Accuracy of density functionals in the description of dispersion interactions and IR spectra of phosphates and phosphorylated compounds

ARTICLE *in* JOURNAL OF MOLECULAR MODELING · SEPTEMBER 2014

Impact Factor: 1.74 · DOI: 10.1007/s00894-014-2426-y · Source: PubMed

3 AUTHORS, INCLUDING:

Gilles Ohanessian
French National Centre for Scientific Resea…
115 PUBLICATIONS **3,713** CITATIONS
SEE PROFILE

Carine Clavaguéra
MINES ParisTech
57 PUBLICATIONS **728** CITATIONS
SEE PROFILE

J Mol Model (2014) 20:2426
DOI 10.1007/s00894-014-2426-y

ORIGINAL PAPER

Accuracy of density functionals in the description of dispersion interactions and IR spectra of phosphates and phosphorylated compounds

Ashwani Sharma · Gilles Ohanessian · Carine Clavaguéra

Received: 12 March 2014 / Accepted: 7 August 2014

Abstract The performances of quantum chemistry methods (i.e., DFT and ab initio) in calculating the structural and vibrational properties of phosphates and phosphorylated compounds have been evaluated. Diethyl-phosphate, phosphonic acid, dihydrogen phosphate anion, phosphoric acid dimer and protonated glycylphosphotyrosine dipeptide were selected for our study. Geometry and harmonic frequency deviations were investigated, pointing out the contribution of dispersion interactions on diethyl-phosphate, $[Gly\text{-}pTyr+H]^+$ and the phosphoric acid dimer. The B3LYP-D functional, followed by CC2 and MP2 methods, revealed significant accuracy for frequency calculations of the majority of the phosphorylated compounds in comparison with available experimental data. These investigations provide a guide to the accurate computation of phosphorylated biological compounds.

Keywords Dispersion interactions · Density functional theory · Infrared spectroscopy · Phosphate · Phosphorylated compounds

Introduction

Phosphorylation of proteins regulates various cellular mechanisms, such as signal transduction, gene regulation, cell survival and transport of metabolites through membranes [1–4]. When phosphorylated, proteins may interact, e.g., with cellular enzymes via intermolecular forces and regulate their enzymatic activities [5, 6]. Therefore, there is a need to determine the structural and spectroscopic properties of phosphorylated proteins in order to understand the mechanisms of their interaction with other biomolecules. In this perspective, gas-phase infrared spectroscopy has emerged as an efficient technique to obtain vibrational signatures of phosphates and phosphorylated compounds [7–16]; however, extensive modeling to relate spectra to structures is required. Commonly, density functional theory (DFT) and ab initio methods are employed for the modeling of chemical compounds or biomolecules [17–19]. Some improvements in density functionals have been made, such as kinetic energy density dependent exchange-correlation functionals (meta-GGA) [20–22] and long-range inclusion of HF exchange [23–25], to prevent the incorrect asymptotic behavior of the exchange term [26]. Furthermore, various corrections have been implemented for overcoming the lack of dispersion interactions in DFT [27–31]. The performances of these density functionals have been evaluated for the description of non-covalent interactions [19, 32–38]. One of the most effective empirical dispersion corrections is known as DFT-D and was developed originally by Grimme in the D2 version [39, 40], or more recently in the D3 version [41–43]. Only few comparable reports are available in the specific cases of phosphorylated compounds [44, 45].

Diethyl-phosphate, phosphonic acid, dihydrogen phosphate anion, phosphoric acid dimer and protonated, C-amidated glycylphosphotyrosine peptide are studied herein using diverse quantum chemistry methods, i.e., DFT and ab initio methods. These compounds correspond to inorganic, organic and biological phosphates as neutral, cationic or anionic molecules. We made the choice to mix simple and more complex compounds based first on the availability of experimental data in gas phase or argon matrix, and second choosing compounds small enough to be amenable to high-level calculations. Thanks to numerical analyses, we investigate the

Electronic supplementary material The online version of this article (doi:10.1007/s00894-014-2426-y) contains supplementary material, which is available to authorized users.

A. Sharma · G. Ohanessian · C. Clavaguéra (✉)
Laboratoire de Chimie Moléculaire, Department of Chemistry, Ecole Polytechnique, CNRS, 91128 Palaiseau Cedex, France
e-mail: carine.clavaguera@polytechnique.edu

Published online: 22 August 2014

capabilities of various functionals, i.e., GGA and hybrid functionals, which are corrected or not by a dispersion term, and ab initio methods, i.e., MP2 and CC2, to evaluate dispersion effects on both geometries and vibrational frequencies. Furthermore, the computed frequencies are compared to experimental frequencies using statistical criteria in order to select the best functionals. To the best of our knowledge, our work provides the first calibration test of DFT/DFT-D and ab initio methods for these phosphorylated compounds.

Computational details

Quantum chemistry

Geometry optimizations and vibrational frequency calculations were performed using density functional theory (DFT) with the split-valence plus polarization (def2-SVP) and triple-ζ valence plus double polarization (def2-TZVPP) basis sets along with resolution of the identity (RI) approximation. We considered two GGA functionals, i.e., PBE and BLYP, one meta-GGA functional, i.e., TPSS, and one hybrid functional, i.e., B3LYP, as a variety of commonly available functionals. Furthermore, we used the empirical dispersion corrections proposed by Grimme, in the named D2 version, [39, 40] for these functionals (i.e., PBE-D, BLYP-D, TPSS-D and B3LYP-D) in order to evaluate dispersion contributions to the geometries and vibrational frequencies. Additionally, the RI-MP2 and RI-CC2 methods with def2-SVP and def2-TZVPP basis sets were considered for geometry optimizations and frequency calculations. These levels of theory were used as a reference in the case of lack of experimental data. These calculations were performed using the TURBOMOLE v6.0 program package [46].

Geometry optimization and vibrational frequency calculations were also performed using the M06 and M06-L functionals associated with two types of basis sets (DZV and TZV) augmented by two different sets of polarization functions ("common" and "Hondo7") using the GAMESS 2010 package [47, 48]. As both sets of polarization functions gave similar results in terms of geometries and frequencies, only the results obtained with the "common" set are presented below.

Analysis tools

Faced with the amount of structural and spectral data generated by the calculations, only a statistical analysis is able to provide an objective comparison of the methods. Dispersion contributions provided by the dispersion-corrected functionals on geometries and frequencies of phosphorylated compounds were analyzed by three numerical tests. First, the root-mean-square deviation (RMSD) between the optimized geometries obtained at DFT and DFT-D levels was calculated to evaluate the dispersion contribution to geometries. If RMSD is greater than 0.08 Å, the effect is considered significant. This cutoff was chosen based on the average of the highest and lowest values of RMSD obtained for the different compounds. Second, the bond lengths between neighboring atoms within the DFT and DFT-D optimized geometries were measured. A bond length difference greater than 0.05 Å was considered to reflect a dispersion contribution. Third, these effects can also be pointed out by the difference in IR vibrational frequency values obtained at DFT and DFT-D levels. We considered that a difference in the frequency value greater than 15 cm^{-1} was evidence of dispersion interactions. The cutoff values for bond lengths and frequencies were selected based on the difference between DFT and DFT-D values. Geometries and frequencies satisfying all three tests were considered to be impacted by dispersion corrections.

Assessment of density functionals for describing IR spectra of phosphorylated compounds was performed using statistical analyses such as analysis of variance (ANOVA), *t*-test, and calculation of standard deviations (SD) and correlation coefficients (CC). The SD and CC methods were combined with the least square fitting method in order to compare the divergence of the calculated frequencies from the experimental values. ANOVA and *t*-test are dependent on the population size of the source data sets and are very sensitive to small changes in these data sets. As only few reported values of experimental frequencies were available for comparison with calculated frequencies, these results contained a large amount of false positive values and were not retained for this analysis. The experimental values of frequencies are denoted by Y_i ($i=1-n$) and calculated frequencies by X_i. The fitting process was carried out to calculate the SD and CC values between the calculated and experimental frequency data. Here, the larger the divergence between the calculated and experimental frequencies, the larger the value of SD. Moreover, the variation between the calculated and experimental IR frequencies was measured by CC (the closer the calculated and experimental frequencies, the higher the value of CC) defined as:

$$\mathrm{cc} = \frac{\sum_{i=1}^{n}\left(\left(x_i-\bar{\bar{x}}\right)\left(y_i-\bar{\bar{y}}\right)\right)^2}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{\bar{x}}\right)^2\sum_{i=1}^{n}\left(y_i-\bar{\bar{y}}\right)}}$$

Scaling factor

In order to correlate the computational vibrational frequencies with experimental data, scaling factors are usually applied to the calculated values with the aim of compensating for the approximations in electronic structure calculations (basis set

and method effects) as well as to account for the anharmonic behavior of the potential energy surface [36, 49, 50]. In the case of phosphate modes, only a few studies for the determination of mode-specific scaling factors are available in the literature [51, 52]. At various levels, different scaling factors were determined that can be either smaller or larger than 1. Consequently, in some cases it was not possible to provide uniform scaling factor for modes of the phosphorylated compounds [11, 13, 52, 53]. As a result, no scaling factor was applied to the computed frequencies provided in this study regardless of the method and basis set used. A discussion about the resulting scaling factors for the compounds under study is provided in "Scaling factor".

Results

The geometries and IR frequencies were calculated via DFT and DFT-D, using B3LYP, BLYP, PBE and TPSS functionals, for the different systems, i.e., diethyl-phosphate, phosphonic acid, dihydrogen phosphate anion, phosphoric acid dimer and protonated C-amidated glycylphosphotyrosine (Fig. 1), with the aim of estimating the role of dispersion both on structures and vibrational spectra. The RMSD on optimized geometries obtained in DFT were compared to CC2 data. The MP2/TZVPP and CC2/TZVPP levels were used as reference because of the lack of experimental values for these structures. The most stable structure was used for each molecule. The results are provided as bar diagrams.

Dispersion corrections on geometries

For diethyl-phosphate, we observed that interactions involving the CH_3 groups and one phosphate oxygen showed significant distance differences when calculated without and with dispersion corrections. The inclusion of a dispersion term leads to shortening of the P–O···H and H···H bond lengths by 0.2 Å – 0.3 Å, i.e., 3.75 Å for the H···O distance and 4.90 Å for the H···H distance vs 3.53 and 4.60 Å at B3LYP/TZVPP and B3LYP-D/TZVPP levels, respectively. In addition, other functionals (BLYP, PBE and TPSS) and smaller basis sets (SVP) produced the same trend of dispersion contribution on H···O and H···H interactions. We also observed that the O–C–C and P–O–C angles are reduced slightly by the addition of dispersion corrections whatever the basis set or functional used. The RMSD between the geometries obtained by DFT or DFT-D in comparison with MP2 and CC2 methods are shown in Fig. 2. For both basis sets, the RMSD was greater than 0.08 Å between DFT and DFT-D (all functionals) and also between DFT and MP2 or CC2, highlighting the effects on the geometry of dispersion corrections. DFT-D is able to produce geometries close to those of CC2 and MP2, showing also an improvement from DFT-D/SVP to DFT-D/TZVPP. In the TPSS case, the basis set effect is more pronounced when comparing DFT-D and MP2 or CC2.

A similar procedure was carried out for $HPO(OH)_2$. Bond length analysis revealed no significant difference (> 0.05 Å) between DFT and DFT-D geometries. Furthermore, RMSD estimates between the DFT or DFT-D and MP2 and CC2 methods revealed that geometries are very similar at all levels, giving RMSD values less than 0.08 Å. (Fig. S1). These results indicate that the addition of dispersion corrections does not have any significant effect on the geometry of phosphonic acid. Similarly, geometry optimizations of the $H_2PO_4^-$ ion show that there is no bond length difference between DFT and DFT-D geometries, as well as between MP2 and CC2 results (Fig. S2). The dispersion interactions on the geometry changes for these two compounds are negligible.

The most stable conformation of the $[Gly\text{-}pTyr+H]^+$ dipeptide (Fig. 1) was selected from a previous study [48]. The

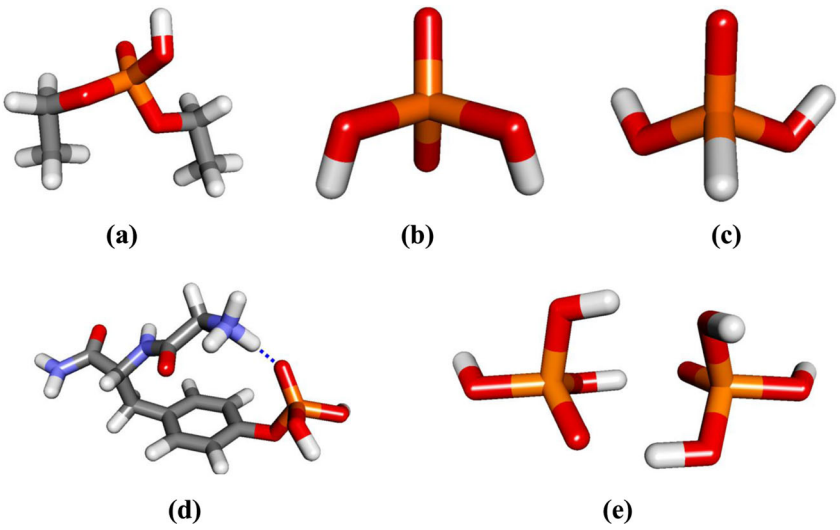

Fig. 1 **a**–**e** Representation of the phosphorylated compounds under study. **a** Diethyl-phosphate, **b** dihydrogen phosphate anion, **c** phosphonic acid, **d** protonated C-amidated glycyl phosphotyrosine peptide, **e** phosphoric acid dimer

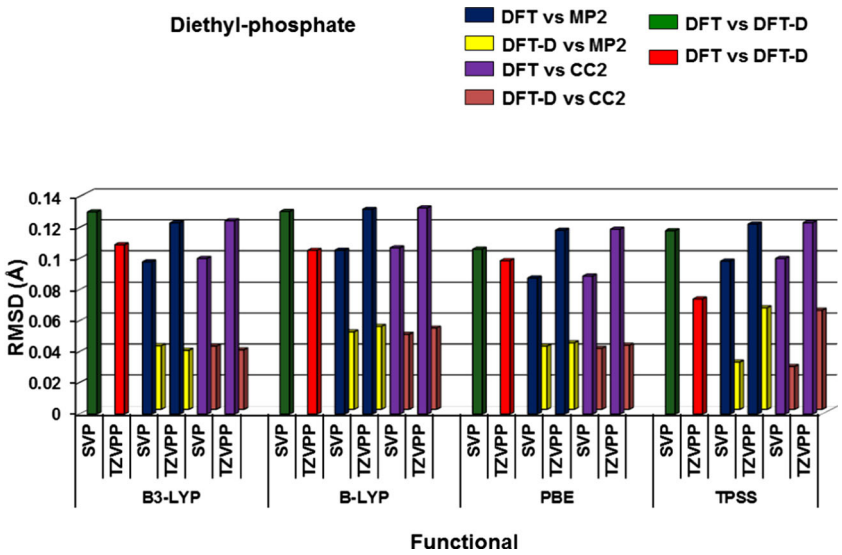

Fig. 2 Root-mean-square deviation (RMSD) calculation between the diethyl-phosphate optimized geometries obtained at DFT and DFT-D, MP2 and CC2 levels using the SVP and TZVPP basis sets

B3LYP/TZVPP and B3LYP-D/TZVPP geometries exhibit several types of interactions for which the difference in distance is larger than 0.05 Å. Indeed, the interactions between the phosphate group and one hydrogen atom of the aromatic cycle are shortened at the B3LYP-D level by more than 0.1 Å, and, similarly, for the interaction between the oxygen of the peptidic bond and one hydrogen atom of the ring. For example, the H···O bond length between the hydrogen atom of the NH_3^+ group of the glycine residue and the oxygen atom of the phosphate group of the tyrosine residue is reduced from 3.43 Å (B3LYP), 3.44 Å (BLYP), 3.33 Å (PBE) and 3.29 Å (TPSS) to 3.26 Å (B3LYP-D), 3.30 Å (BLYP-D), 3.20 Å (PBE-D) and 3.11 Å (TPSS-D) using the TZVPP basis set. This shortening of the HO bond length helped to make a strong hydrogen bond interaction between the NH_3^+ and O= P groups. In addition, the C···C bond length between the carbon atom of the glycine residue and the carbon atom of the aromatic ring was also reduced from 4.79 Å (B3LYP), 4.86 Å (BLYP), 4.76 Å (PBE) and 4.80 Å (TPSS) to 4.62 Å (B3LYP-D), 4.66 Å (BLYP-D), 4.63 Å (PBE-D) and 4.72 Å (TPSS-D) with TZVPP basis set. Similar effects were observed with the SVP basis set. However, the BLYP functional failed to point out any significant dispersion contributions to the peptide geometry using the small basis sets. The small decrease of some bond angles (P–O–C, C–N–C, N–C–C and C–C–C) also supports the effect of dispersion corrections. Furthermore, the RMSD between the optimized geometries at DFT and DFT-D levels is greater than 0.08 Å for all functionals (Fig. 3), and comparison of the MP2 and CC2 geometries with DFT geometries also shows an RMSD greater than 0.08 Å. On the contrary, MP2 and CC2 geometries are very close to the DFT-D geometries (except for TPSS). These results reveal the effect of dispersion interactions on the geometry of $[Gly\text{-}pTyr+H]^+$.

The $(H_3PO_4)_2$ dimer is the only system under study for which two entities interact only by intermolecular forces (Fig. 1). The DFT and DFT-D geometries show that the hydrogen bond lengths between the two phosphate groups are impacted by dispersion with the O–H···O interactions, being shortened by up to 0.2 Å using the TZVPP basis set, i.e., from 2.25 Å (B3LYP), 2.26 Å (BLYP), 2.09 Å (PBE) and 2.11 Å (TPSS) to 2.06 Å (B3LYP-D), 2.04 Å (BLYP-D), 1.99 Å (PBE-D) and 1.96 Å (TPSS-D), respectively. This effect is much diminished when using the SVP basis set, probably due to cancellation of errors, i.e., from 1.99 Å (B3LYP) and 1.97 Å (BLYP) to 1.91 Å (B3LYP-D) and 1.89 Å (BLYP-D). The presence of the dispersion term reduces the values of the H–O–P angles by 1°, from 111.6° (B3LYP) to 110.7° (B3LYP-D) and 111.2° (BLYP) to 110.2° (BLYP-D) with both basis sets. Furthermore, RMSD analysis between DFT and DFT-D geometries revealed that the RMSD value is between 0.04 and 0.08 Å depending on the functional (Fig. 4). DFT-D geometries are close to those of CC2 and MP2 used as a reference, which confirms the importance of the dispersion interactions in this dimer.

Dispersion corrections on vibrational frequencies

Dispersion interactions were estimated by analysis of the vibrational frequency differences between DFT and DFT-D values using both basis sets for the five compounds. Regarding the diethyl-phosphate molecule, several frequency values (Table S1 in supporting information) differ by more than 15 cm^{-1} in the 800–3,200 cm^{-1} frequency range, whatever the functional. These results reveal the contribution of dispersion to diethyl-phosphate IR frequencies, especially on C–H stretching modes, as a similar contribution to the geometry was found. Analogously for the $[Gly\text{-}pTyr+H]^+$ dipeptide,

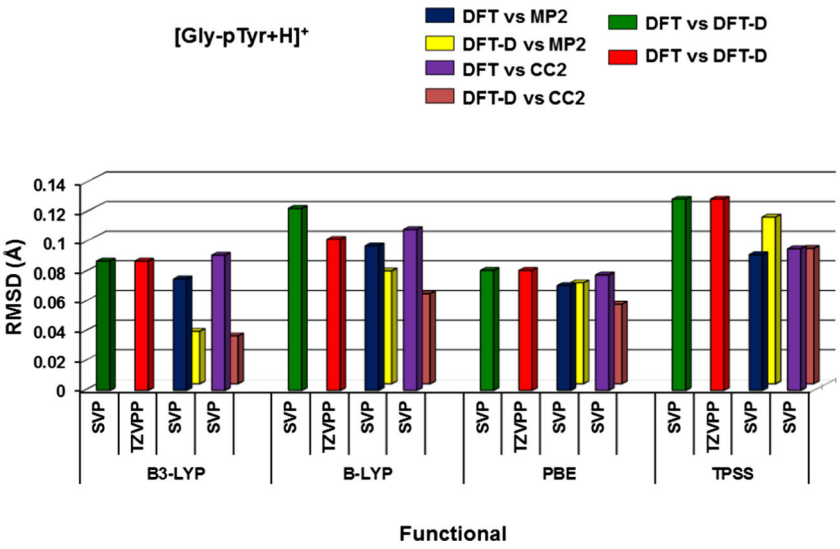

Fig. 3 RMSD calculation between the $[Gly\text{-}pTyr+H]^+$ optimized geometries obtained at DFT and DFT-D, MP2 and CC2 levels using the SVP and TZVPP basis sets

DFT and DFT-D results also lead to several frequency differences larger than 15 cm^{-1} with the four functionals (Table S2). The dispersion contribution was more pronounced with the CCH bend and C–H stretch modes, and both on peptidic as well as ammonium N–H stretch modes. The largest effect was observed for the N–H stretch mode of the ammonium N–H hydrogen-bonded with P=O because the strength of this interaction is very sensitive to the level of the calculation. For the $(H_3PO_4)_2$ dimer, the dispersion correction on frequencies is significant, with frequency differences that can reach 60 cm^{-1} in some cases (Table S3). The POH bend modes and O–H stretch modes are the most impacted by dispersion corrections. Finally, no frequency difference (greater than 15 cm^{-1}) was detected between DFT and DFT-D calculations for phosphonic acid and the $H_2PO_4^-$ anion. Analogous conclusions as for geometries can be established for these two systems, for which frequency analyses show that the dispersion contribution is negligible. These conclusions were to be expected for this size of phosphate molecules.

The combination of the three tests described above on geometries and frequencies provides a tool to evaluate numerically how to take dispersion into account in phosphorylated compounds. The results highlight the ability of the DFT-D framework to reproduce these effects as ab initio methods.

Performance of DFT and ab initio methods to reproduce vibrational frequencies

The experimental frequencies were collected from gas phase and argon matrix experiments [54–57] and are provided with mode assignments in Table 1. The experimental data are in the 900–1,700 cm^{-1} frequency range, except for the P–H stretch mode around 2,500 cm^{-1}. For diethyl-phosphate, phosphonic acid and dihydrogen phosphate, the available experimental

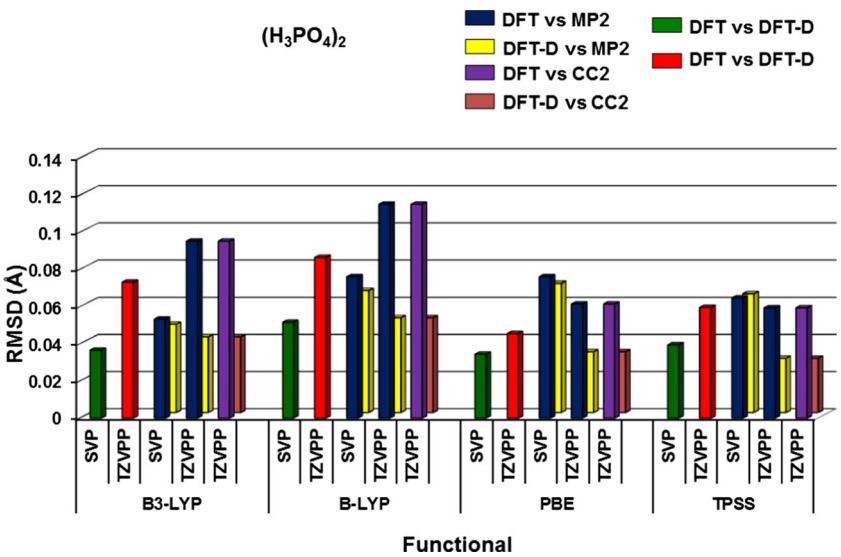

Fig. 4 RMSD calculation between the $(H_3PO_4)_2$ optimized geometries obtained at DFT and DFT-D, MP2 and CC2 levels using the SVP and TZVPP basis sets

Table 1 Comparison of available experimental and computed vibrational frequencies (in cm^{-1}) at B3LYP/TZVPP, B3LYP-D/TZVPP and M06-L/TZV levels with corresponding modes assignments

Exp.	B3LYP	B3LYP-D	M06-L	Vibrational mode
Diethyl-phosphate				
906	904	902	909	CCH bend, C-C str, P-OH str, P-OH bend
970	962	977	966	CCH bend, C-C str, P-OH str, P-OH bend
1,050	1,048	1,045	1,064	P-OH bend
	1,054	1,049	1,073	PO-C str, C-C str, P-OH str, P-OH bend
1,154	1,126	1,119	1,136	CH_3 bend
	1,184	1,183	1,182	CH_3 bend, CH_2 bend
1,246	1,297	1,296		CH_2 bend, P=O str
1,306	1,314	1,314	1,307	CH_2 bend
Phosphonic acid				
873	853	853	861	P-OH str, P-OH bend
902	905	907	908	P-OH str, P-OH bend
1,299	1,300	1,299	1,326	P-OH bend, OP-H bend, P=O str
2,487	2,524	2,523	2,504	P-H str
Dihydrogen phosphate anion				
764	758	755	752	P-OH str
1,106	1,108	1,112	1,112	P-OH bend, P=O str
1,297	1,330	1,329	1,344	P-OH bend, P=O str
$[Gly\text{-}pTyr+H]^+$				
937	926	927	939	P-OH str, P-OH bend
967	969	961	966	Aromatic CCH bend, P-OH bend
1,005	1,006	1,006	1,001	CH_2 bend, aromatic CCH bend
1,219	1,222	1,217	1,218	CH_2 bend, aromatic CCH bend, PO-C str
1,248	1,240	1,244	1,247	PO-C str, P=O str, C-C str, C=C str, CNH bend
1,289	1,281	1,289	1,296	C-N str, CH_2 bend, NH_3 bend
1,433	1,422	1,439	1,447	CH_2 bend, NH_3 bend
1,470	1,464	1,472	1,470	CH_2 bend, NH_3 bend
1,503			1,491	CH_2 bend
1,540	1,544	1,540	1,543	PO-C str, C=C str, C-C str, aromatic CCH bend
1,590	1,600	1,587	1,600	NH_3 bend
1,668	1,667	1,665	1,665	PO-C str, C=C str, C-C str, aromatic CCH bend

data are related mainly to phosphate modes such as P=O and P–O stretch and POH bend modes. Depending on the molecule, P–H stretch, CH_n and OPH bend modes are additionally accessible. The frequencies of $[Gly\text{-}pTyr+H]^+$ and $H_2PO_4^-$ were obtained from IRMPD experiments [54, 55, 57] providing infrared signatures of specific chemical groups: phosphate modes (POH bend, P–O and P=O stretch modes) but also C–C stretch, XCH and XNH bend modes. This selection of experimental data was compared with the calculated frequencies at various levels. The methods were ranked with SD and CC analysis to measure the variability of the computed frequencies with the experimental values.

The trends of SD as a function of the method used for frequency calculations (ordered at the right side of the graph) with respect to experiment are shown in Fig. 5 for diethyl-phosphate and phosphonic acid. For diethyl-phosphate, the comparison is represented in blue (closed circles with solid line) for DFT functionals and ab initio methods, and in pink (closed circles with dashed line) for dispersion-corrected functionals. The lowest values of SD were produced by CC2/SVP and CC2/TZVPP calculations. B3LYP-D, BLYP, BLYP-D, TPSS and TPSS-D functionals associated with the TZVPP basis set also perform well in comparison with other functionals. In addition, the values of CC, which represent higher closeness between the calculated and experimental data, are consistent with SD results (Fig. S3) for CC2, B3LYP-D, BLYP and BLYP-D. It is found that M06-L and M06 results are less accurate than those obtained with the other methods to calculate the IR frequencies, whatever the basis set. MP2 results are intermediate between these two sets of methods mentioned above.

On the contrary, for phosphonic acid (red and green triangles in Fig. 5), MP2/SVP provides frequencies closer to those of experiments than CC2 (SVP or TZVPP) and MP2/TZVPP,

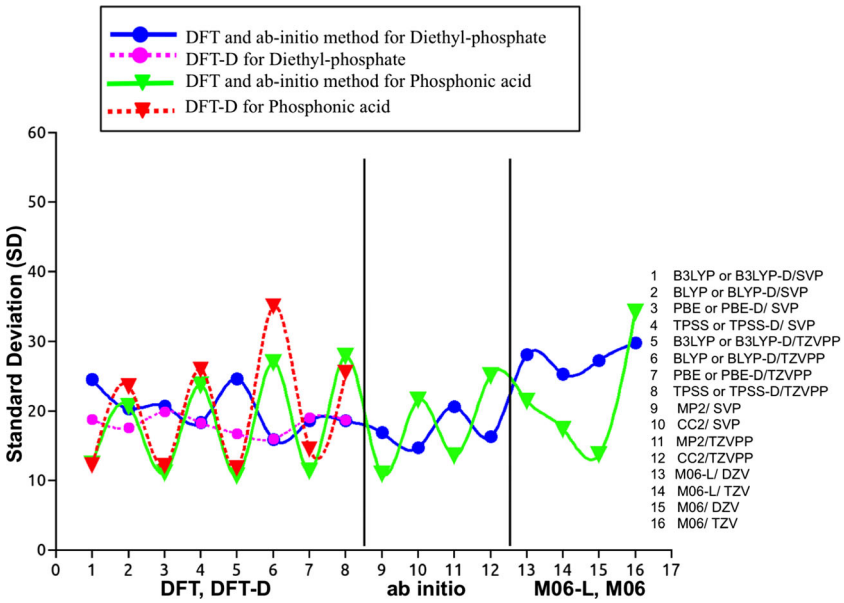

Fig. 5 Trends of standard deviation (SD) for vibrational frequencies of diethyl-phosphate and phosphonic acid as a function of the quantum mechanics level used for the frequency calculation, with respect to experimental frequencies. *Solid lines* DFT and ab initio, *dashed lines* DFT-D

probably due to cancellation of errors. The SD analysis shows that B3LYP, B3LYP-D, PBE and PBE-D functionals (with both SVP and TZVPP basis sets) are able to give an accurate estimate of the frequencies. However, M06 and M06-L functionals give comparable SD values but with a larger SD for M06/TZV. BLYP, BLYP-D, TPSS and TPSS-D functionals give larger errors. Analysis of CC produced uniform trends for all values of frequencies (Fig. S3) and this is may be due to very small differences in the CC values.

Figure 6 reports the SD results for $H_2PO_4^-$ (blue and pink circles). The values of SD for the B3LYP (blue filled circles) and B3LYP-D calculations (pink filled circles), both with SVP and TZVPP basis sets (number 1 and 5 on *x*-axis), are lower than for other methods. Accordingly, the CC value is also the highest in these cases (Fig. S4). Furthermore, the MP2 and CC2 methods produce higher values of SD as compared to B3LYP and B3LYP-D functionals. M06-L/TZV performs also well in comparison with M06 (TZV and DZV) or M06-L/DZV, but still with a SD value larger than for B3LYP and B3LYP-D.

The comparative graph for $[Gly\text{-}pTyr+H]^+$ is also given in Fig. 6 (green and red triangles). The values of SD at B3LYP/SVP and B3LYP-D/SVP levels (number 1 on *x*-axis) are low for this system in comparison with other methods. The lowest

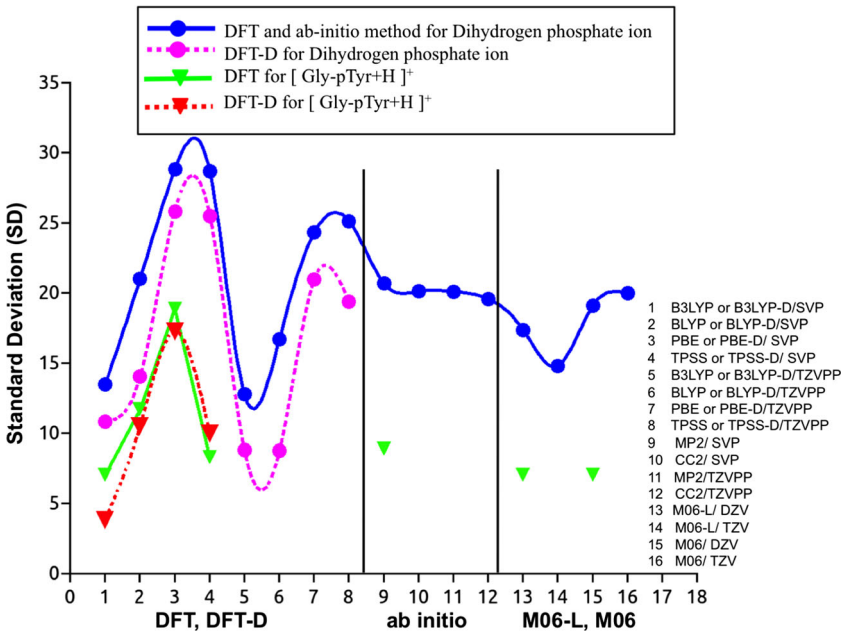

Fig. 6 Trends of SD for vibrational frequencies of dihydrogen phosphate anion and $[Gly\text{-}pTyr+H]^+$ dipeptide as a function of the quantum mechanics level used for the frequency calculations, with respect to experimental frequencies. *Solid lines* DFT and ab initio, *dashed lines* DFT-D

value is obtained using the B3LYP-D functional, which indicates that dispersion corrections have a significant effect on the vibrational frequencies of $[Gly\text{-}pTyr+H]^+$. MP2/SVP, M06-L/DZV and M06/DZV calculations also give accurate frequencies in comparison with experiments, however SD values are larger than for B3LYP-D.

As the experimental frequencies used in this study correspond for the most part to phosphate modes in the 900–1,700 cm^{-1} range, we decided to avoid the use of scaling factors to correct the harmonic calculated frequencies. Indeed, the required scaling factor, calculated as the ratio between the computed and experimental frequencies in Table 1, is very close to 1.0 in most cases. This value differs clearly from the standard scaling factors used commonly for other modes [36, 49, 50]. Consequently, this study justifies the use of mode-specific factor values for phosphate vs non-phosphate modes for phosphorylated compounds. The SD analysis tool highlights the general performance of the B3LYP functional for small phosphate compounds. Furthermore, as discussed above, dispersion contributions were highlighted on geometries when non-covalent interactions are present. In such cases, B3LYP-D is the best functional to estimate vibrational frequencies. When the calculation is computationally practicable, CC2, and to a lesser extent MP2, can produce very accurate results. Furthermore, the M06 series was found to perform well in some cases but with large errors in others, the results remaining very dependent on the compound.

Conclusions

The aim of the present study was to explore the performances of various quantum chemistry methods, including DFT, dispersion-corrected DFT, MP2 and CC2, regarding geometries and IR vibrational frequencies of phosphorylated compounds. Accordingly, several tools were tested to give numerical estimates of these performances. RMSD calculations, bond length and frequency difference analyses provide evaluations of these quantum chemistry methods. Comparison between DFT and DFT-D revealed the importance of taking into account dispersion interactions in diethyl-phosphate, protonated C-amidated glycylphosphotyrosine dipeptide and phosphoric acid dimer, in order to reproduce geometries and IR frequencies. Although PBE performs well for geometries and vibrational frequencies for small phosphates, it is less efficient for IR frequencies when non-covalent interactions are involved. TPPS displays rather irregular performance. Thus, these two functionals may not be recommended for computations of phosphates with non-covalent interactions. SD calculations between computed and experimental vibrational frequencies emphasize a good performance of B3LYP-D especially with large basis sets, i.e., TZVPP, in comparison with other density functionals for most of our compounds. This confirms that the combination of dispersion corrections and Hartree-Fock exchange is most often essential in the calculation of vibrational frequencies for phosphorylated compounds. These results also highlight the accuracy of CC2 with SVP or TZVPP basis sets to reproduce experimental vibrational frequencies, even if calculations with this method are achievable only for small systems. Our future work will focus on phosphorylated peptides of larger size in order to detail the role of phosphorylation in peptide structure.

Acknowledgments A.S. thanks the vice-presidency for external relations in Ecole Polytechnique for support from the international internship program, and for a 1-year post-doctoral fellowship.

References

1. Blom N, Kreegipuu A, Brunak S (1998) Nucleic Acids Res 26:382
2. Diella F, Cameron S, Gemünd C, Linding R, Via A, Kuster B, Sicheritz-Pontén T, Blom N, Gibson T (2004) J BMC Bioinforma 5:79
3. Johnson LN, Lewis RJ (2001) Chem Rev 101:2209
4. Kreegipuu A, Blom N, Brunak S (1999) Nucleic Acids Res 27:237
5. Engel D, Nudelman A, Levovich I, Gruss-Fischer T, Entin-Meer M, Phillips DR, Cutts SM, Rephaeli A (2006) J Cancer Res Clin Oncol 132:673
6. Park J, Singh B, Maj MC, Gupta RS (2004) Protein J 23:167
7. Cuisset A, Mouret G, Pirali O, Roy P, Cazier F, Nouali H, Demaison J (2008) J Phys Chem B 112:12516
8. Smirnova IN, Cuisset A, Hindle F, Mouret G, Bocquet R, Pirali O, Roy P (2010) J Phys Chem B 114:16936
9. Salpin JY, Guillaumont S, Ortiz D, Tortajada J, Maître P (2011) Inorg Chem 50:7769
10. Chiavarino B, Crestoni ME, Fornarini S, Lanucara F, Lemaire J, Maitre P, Scuderi D (2008) Int J Mass Spectrom 270:111
11. Fales BS, Fujamade NO, Oomens J, Rodgers MT (2011) J Am Soc Mass Spectrom 22:1862
12. Nei YW, Hallowita N, Steill JD, Oomens J, Rodgers MT (2013) J Phys Chem A 117:1319
13. Rummel JL, Steill JD, Oomens J, Contreras CS, Pearson WL, Szczepanski J, Powell DH, Eyler JR (2011) Anal Chem 83:4045
14. Fales BS, Fujamade NO, Nei YW, Oomens J, Rodgers MT (2011) J Am Soc Mass Spectrom 22:81
15. Nei YW, Crampton KT, Berden G, Oomens J, Rodgers MT (2013) J Phys Chem A 117:10634
16. Lanucara F, Crestoni ME, Chiavarino B, Fornarini S, Hernandez O, Scuderi D, Maitre P (2013) RSC Adv 3:12711
17. Dahlke EE, Orthmeyer MA, Truhlar DG (2008) J Phys Chem B 112:2372
18. Riley KE, Hobza P (2007) J Phys Chem A 111:8257
19. Van Mourik T (2008) J Chem Theory Comput 4:1610
20. Perdew JP, Kurth S, Zupan A, Blaha P (1991) Phys Rev Lett 82:2544
21. Adamo C, Ernzerhof M, Scuseria GE (2000) J Chem Phys 112:2643
22. Peverati R, Truhlar DG (2011) J Phys Chem Lett 2:2810
23. Savin A (1996) In: Seminario JM (ed) Recent developments and applications of modern density functional theory. Elsevier, Amsterdam, pp 327–357
24. Song JW, Watson MA, Hirao K (2009) J Chem Phys 131:144108
25. Iikura H, Tsuneda T, Yanai T, Hirao K (2001) J Chem Phys 115:3540

26. Jacquemin D, Perpète EA, Scalmani G, Frisch MJ, Kobayashi R, Adamo C (2007) J Chem Phys 126:144105
27. Becke AD, Johnson ER (2005) J Chem Phys 122:154104
28. Tkatchenko A, Scheffler M (2009) Phys Rev Lett 102:073005
29. Gritsenko O, Baerends EJ (2006) J Chem Phys 124:1
30. Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Phys Rev Lett 92:246401
31. Waller MP, Robertazzi A, Platts JA, Hibbs DE, Williams PA (2006) J Comput Chem 27:491
32. Černý J, Hobza P (2005) Phys Chem Chem Phys 7:1624
33. Cohen AJ, Mori-Sánchez P, Yang W (2008) Science 321:792
34. Hohenstein EG, Chill ST, Sherrill CD (2008) J Chem Theory Comput 4:1996
35. Jacquemin D, Perpète EA, Ciofini I, Adamo C, Valero R, Zhao Y, Truhlar DG (2010) J Chem Theory Comput 6:2071
36. Bouteiller Y, Poully JC, Desfrancois C, Gregoire G (2009) J Phys Chem A 113:6301
37. Foster ME, Sohlberg K (2010) Phys Chem Chem Phys 12:307
38. Morgado CA, McNamara JP, Hillier IH, Burton NA, Vincent MA (2007) J Chem Theory Comput 3:1656
39. Grimme S (2004) J Comput Chem 25:1463
40. Grimme S (2006) J Comput Chem 27:1787
41. Grimme S, Antony J, Ehrlich S, Krieg HA (2010) J Chem Phys 132: 154104
42. Hujo W, Grimme S (2011) Phys Chem Chem Phys 13:13942
43. Ehrlich S, Moellmann J, Grimme S (2013) Acc Chem Res 46:916
44. Potrzebowski MJ, Assfeld X, Ganicz K, Olejniczak S, Cartier A, Gardiennet C, Tekely P (2003) J Am Chem Soc 125:4223
45. Mládek A, Šponer JE, Jurečka P, Banáš P, Otyepka M, Svozil D, Šponer J (2010) J Chem Theory Comput 6:3817
46. Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Chem Phys Lett 162:165
47. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, John A, Montgomery JJ (1993) J Comput Chem 14:1347
48. Gordon MS, Schmidt MW (2005) In: Dykstra CE, Frenking G, Kim KS, Scuseria GE (eds) Theory and applications of computational chemistry: the first forty years; Elsevier, Amsterdam, pp 1167
49. Scott AP, Radom L (1996) J Phys Chem 100:16502
50. Wong MW (1996) Chem Phys Lett 256:391
51. Carauta ANM, De Souza V, Hollauer E, Téllez SCA (2004) Spectrochim Acta Part A 60:41
52. Katsyuba S, Vandyukova E (2003) Chem Phys Lett 377:658
53. Correia CF, Balaj PO, Scuderi D, Maitre P, Ohanessian G (2008) J Am Chem Soc 130:3359
54. Correia CF, Clavaguera C, Erlekam U, Scuderi D, Ohanessian G (2008) Chem Phys Chem 9:2564
55. Scuderi D, Correia CF, Balaj OP, Ohanessian G, Lemaire J, Maitre P (2009) Chem Phys Chem 10:1630
56. Withnall R, Andrews L (1987) J Phys Chem 91:784
57. Jiang L, Sun ST, Heine N, Liu JW, Yacovitch TI, Wende T, Liu ZF, Neumark DM, Asmis KR (2014) Phys Chem Chem Phys 16:1314