, **109**, 5625–5635.
11 I. Rossi and D. G. Truhlar, *Chem. Phys. Lett*., 1995, **233**, 231–236.
12 P. A. Bash, L. L. Ho, A. D. MacKerell Jr., D. Levine and P. Hallstrom, *Proc. Natl. Acad. Sci. U. S. A*., 1996, **93**, 3698–3703.
13 X. Lopez and D. M. York, *Theor. Chem. Acc*., 2003, **109**, 149–159.
14 M. Sundararajan, J. P. McNamara, I. H. Hillier, H. Wang and N. A. Burton, *Chem. Phys. Lett*., 2005, **404**, 9–12.
15 M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem*., 1990, **11**, 700–733.
16 G. Menegon, M. Loos and H. Chaimovich, *J. Phys. Chem. A*, 2002, **106**, 9078–9084.
17 J. Cumming and P. Kebarle, *Can. J. Chem*., 1978, **56**, 1–9.
18 D. Osborn, D. Leahy, E. Kim, E. deBeer and D. Neumark, *Chem. Phys. Lett*., 1998, **292**, 651–655.
19 R. Gunion, M. Gilles, M. Polak and W. Lineberger, *Int. J. Mass Spectrom. Ion Processes*, 1992, **117**, 601–620.
20 R. C. Lum and J. J. Grabowski, *J. Am. Chem. Soc*., 1992, **114**(22), 8619–8627.
21 The H^+ dissociation enthalpy of $PhOPO_3H_2$ was estimated in 320 kcal mol^{-1}. No experimental measurements were found. This estimate is based on the enthalpy values for H^+ dissociation of H_3PO_4 and $(CH_3O)_2PO_2H$, 330.5 and 331.6 kcal mol^{-1}, respectively.[20,52] The 8.5 kcal mol^{-1} difference between the H^+ dissociation enthalpy of benzoic acid (PhCOOH) and acetic acid (CH_3COOH)[17] was used to correct the H_3PO_4 and $(CH_3O)_2PO_2H$ values. The error for the $PhOPO_3H_2$ enthalpy was estimated to 7 kcal mol^{-1}.
22 R. Schwartz, G. Davico and W. Lineberger, *J. Electron Spectrosc. Relat. Phenom*., 2000, **108**, 163–168.
23 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery Jr, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle and J. A. Pople, *GAUSSIAN 98 (Rev. A.9)*, Gaussian Inc., 1998.
24 D. Rinaldi, P. Hoggan, A. Cartier, K. Baka, G. Monard, M. Loos, A. Mokrane, V. Dillet and V. Thery, GEOMOP Nancy, France, 2001.
25 H. B. Schlegel, *Adv. Chem. Phys*., 1987, **67**, 249–286.
26 M. J. Dewar and W. Thiel, *J. Am. Chem. Soc*., 1977, **99**, 4899–4907.
27 M. J. S. Dewar, M. L. McKee and H. S. Rzepa, *J. Am. Chem. Soc*., 1978, **100**, 3607.
28 M. J. S. Dewar and C. H. Reynolds, *J. Comput. Chem*., 1986, **7**, 140–143.
29 M. J. Dewar, E. G. Zoebisch, H. F. Healy and J. P. P. Stewart, *J. Am. Chem. Soc*., 1985, **107**, 3902–3909.
30 J. J. P. Stewart, *J. Comput. Chem*., 1989, **10**, 209–220.
31 G. M. Arantes, PhD thesis, Institute of Chemistry, University of São Paulo, Brazil, 2004.
32 R. Judson, in *Reviews in Computational Chemistry*, ed. K. Lipkovitz and D. Boyd, VHC, New York, 1997, vol. 10, pp. 1–74.
33 W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. Flannery, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 2nd edn, 1992.
34 G. Menegon, K. Shimizu, J. Farah, L. Dias and H. Chaimovich, *Phys. Chem. Chem. Phys*., 2002, **4**, 5933–5936.
35 D. L. Carroll, *Genetic algorithm driver, version 1.70*, University of Illinois, 1998.
36 C. Alhambra, L. Wu, Z.-Y. Zhang and J. Gao, *J. Am. Chem. Soc*., 1998, **120**, 3858–3866.
37 F. B. Van, Duijneveldt, J. G. C. M. Van, de Rijdt and J. H. Van Lenthe, *Chem. Rev*., 1994, **94**, 1873–1885.
38 M. J. Field, M. Albe, C. Bret, F. P.-D. Martin and A. Thomas, *J. Comput. Chem*., 2000, **21**, 1088–1100.
39 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc*., 1996, **118**, 11225–11236.
40 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Phys. Chem*., 1983, **79**, 926–935.
41 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc*., 1995, **117**, 5179–5197.
42 S. Chalmet and M. F. Ruiz-López, *Chem. Phys. Lett*., 2000, **329**, 154–159.
43 V. Dillet, R. L. Van, Etten and D. Bashford, *J. Phys. Chem. B*, 2000, **104**, 11 321–11 333.
44 F. P.-D. Martin, R. Dumas and M. J. Field, *J. Am. Chem. Soc*., 2000, **122**, 7688–7697.
45 A. Thomas, D. Jourand, C. Bret, P. Amara and M. J. Field, *J. Am. Chem. Soc*., 1999, **121**, 9693–9702.
46 M. J. Field, *A Practical Introduction to the Simulation of Molecular Systems*, Cambridge University Press, Cambridge, 1st edn, 1999.
47 M. Allen and D. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, New York, 1st edn, 1987.
48 B. Roux, *Comput. Phys. Commun*., 1995, **91**, 275–282.
49 D. G. Truhlar, B. C. Garrett and S. J. Klippenstein, *J. Phys. Chem*., 1996, **100**, 12771–12800.
50 Z.-Y. Zhang, L. Wu and L. Chen, *Biochemistry*, 1995, **34**, 16088–16096.
51 J. E. Coleman, *Annu. Rev. Biophys. Biomol. Struct*., 1992, **21**, 441–483.
52 R. A. Morris, W. B. Knighton, A. A. Viggiano, B. C. Hoffman and H. F. Schaeffer III, *J. Chem. Phys*., 1997, **106**, 3545–3547.