JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Accuracy of Density Functionals in the Prediction of Electronic Proton Affinities of Amino Acid Side Chains

Natércia F. Brás,[†,§] Marta A. S. Perez,[†,§] Pedro A. Fernandes,[†] Pedro J. Silva,[‡] and Maria J. Ramos*[,†]

[†]REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal

[‡]REQUIMTE, Fac. de Ciências da Saúde, Universidade Fernando Pessoa, Rua Carlos da Maia, 296, 4200-150 Porto, Portugal

Supporting Information

ABSTRACT: The ionization states of amino acids influence the structure, function, stability, solubility, and reactivity of proteins and are difficult to determine unambiguously by experimental means. Thus, it is very important to be able to determine them theoretically and with high reliability. We have analyzed how well DFT functionals, often used to characterize complex and large models such as proteins, describe the zero-point-exclusive proton affinity at 0 K, PA_{el}^{0K}, for the ionizable side chains of lysine (Lys), histidine (His), arginine (Arg), and aspartate (Asp^-) as well as the cysteine (Cys^-), serine (Ser^-), and tyrosine (Tyr^-) anions. The reference values PA_{el}^{0K} were determined at the very accurate CCSD(T)/CBS level. Those values were obtained by the sum of the complete basis set limit of the MP2 energies plus a CCSD(T) correction term evaluated with the aug-cc-pVTZ basis set. The complete basis set limit of MP2 energies was determined using the Truhlar and Helgaker extrapolation schemes. A new, important, and consistent DFT benchmarking database for PA_{el}^{0K} and for proton transfer between two different ionizable side chains, ΔPA_{el}^{0K}, is provided, making this work relevant to all studies with ionizable amino acids side chains that use DFT. Among the 64 density functionals tested, the MPW1B95-D3, XYG3, MPW1B95, B1B95-D3, BMK, BMK-D3, M06-2X, B1LYP, B1B95, PBE1PBE, CAM-B3LYP, B97-1, PBE1KCIS, B3P86, CAM-B3LYP-D3, B3LYP, B98, M06-L, and M06 provide the most accurate PA_{el}^{0K} values for all ionizable amino acids studied, with errors below 1.5 kcal/mol, which translate into an error of less than 1 pK_a unit in solution. Furthermore, among the best rated to predict PA_{el}^{0K}, we have found that M06-2X was the most accurate density functional for proton transfers between different amino acids.

INTRODUCTION

Amino acids are critical to life. They are the building units of proteins and have countless fundamental functions in the metabolism of all living beings. The amino acid residues with ionizable side chains make up, on average, 29% of the residues in proteins.[1] Their ionization state influences the structure, function, stability, solubility, and reactivity of proteins.[2,3] Solvent exposure, Bohr effect, charge−dipole interactions, and charge−charge interactions are important environmental effects that change the ionization state of the amino acid residues in proteins[4] and, consequently, the charge of the proteins. The enzymatic reactivity strongly depends on the ionization state of the active site residues and on the proton transfers taking place between them. The recognition of ligands and drugs by biological receptors is strongly influenced by the charge of the receptor. The knowledge of the ionization state of amino acid residues in proteins is thus of great importance to the chemical and pharmaceutical industries.

Experimental methods that assign the ionization state to the amino acid residues are typically indirect and difficult to perform.[5,6] While the proton affinities of a large variety of organic molecules have been measured, the data on amino acid residues in proteins is less complete, in part because they tend to be relatively nonvolatile and thermally labile.[5] Moreover, the methods of measuring proton affinities (equilibrium measurements, reaction bracketing, and kinetic methods) provide only relative values. As there is still significant uncertainty in the reference scale of proton affinities, it becomes difficult to rank them on an absolute scale.[5] The proton affinity of a given amino acid can vary significantly in different protein environments, which makes difficult the unambiguous determination of its ionization state. The fact that the protons are not detectable by X-ray crystallography further complicates this issue.[5,6] As experimental methods face severe impracticalities in many situations it becomes very important to calculate the protonation state of the ionizable amino acid side chains.

There are numerous studies that report different kinds of theoretical approaches and procedures, which predict the protonation state of ionizable residues.[7−12] Most of them predict the pK_a and/or the proton affinities (PA) and/or the gas-phase basicities (GPB) in a specific environment and compare the predictions with experimental values. Many of these methods are knowledge-based.[11,12] Even though this philosophy is quite useful, knowledge-based methods always have the limitation of depending on the training set that is used for the calibration and typically fail in the most complex and unusual cases, which constitute the most interesting ones from a biochemistry point of view. Moreover, given the extraordinary diversity of microenvironments that proteins can build up, knowledge-based

Received: May 4, 2011
Published: October 25, 2011

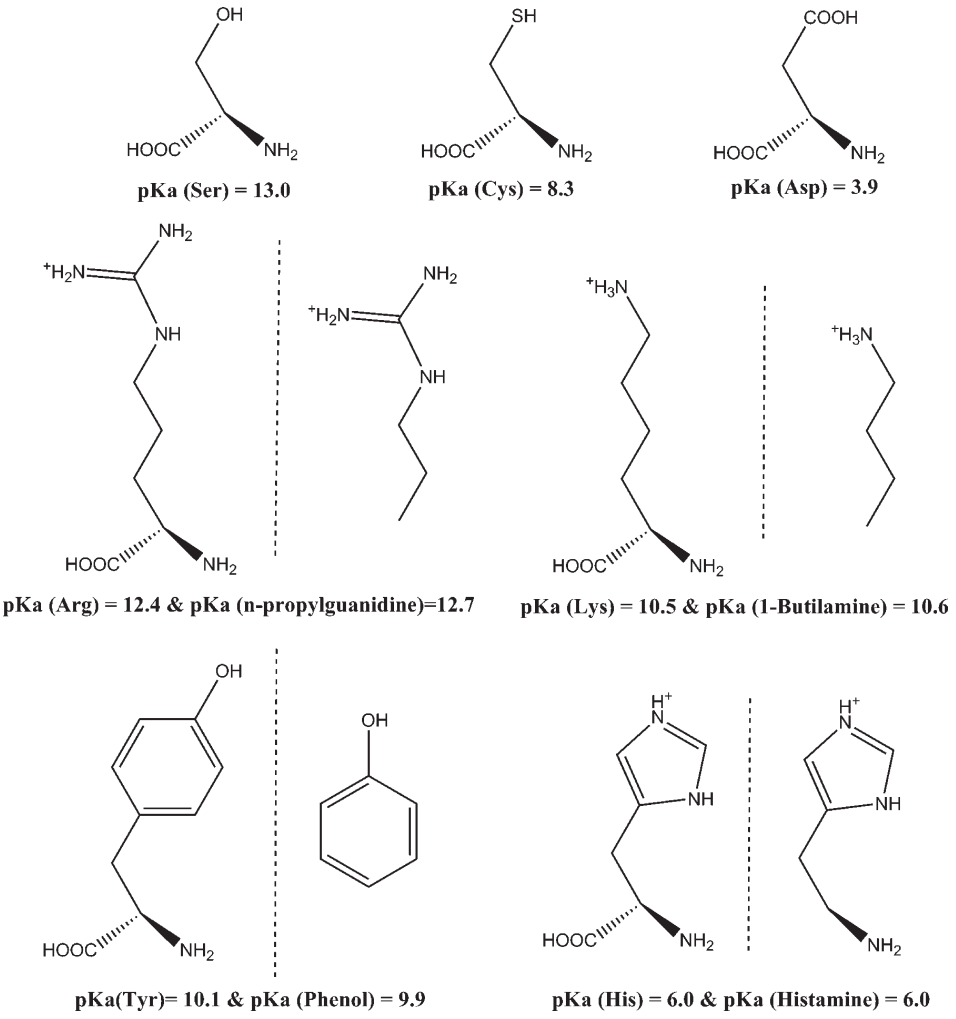

Figure 1. Representation of the structures and pK_a values of all ionizable amino acids and corresponding models.

approaches can never be as robust as would be desirable in this context. This said, we admit that in many instances, within the actual level of development of electronic structure calculations, knowledge-based methods can lead to smaller average errors, as the calculation of long-range interactions and the inclusion of flexibility in proteins is prohibitively time-consuming with electronic structure methods. However, electronic structure calculations have the potential of retrieving extremely precise results, both in usual and unusual environments with comparable accuracy. This is the basic reason why the scientific community is investing so much effort to predict protonation states with high-level electronic structure calculations. Although some studies have been made with wave function methods,[7,13,14] the most frequently used methods to determine the protonation state of the residues are based on Density Functional Theory (DFT),[10,15] which can be quite effective with much less computational time if the proper functionals are chosen.

However, it is not trouble-free to properly choose a DFT functional among the very large number currently available. Moreover, new density functionals continue to appear, taking the performance of DFT to new levels of accuracy. The selection of the functional to use will have to depend on the problem at hand. To correctly choose the DFT functionals, we should look for new benchmarking studies that test several density functionals against experimental data or higher-level methods on the specific system under consideration.[15−20] Our goal here is to analyze how well DFT functionals describe the protonation state of the ionizable side chains of the amino acids Lys, His, Arg, Asp, Cys, Ser, and Tyr.

The proton affinity at a given temperature is, by definition, the enthalpy of reactions 1 or 2, depending on whether the base is negatively charged or not.

$$BH^+ \underset{\Delta H_R = PA}{\rightarrow} B + H^+ \quad (1)$$

$$B \underset{\Delta H_R = PA}{\rightarrow} B^- + H^+ \quad (2)$$

In this work, instead of calculating the proton affinity (PA), we have calculated the zero-point-exclusive proton affinity at 0 K (the electronic proton affinity, PA_{el}^{0K}) for the ionizable amino acid side chains. This corresponds to the difference in the electronic energy between the deprotonated and protonated side chains only.

We have not calculated the entropy of the process (which would be needed in order to calculate gas-phase acidities/basicities) because its translational and rotational contributions are equivalent for all of the functionals, and the zero point energy

(ZPE) (and consequently the vibrational entropy) is not the quantity that we intend to benchmark. Additionally, by using the same geometry for all functionals, we can compare more clearly the results with high-level CCSD(T) calculations, without contributions from different geometries or different vibrational frequencies.

The difference in the electronic energy between the deprotonated and protonated side chains is by far the dominant and the most sensitive factor in the proton affinity. It is also the quantity that has the slower convergence along with the quality of the theoretical level.

We have used two *ab initio* methods, MP2 (second Moller–Plesset perturbation theory) and CCSD(T) (single and double excitation coupled cluster theory with perturbative triple correction), and extrapolated the energy to the complete basis set limit (CBS) to compute the PA_{el}^{0K} at the very accurate CCSD(T)/CBS level.

The comparison of the extrapolated CCSD(T)/CBS PA_{el}^{0K} with the results obtained from the calculations with 64 density functionals provides the capabilities and limitations of current DFT functionals to describe the ionizable amino acid side chains. Therefore, we deliver an important and consistent DFT benchmarking database, making this work relevant to all studies with ionizable amino acids side chains in which the use of very accurate post-HF methods is not yet feasible, e.g., most mechanistic studies with biological systems.

COMPUTATIONAL DETAILS

Models of the Ionizable Amino Acids Side Chains. We have used the complete molecules of the smallest amino acids Asp, Ser, and Cys, while for the bigger Tyr, Arg, Lys, and His (for which the use of the CCSD(T) methods with very large basis sets is impractical), we have used models that represent the ionizable amino acid side chains. All molecules have been modeled with neutral terminations to mimic better their electronic structure when they are inserted in a protein backbone. Figure 1 shows the molecules used here. The models that represent the Tyr, Lys, Arg, and His side chains have an experimental pK_a very close to the respective amino acid, therefore constituting excellent models to study the pK_a of the side chains of the real residues. A phenol molecule (pK_a = 9.9)[21] models the Tyr side chain (pK_a = 10.1).[22] An *n*-propylguanidine (pK_a = 12.7) models the Arg side chain (pK_a = 12.4),[23] and a 1-butilamine (pK_a = 10.6)[24] models the Lys side chain (pK_a = 10.5).[25] Concerning His, we have chosen a histamine molecule as a model because both have the same pK_a value (pK_a = 6.0).[26] The side chain of this amino acid has one δ nitrogen atom (near the backbone) and one ε nitrogen atom (far from backbone), which have a similar tendency to lose their protons (the difference in energy between the two protonated species is approximately 0.5 kcal/mol).[7] Therefore, we have studied only the deprotonation of the ε nitrogen atom. This article is focused on these structures/models, and therefore whenever the amino acids are mentioned throughout the text, these models are meant. Figure 1 shows all of the molecules used in the calculations.

Post-HF *ab Initio* Calculations. We have started by fully optimizing the geometry of each amino acid in implicit solvent at the MP2/6-311+G(d,p) level. It would have been more consistent to calculate the proton affinities with the minimum-energy geometry given by each of the density functionals instead of using MP2 geometries. However, the optimization at the MP2/6-311+G(d,p) theoretical level is known to give good geometries, and it is not biased toward any functional. The solvent was modeled through the Polarizable Continuum Model (PCM) using the polarizable conductor calculation model (CPCM)[27,28] and the dielectric constant and radius of water. The purpose was to avoid artificial intramolecular interactions due to the lack of environment. The optimized geometries are shown in the Supporting Information. There are no internal hydrogen bonds within the molecules.

For each optimized structure, we have carried out single point energy calculations in a vacuum at the frozen-core CCSD(T) and frozen-core MP2 levels using the Pople basis sets 6-31G*(0.25), 6-31+G(d), and 6-31+G(d,p) and the Dunning basis sets aug-cc-pVDZ and aug-cc-pVTZ. The modified 6-31G*(0.25) basis set includes d-polarization functions with an exponent of 0.25, which makes them more diffuse than the standard d orbitals and simultaneously accounts for polarization and diffusion. We have also calculated the energies with the aug-cc-pVQZ basis set at the MP2 level. Considering the large size of our models, it was unfeasible to use the aug-cc-pVQZ basis set with CCSD(T). The aug-cc-pVTZ basis set may give less accurate results for the sulfur atom present in the Cys side chain.[29] However, we will only analyze energy differences and not the absolute energies. Therefore, most of these eventual inaccuracies will cancel out when we calculate PA_{el}^{0K}.

Extrapolation of the Energies to the Basis Set Limit. The Dunning basis sets constitute hierarchical sequences that enable us to extrapolate the correlation energy to the CBS limit.

To extrapolate the energies to the MP2/CBS level, we have employed both the Truhlar and the Helgaker extrapolation schemes. The total energy at the CBS limit ($E_{MP2/CBS}$) was obtained as the sum of the CBS Hartree–Fock energy ($E_{HF/CBS}$) and the CBS correlation energy ($E_{corr/CBS}$) (eq 3).

$$E_{MP2/CBS} = E_{HF/CBS} + E_{corr/CBS} \tag{3}$$

The extrapolations of the Hartree–Fock energy and of the correlation energy were made independently because the convergence characteristics are very different, with the Hartree–Fock energy converging much faster than the correlation energy. The Truhlar extrapolation employed here uses the aug-cc-pVXZ (X = 2 and 3) basis sets (see eq 4).[30,31] The parameters α and β in eq 4 have the values of 4.93 and 2.13.[32] $E_{corr/DZ}$ energies were obtained with MP2/aug-cc-pVDZ, and $E_{corr/TZ}$ energies were obtained with MP2/aug-cc-pVTZ.

$$E_{MP2/CBS}^{Truhlar} = \frac{3^{\alpha}}{3^{\alpha}-2^{\alpha}}E_{HF/TZ} - \frac{2^{\alpha}}{3^{\alpha}-2^{\alpha}}E_{HF/DZ} + \frac{3^{\beta}}{3^{\beta}-2^{\beta}}E_{corr/TZ} - \frac{2^{\beta}}{3^{\beta}-2^{\beta}}E_{corr/DZ} \tag{4}$$

While the HF energy converges rather quickly with the basis set size, the slow basis set convergence of the correlation energy is usually considered to be the major obstacle in obtaining CBS energies.[31−35] However, even though the correlation energy converges slower, it seems to do so in a more systematic way, and thus it may give rise to a lower extrapolation error.

To extrapolate the correlation energy to the CBS limit using larger basis sets (beyond the triple-ζ quality), we have also used the Helgaker et al. extrapolation scheme.[36] The corresponding equation when the aug-cc-pVTZ and aug-cc-pVQZ basis sets are used is presented in eq 5, in which $E_{corr/TZ}$ is the correlation

energy at the MP2/aug-cc-pVTZ level and $E_{corr/QZ}$ is the correlation energy at the MP2/aug-cc-pVQZ level.

$$E_{corr/CBS} = \frac{4^3 E_{corr/QZ} - 3^3 E_{corr/TZ}}{4^3 - 3^3} \quad (5)$$

We could have used the correlation energies resulting from the aug-cc-pVDZ and aug-cc-pVTZ basis sets in this extrapolation. However, this would lead to very large errors in the extrapolated results.[37]

It is generally considered that the extrapolation error is dominated by the correlation energy. This tendency is also clearly present in our study (please see the Supporting Information). Therefore, the choice of the method to extrapolate the HF energy (typically polynomial or exponential) does not strongly affect the results. Here, we have used the Truhlar scheme to extrapolate the HF energy in all cases. We also tested the method of Helgaker.[38] However, its coefficient was optimized with results calculated using the cc-pVXZ basis sets, and therefore it was less appropriate for our calculations than the method of Truhlar. Nevertheless, the difference between the HF/CBS values using both extrapolation schemes was small (mean unsigned difference of 0.10 kcal/mol). Probably a significant part of the difference does not come from the extrapolation schemes themselves but from the use of a coefficient optimized for the cc-pVXZ basis sets.

The final CBS energy corresponds to the sum of the CBS HF energy, extrapolated using the Truhlar scheme, and the CBS correlation energy, extrapolated with the Helgaker et al.:

$$E^{Helgaker}_{MP2/CBS} = E_{HF/CBS} + E^{Helgaker}_{corr/CBS} \quad (6)$$

We have calculated the MP2/CBS correlation energies with both extrapolation schemes discussed above, for all amino acid side chains considered in this work.

The difference between the CCSD(T) and MP2 energies ($\Delta E_{CCSD(T)-MP2}$) has a small basis set dependence in many chemical systems (see, e.g., reference[39]). To estimate the CCSD(T)/CBS energy, we have calculated $\Delta E_{CCSD(T)-MP2}$ and added this difference to the MP2/CBS energy. This is a standard procedure, based on the assumption that the difference between MP2 and CCSD(T) has a small dependence on the basis set. The CCSD(T)/CBS electronic energy values can thus be given by eq 7.

$$E_{CCSD(T)/CBS} = E^{Helgaker}_{MP2/CBS} + \Delta E_{(CCSD(T)-MP2)/TZ} \quad (7)$$

To evaluate how dependent from the basis set this term is, we have calculated $\Delta E_{CCSD(T)-MP2}$ with the 6-31G*(0.25), 6-31+G(d), and 6-31+G(d,p) and the aug-cc-pVDZ and aug-cc-pVTZ basis sets.

We could have calculated the CCSD(T)/CBS energy extrapolating directly from the CCSD(T)/aug-cc-pVDZ and aug-cc-pVTZ values (without using the MP2 energies), using the method of Helgaker et al.[37] However, the procedure was shown to lead to significant errors, due to the use of the CCSD(T)/aug-cc-pVDZ energy in the extrapolation, which is still too far from the CBS limit to be beneficial for the extrapolation.[37,39] We tested this method, and indeed we could find differences of about 2 kcal/mol in relation to the more reliable extrapolation based on eq 7. To benchmark the density functionals, we have used $\Delta E_{CCSD(T)-MP2}$ calculated with the aug-cc-pVTZ basis set. The other basis sets were used only for the purpose of studying the convergence of $\Delta E_{CCSD(T)-MP2}$. Therefore, whenever we refer simply to the CCSD(T)/CBS energy in this work, we are assuming that $\Delta E_{CCSD(T)-MP2}$ was calculated with the largest basis set. The reference value of PA^{0K}_{el} for the DFT bechmarking is defined as eq 8:

$$PA^{0K}_{el} = E^{A^-}_{CCSD(T)/CBS} - E^{HA}_{CCSD(T)/CBS} \quad (8)$$

The differences in PA^{0K}_{el} between amino acid pairs (ΔPA^{0K}_{el}) were also calculated to evaluate the accuracy of the DFT calculation of the energy involved in proton transfers between side chains, an extremely common situation in, e.g., reactions catalyzed by enzymes. ΔPA^{0K}_{el} may eventually benefit from the cancellation of the systematic errors resulting from the calculation of PA^{0K}_{el} and is the quantity of interest in many mechanistic studies.

DFT Calculations. As the accuracy of DFT strongly depends upon the choice of the functional used for the specific system and property, single point energy calculations with 64 density functionals (2 of the local spin density approximation (LSDA) type, 16 of the generalized gradient approximation (GGA) type, 8 of the meta-GGA (m-GGA) type, 21 of the hybrid-GGA (h-GGA) type, 14 of the hybrid-meta GGA (hm-GGA) type, and 3 double hybrid-GGA (hh-GGA) type; please see Table 5 for a complete description of the functionals used) were carried out using the 6-311++G(2d,2p) basis set. We have chosen the 6-311++G(2d,2p) basis set because it was the most complete basis set that can be used with most biological systems that typically include over 100 atoms. Calculations with more complete basis sets would become intolerably time-consuming in large systems. Moreover, this basis set is usually close to the CBS limit with DFT, which converges rather quickly with basis set size,[40] and differences in energy arising from the use of larger basis sets usually amount to a few tenths of kilocalorie per mole. The accuracy of DFT results is thus determined by both the truncation error of the basis set employed and the intrinsic error due to the approximate nature of each functional.

Post-HF and DFT calculations were carried out using the Gaussian 03 suite of programs,[41] except for the double hybrids and dispersion-corrected (B3LYP-D and B97-D) functionals that were carried out using the Gaussian 09 suite of programs.[42] We have used the default grid sizes. The meta-hybrid density functionals are mostly sensitive to the grid size (in particular the M06 suite, which is known to be extremely sensitive in this regard).[43] To check whether the default grid was accurate enough, we tested the default grid (a pruned (75,302) grid) and the ultrafine grid (a pruned (99 590) grid). We have repeated the calculations with the M06 and M06-2X functionals, and the largest difference obtained was only 0.03 kcal/mol in PA^{0K}_{el} (average absolute difference of 0.02 kcal/mol). The influence of the grid size used was found to be insignificant on the systems and properties studied here.

Grimme's D3 correction[44] was applied to a few density functionals (B3LYP-D3, MPW1B95-D3, B1B95-D3, CAM-B3LYP-D3, and BMK-D3). The D3 correction was calculated using the program available online at Grimme's Web site[45] and using the Becke−Johnson damping.[46−48]

RESULTS AND DISCUSSION

Calculation of PA^{0K}_{el} at the MP2/CBS Level. There are few methods to accurately determine thermochemical parameters, and the extrapolation to the complete basis set (CBS) using a

Table 1. PA_{el}^{0K} at the MP2 Level and the Correlation-Consistent aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ Basis Sets (All Values in kcal/mol)

	PA_{el}^{0K} for each basis set		
amino acid	aug-cc-pVDZ	aug-cc-pVTZ	aug-cc-pVQZ
Asp^-	346.36	347.50	347.42
Cys^-	351.94	353.92	353.87
His	235.29	236.23	236.27
Ser^-	374.11	375.49	375.53
Lys	226.33	226.85	226.76
Arg	245.96	247.50	247.71
Tyr^-	352.49	354.21	354.36

series of correlation consistent bases together with high-level post-HF methods is among those that provide the best results. Here, we have used high-level post-HF *ab initio* methods (MP2 and CCSD(T)), which were corrected using basis set extrapolation techniques to obtain reference CCSD(T)/CBS energies. Two extrapolation schemes were used and compared. Subsequently, we performed a comparison between the results of 64 DFT functionals and the extrapolated post-HF *ab initio* methods to examine the performance of various DFT functionals in predicting the PA_{el}^{0K} values for every ionizable amino acid side chain. The absolute electronic energies calculated at the HF/aug-cc-pVXZ (X = 2−4), MP2/aug-cc-pVXZ (X = 2−4), and CCSD(T)/aug-cc-pVXZ (X = 2−3) theoretical levels can be found in Table S1 of Supporting Information. Table 1 shows the PA_{el}^{0K} values obtained at the MP2 level and the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.

Table 2 summarizes the PA_{el}^{0K} values at the MP2/CBS level (the separate HF and correlation energies can be found in Table S2 of the Supporting Information) obtained with the extrapolation scheme of Truhlar. The differences between PA_{el}^{0K} at the MP2/CBS and MP2/aug-cc-pVXZ (X = 2−4) levels are also shown. The results reveal that the MP2/CBS values are not closer to the quadruple-ζ result than to the triple-ζ result (both the MUE and the MSE to the CBS values are equivalent for the triple and quadruple basis sets). The convergence of PA_{el}^{0K} with the basis set size is impressively fast. The results with the aug-cc-pVTZ basis set are already at 0.17 kcal/mol from the CBS limit (in average), and the maximum difference between the triple-ζ and the CBS values is 0.29 kcal/mol. We conclude that the extrapolated values for these systems and property are not obviously more accurate than the triple-ζ values. However, this conclusion only holds for PA_{el}^{0K}, which is based on energy difference. If we consider the absolute electronic energies of the protonated and deprotonated species, then the extrapolated values are much closer to the quadruple-ζ results (average MUE = 0.08 kcal/mol, average MSE = −0.08 kcal/mol) than to the triple-ζ results (average MUE = 0.18 kcal/mol, average MSE = −0.18 kcal/mol). Further details can be found in Tables S2 and S3 in the Supporting Information.

We have calculated also the PA_{el}^{0K} at the MP2/CBS level using the Helgaker extrapolation scheme. This extrapolation involved the MP2/aug-cc-pVTZ and MP2/aug-cc-pVQZ electronic energies, as well as the Truhlar extrapolated HF energies. The values provided by this method are expected to be more accurate and reliable because they are based in larger basis sets than those used in the Truhlar method (they are also

Table 2. PA_{el}^{0K} Calculated at the MP2/CBS Level Using the Truhlar Extrapolation Scheme[a]

	PA_{el}^{0K}			
amino acids	MP2/CBS	MP2/CBS-MP2/aug-cc-pVDZ	MP2/CBS-MP2/aug-cc-pVTZ	MP2/CBS-MP2/aug-cc-pVQZ
Asp^-	347.56	1.20	0.06	0.14
Cys^-	354.08	2.14	0.15	0.20
His	235.99	0.70	−0.24	−0.28
Ser^-	375.68	1.57	0.19	0.15
Lys	226.71	0.38	−0.15	−0.05
Arg	247.42	1.45	−0.09	−0.30
Tyr^-	354.51	2.02	0.29	0.14
MUE		**1.35**	**0.17**	**0.18**
MSE		**1.35**	**0.03**	**0.00**

[a] The difference between PA_{el}^{0K} values obtained with a CBS and with the aug-cc-pVXZ (X = 2−4) is also shown. All values in kcal/mol.

Table 3. PA_{el}^{0K} Values at the MP2/CBS Level Obtained with the Helgaker Extrapolation Scheme Using the aug-cc-pVTZ and aug-cc-pVQZ Basis Sets, As Well As the Difference between the Truhlar and Helgaker et al. Extrapolation Methods (ΔPA_{el}^{0K}; All Values in kcal/mol)

amino acids	PA_{el}^{0K} (MP2/CBS)	ΔPA_{el}^{0K}
Asp^-	347.19	0.38
Cys^-	353.22	0.86
His	236.27	−0.28
Ser^-	375.37	0.31
Lys	226.69	0.02
Arg	247.84	−0.43
Tyr^-	354.31	0.20

much more expensive in terms of CPU power and seldom feasible in biological systems of interest with up to 100 atoms, in which the use of the aug-cc-pVTZ basis set is already a challenge). Table 3 shows the MP2/CBS PA_{el}^{0K} values obtained with the Helgaker method, as well as the difference between both extrapolated methodologies.

The data shows that the difference between the PA_{el}^{0K} values at the MP2/CBS level extrapolated with both methods is very small in most cases (absolute average of 0.35 kcal/mol only). The only exception is Cys, which shows a more significant deviation (−0.86 kcal/mol). Even though the use of such large basis sets might be very useful in applications involving a very limited number of atoms/electrons, if we consider typical (larger) biological systems, the extrapolation using quadruple-ζ basis sets will probably be detrimental for the quality of the overall results. The application of this large basis set would force the researcher to reduce significantly the size of the molecular models, introducing errors much larger than the few tenths of a kilocalorie per mole that can result in terms of accuracy (assuming that the difference between both methods is translated into a gain in accuracy for the extrapolation using larger basis sets).

Calculation of PA_{el}^{0K} at the CCSD(T)/CBS Level. The CCSD(T)/CBS energies were obtained adding to the MP2/CBS energies the energy difference between CCSD(T) and MP2

Table 4. $\Delta E_{CCSD(T)-MP2}$ with the Pople Basis Sets (6-31G*(0.25), 6-31+G(d), and 6-31+G(d,p)) and Correlation-Consistent Basis Sets (aug-cc-pVDZ and aug-cc-pVTZ)[a]

	$\Delta E_{CCSD(T)-MP2}$					PA_{el}^{0K}
amino acids	6-31G*(0.25)	6-31+G(d)	6-31+G(d,p)	aug-cc-pVDZ	aug-cc-pVTZ	CCSD(T)/CBS
Asp^-	1.23	2.58	2.76	2.60	2.61	349.80
Cys^-	2.29	1.72	1.77	2.81	2.34	355.55
His	2.39	2.35	2.40	2.43	2.58	238.84
Ser^-	3.20	2.30	2.53	2.37	2.53	377.90
Lys	1.27	1.09	1.26	1.18	1.34	228.03
Arg	2.68	1.18	1.39	1.40	1.62	249.47
Tyr^-	3.28	3.17	3.33	3.05	3.05	357.36
MUE	**0.52**	**0.27**	**0.21**	**0.17**	**0.00**	
MaxE	**1.38**	**0.62**	**0.57**	**0.47**	**0.00**	

[a] The mean unsigned error (MUE) and maximum error (MaxE) between the $\Delta E_{CCSD(T)-MP2}$ values of each basis set and the largest basis set (aug-cc-pVTZ) are also shown. The $\Delta E_{CCSD(T)-MP2}$ value with the aug-cc-pVTZ was used to calculate the PA_{el}^{0K} for each amino acid extrapolated to the CCSD(T)/CBS level (last column on the right of the table). All values in kcal/mol.

($\Delta E_{CCSD(T)-MP2}$). This last term was calculated with several basis sets to check for its convergence with basis set size. The $\Delta E_{CCSD-MP2}$ and PA_{el}^{0K} values obtained in this way are shown in Tables 4 and 5.

The difference between CCSD(T) and MP2 is significant, as expected. However, such a difference seems to be efficiently accounted for by the $\Delta E_{CCSD-MP2}$ term with the aug-cc-pVTZ basis set. The average MUE $\Delta E_{CCSD-MP2}$ difference between the double and triple-ζ basis sets is only 0.17 kcal/mol, and the MaxE is 0.47 kcal/mol. We cannot calculate exactly the accuracy of the $\Delta E_{CCSD-MP2}$ term because we would need to know the CCSD(T)/CBS energy extrapolated from the CCSD(T) energies only. As we have only the double- and triple-ζ basis sets results with this method, we cannot obtain an accurate extrapolation of the CCSD(T) just from them. However, the data in Table 4 clearly show that the accuracy of $\Delta E_{CCSD-MP2}$ increases monotonically with basis set size (both MUE and MaxE) and that the 6-31G*(0.25) basis set should be avoided in these systems as the deviation for the largest basis set is significantly large. The aug-cc-pVTZ basis set should be used, at least to obtain predictions that safely converge within chemical accuracy. However, this basis set is prohibitively large for many applications at the CCSD(T) level.

On the basis of these observations, we have used the $\Delta E_{CCSD-MP2}$ correction with the aug-cc-pVTZ basis set and the MP2/CBS values obtained with the Helgaker et al. extrapolation scheme to obtain the PA_{el}^{0K} at the CCSD(T)/CBS level.

The MP2/CBS extrapolated energy differences converge much better and faster than absolute energies but are still approximations. The frozen-core approximation and the conversion of MP2/CBS into CCSD(T)/CBS through eq 7 also introduce small errors in the final values. Others sources of error (e.g., the use of aug-cc-pVTZ to describe the sulfur atom or the truncation of the expansion of the cluster operator) are present as well. Altogether, these values may result in an uncertainty of a few tenths of a kilocalorie per mole (the exact value is difficult to estimate). As these approximated CCSD(T)/CBS energy differences were used to benchmark the density functionals, a part of the error attributed to DFT might arise from the uncertainty in the reference CCSD(T)/CBS values, which may slightly affect the order in which the functionals are classified.

Benchmarking of Density Functionals. This assessment does not represent the global quality of the functionals, which must be measured through the calculation of diverse sets of properties in a representative set of molecular systems. Instead, this study just evaluates specifically the accuracy of the description of the zero-point exclusive proton affinity at 0 K for amino acid side chains. We have benchmarked 64 density functionals (2 LSDA, 16 GGA, 8 m-GGA, 21 h-GGA, 14 hm-GGA, and 3 hh-GGA; please see Table 5 for a complete description of the functionals used) through single point energy calculations at the DFT/6-311++G(2d,2p)//MP 2/6-311++G(d,p) level. The mean unsigned error (MUE PA_{el}^{0K}) and maximum error (MaxE PA_{el}^{0K}) between these values and the CCSD(T)/CBS value is shown in Table 5.

All of these values are important in order to understand the accuracy provided by each functional in, e.g., biological acid/base catalysis, in which some of these amino acids are involved in the catalytic cycles. We have divided the density functionals into three groups according to the ΔPA_{el}^{0K} MUE. In group I, the MUE ranges from 0 to 1.42 kcal/mol (which corresponds to an error below 1 pK_a unit at 310.15 K). Group II includes the functionals with MUE between 1.42 and 2.84 kcal/mol (which corresponds to an error between 1 and 2 pK_a units at 310.15 K). Group III includes functionals with a MUE above 2.84 kcal/mol (errors above 2 pK_a units). The density functionals that provide the more accurate PA_{el}^{0K} values are MPW1B95-D3, XYG3, MPW1B95, B1B95-D3, BMK, BMK-D3, M06-2X, and B1LYP. They have a MUE close to chemical accuracy and MaxE of circa 2 kcal/mol. The remaining functionals of group I (B1B95, PBE1PBE, CAM-B3LYP, B97-1, PBE1KCIS, B3P86, CAM-B3LYP-D3, B3LYP, B98, and M06-L) and both M06 and B3LYP-D (marginally in group II but better than any other of this group) also show very satisfactory performance with MUEs below 1.48 kcal/mol and MaxE's of *circa* 3 kcal/mol. MPW1B95 in particular has the lowest MUE and one of the lowest MaxE's. All of these functionals are nonlocal, even though there is no correlation between the MUE and the fraction of HF exchange. The very popular B3LYP functional ranks 16th in terms of MUE and 11th in terms of MaxE, belonging to the group of the most accurate functionals. The performance of the dispersion-corrected B3LYP-D functional ranks 19th and 18th in terms of MUE and MaxE.

Table 5. Differences in the PA_{el}^{0K} Values Calculated at the DFT/6-311++G(2d,2p)//MP2/6-311++G(d,p) Level and at the CCSD(T)/CBS Level, for Each Amino Acid[a]

functional	type	%HFχ	MUE	MaxE	Asp^-	Cys^-	His	Ser^-	Arg	Lys	Tyr^-
MPW1B95-D3[44,49−52]	hm-GGA	31	0.79	2.25	−0.24	0.50	1.05	−0.72	2.25	0.55	−0.24
XYG3[53,54]	hh-GGA	80.33	0.83	1.87	−1.04	0.70	−0.50	−1.28	0.15	0.29	−1.87
MPW1B95[49−52]	hm-GGA	31	0.85	2.49	−0.03	0.78	1.29	−0.56	2.49	0.78	−0.03
B1B95-D3[44,50,55]	hm-GGA	28	0.91	2.29	−0.13	0.67	1.66	−0.73	2.29	0.66	−0.25
BMK[56]	hm-GGA	42	0.93	1.83	−0.72	−1.66	1.06	0.10	1.83	0.25	−0.90
BMK-D3[44,56]	hm-GGA	42	1.00	2.21	−1.10	−2.21	0.60	−0.19	1.40	−0.20	−1.29
M06-2X[17,57]	hm-GGA	54	1.08	2.09	−0.47	−2.09	−1.64	−0.25	−0.81	−1.52	−0.78
B1LYP[55,58,59]	h-GGA	25	1.10	1.92	−1.17	−0.06	1.27	−1.92	1.88	0.34	−1.05
B1B95[50,55]	hm-GGA	28	1.11	2.77	0.28	1.28	1.66	−0.40	2.77	1.16	0.18
PBE1PBE[60]	h-GGA	25	1.17	2.98	0.17	0.79	1.79	−0.69	2.98	1.39	−0.40
CAM-B3LYP[61]	h-GGA	19/65**	1.18	2.00	−2.00	−0.57	0.25	−1.72	1.18	−0.99	−1.51
B97−1[62]	h-GGA	21	1.31	2.91	0.24	0.73	2.23	−0.96	2.91	1.81	−0.26
PBE1KCIS[60,63,64]	hm-GGA	22	1.31	2.85	−0.24	0.93	1.96	−1.31	2.85	1.30	−0.57
B3P86[50,51,55]	h-GGA	20	1.32	3.11	0.03	1.21	2.13	−0.95	3.11	1.49	−0.32
CAM-B3LYP-D3[44,61]	h-GGA	19/65**	1.33	2.30	−2.30	−0.96	−0.11	−1.96	0.83	−1.35	−1.82
B3LYP[50,55,59]	h-GGA	20	1.34	2.78	−1.60	−0.18	1.21	−2.78	1.78	0.22	−1.61
B98[65]	h-GGA	21.98	1.35	3.17	0.44	0.79	2.47	−0.57	3.17	1.95	0.06
M06-L[66]	m-GGA	0	1.38	2.95	1.14	2.95	1.27	−1.19	2.26	0.56	−0.30
B3LYP-D[67−69]	h-GGA	20	1.48	3.00	−1.08	0.29	1.87	−2.13	3.00	1.04	−0.93
M06[17,57]	hm-GGA	27	1.48	2.95	−1.42	−0.06	−1.09	−2.95	0.32	−2.26	−2.30
B3LYP-D3[44,50,55,59]	h-GGA	20	1.56	3.28	−2.23	−1.05	0.42	−3.28	1.01	−0.60	−2.29
B3PW91[50,55,70]	h-GGA	20	1.63	3.68	0.58	1.89	2.67	−0.42	3.68	2.08	0.11
MPW3LYP[49,59,70]	h-GGA	21.8	1.65	3.67	−2.35	−0.96	0.63	−3.67	1.19	−0.39	−2.35
MPW1KCIS[16,49,64,70]	hm-GGA	15	1.66	2.85	−0.60	1.22	2.19	−2.26	2.85	1.49	−1.02
MPW1PW91[49,70]	h-GGA	25	1.72	3.77	0.88	1.90	2.61	0.25	3.77	2.14	0.51
TPSS1KCIS[64,71,72]	hm-GGA	13	1.75	2.84	−0.48	1.46	2.32	−2.26	2.84	2.14	−0.75
O3LYP[59,73,74]	h-GGA	11.61	1.86	4.01	0.77	2.35	3.09	0.06	4.01	2.73	−0.03
TPSSh[71]	hm-GGA	10	1.89	3.12	−0.14	1.96	2.79	−2.02	3.12	2.81	−0.40
MPW1S[49,70,75]	h-GGA	[b]	1.97	4.91	−2.11	−0.05	1.28	−4.91	1.82	0.69	−2.95
VSXC[76]	m-GGA	0	2.10	5.16	−0.97	0.51	3.27	−5.16	3.05	1.47	−0.27
BHandH[55,59]	h-GGA	50	2.15	4.19	−2.52	−4.19	−1.31	−1.81	0.66	−2.26	−2.29
BPBE[55,60]	GGA	0	2.23	5.55	−2.13	0.23	1.56	−5.55	1.93	1.04	−3.18
G96LYP[59,77]	GGA	0	2.25	6.07	−2.88	−0.36	1.40	−6.07	1.35	0.36	−3.33
MPWB1K[78]	hm-GGA	44	2.26	3.78	1.73	1.85	2.19	2.31	3.78	1.76	2.16
BPW91[55,70]	GGA	0	2.26	5.46	−2.08	0.39	1.68	−5.46	2.04	1.10	−3.04
HCTH147[62]	GGA	0	2.27	3.92	−0.53	1.39	2.87	−3.92	3.44	2.26	−1.46
TPSSTPSS[71]	m-GGA	0	2.33	4.71	−1.65	1.09	2.28	−4.71	2.24	2.32	−2.03
OLYP[59,73]	GGA	0	2.34	4.39	−0.58	1.42	2.63	−4.39	3.23	2.42	−1.74
HCTH407[62]	GGA	0	2.41	4.10	−0.38	1.31	3.11	−4.08	4.10	2.49	−1.43
TPSSLYP1W[59,71,79]	m-GGA	0	2.42	6.08	−3.00	0.29	2.00	−6.08	1.30	1.61	−2.62
B97-D[68]	GGA	0	2.43	4.30	−0.30	1.55	3.44	−3.46	4.30	3.04	−0.89
MPWKCIS[49,64,70]	m-GGA	0	2.45	6.81	−3.20	−0.58	1.10	−6.81	1.25	0.31	−3.90
MPWPW91[49,70]	GGA	0	2.45	6.81	−3.19	−0.80	0.85	−6.81	1.18	0.22	−4.11
BB1K[50,55,80]	hm-GGA	42	2.56	4.04	2.06	2.29	2.50	2.57	4.04	2.08	2.40
PBE1W[60,79]	GGA	0	2.66	7.52	−3.73	−1.46	0.42	−7.52	0.63	−0.27	−4.60
wB97X-D[81,82]	h-GGA	22.20/100[c]	2.67	4.82	1.59	2.39	3.55	1.85	4.82	2.83	1.67
BP86[51,55]	GGA	0	2.70	7.48	−3.83	−1.96	0.01	−7.48	0.15	−0.55	−4.88
HCTH93[62]	GGA	0	2.70	4.79	0.99	3.09	4.08	−2.18	4.79	3.62	−0.11
BB95[50,55]	m-GGA	0	2.74	8.01	−4.03	−1.43	−0.07	−8.01	0.12	−0.75	−4.75
BHandHLYP[55,59]	h-GGA	50	2.74	4.26	1.98	1.82	2.88	3.29	4.26	2.05	2.91
B2PLYP[83]	hh-GGA	53	2.78	4.13	2.18	1.63	2.87	3.51	4.13	2.19	2.94
B97−2[62]	h-GGA	21	2.86	4.78	2.10	3.28	3.71	1.21	4.78	3.34	1.65
PBEPBE[60]	GGA	0	3.02	8.15	−4.22	−2.28	−0.23	−8.15	0.12	−0.79	−5.34

Table 5. Continued

functional	type	%HFχ	MUE	MaxE	Asp^-	Cys^-	His	Ser^-	Arg	Lys	Tyr^-
mPW2PLYP[84]	hh-GGA	55	3.28	5.17	2.48	1.79	3.30	4.00	5.17	2.75	3.47
PBELYP1W[59,60,79]	GGA	0	3.29	8.83	−4.94	−2.37	−0.11	−8.83	−0.34	−1.19	−5.30
BLYP[55,59]	GGA	0	3.52	8.87	−5.10	−2.69	−0.38	−8.87	−0.63	−1.44	−5.55
MPW1N[49,70,85]	h-GGA	*b*	3.62	5.33	3.00	3.20	3.68	3.73	5.33	3.29	3.11
MPWKCIS1K[16,49,64,70]	hm-GGA	41	3.86	5.49	3.11	3.57	4.01	3.96	5.49	3.44	3.47
MPW1K[49,70,75,86]	h-GGA	42.8	3.87	5.54	3.28	3.37	3.83	4.18	5.54	3.45	3.46
MPWLYP1W[49,59,70,79]	GGA	0	3.95	9.52	−5.60	−3.16	−0.68	−9.52	−0.93	−1.77	−5.99
MPWB95[49,50,70]	m-GGA	0	4.05	9.37	−5.14	−2.65	−0.92	−9.37	−2.81	−1.63	−5.84
MPWLYP[49,59,70]	GGA	0	4.57	10.23	−6.19	−3.92	−1.22	−10.23	−1.48	−2.31	−6.62
SVWN3[87,88]	LSDA	0	7.45	13.08	−9.02	−8.00	−3.84	−13.08	−2.79	−5.36	−10.03
SVWN5[87,88]	LSDA	0	8.25	14.04	−9.74	−9.07	−4.49	−14.04	−3.54	−5.97	−10.91

[a] The mean unsigned error (MUE PA_{el}^{0K}) and maximum error (MaxE PA_{el}^{0K}) are also shown. All values in kcal/mol. [b] Not available. [c] The first value is %HFχ at short range, and the second value is %HFχ at long range.

The inclusion of dispersion effects (D correction) in B3LYP systematically increased the PA value and worsened the overall results (+0.14 kcal/mol in the MUE and +0.23 kcal/mol in the MaxE). The opposite tendency was seen with the D3 correction. Here, the correction, in all of the five tested functionals and in every amino acid, decreased the PA values by 0.2−0.7 kcal/mol. In three functionals, the D3 correction also worsened the results (B3LYP-D3, +0.22 in MUE and +0.50 in MaxE; CAM-B3LYP-D3, +0.16 in MUE and +0.30 in MaxE; BMK-D3, +0.07 in MUE and +0.37 in MaxE), while in the other two functionals, the results were improved (MPW1B95-D3, −0.06 in MUE and −0.24 in MaxE; B1B95-D3, −0.19 in MUE and −0.48 in MaxE). In general, the influence of the dispersion corrections in these systems is small (always below 1.2 kcal/mol, 0.5 kcal/mol in average). As we deal with single, small molecules, there are few dispersive interactions in the zero overlap region, where dispersion is unaccounted for by density functionals. Moreover, as the geometries of the protonated and unprotonated amino acids are very similar, and the difference is just a proton, most of the dispersion cancels out when we calculate PA_{el}^{0K}.

In general, there is a positive correlation between MUE and MaxE. The MaxE for each density functional is not always provided by the same amino acid. If we consider only groups I and II (MU E < 2.84 kcal/mol, pK_a errors < 2), we can see that there are no significant differences in the MUEs of each amino acid (averaged over all of those functionals), except for serine, in which the MUE is larger on average (MUE(Asp^-) = 1.4 kcal/mol, MUE(Cys^-) = 1.3 kcal/mol, MUE(His) = 1.8 kcal/mol, MUE(Ser^-) = 2.9 kcal/mol, MUE(Arg) = 2.4 kcal/mol, MUE(Lys) = 1.4 kcal/mol, MUE(Tyr^-) = 1.6 kcal/mol). With very few exceptions M06, M06-2X, CAM-B3LYP-D3, CAM-B3LYP, and BMK-D3 systematically underestimated PA_{el}^{0K} while B98, B1B95, B97-1, B3P86, and M06-L overestimated this value.

In recent papers, it was suggested that M06-2X and M06 were the best functionals for a combination of main-group thermochemistry, kinetics, and noncovalent interactions.[17,57] Other studies have concluded that B3LYP and B98 were very precise in the calculation of proton affinities for small models such as pirrole, quinoline,[89] γ-butyrolactone, and 2-pyrrolidinone.[14] Our results show that all of these hybrid and hybrid-meta functionals are very good to describe the PA_{el}^{0K} of the ionizable amino acid side chains, with errors below 1 pK_a unit.

Many biological mechanisms involve proton transfers between ionizable amino acid side chains (e.g., reaction mechanisms in enzymes). Therefore, it becomes very important to evaluate the performance of the density functionals in systems that involve proton transfers between different amino acids. For that purpose, we have calculated the accuracy of the proton transfer energies (ΔE_{PT}) between every pair of the seven studied amino acids for the 20 more accurate density functionals (MPW1B95, PBE1KCIS, M06-2X, B1LYP, B1B95, PBE1PBE, B97-1, B3P86, B3LYP, B98, M06, M06-L, MPW1B95-D3, B1B95-D3, BMK, BMK-D3, CAM-B3LYP, CAM-B3LYP-D3, B3LYP-D, and XYG3). The accuracy was measured by the MUE between the ΔE_{PT} energies calculated at the DFT level and at the CCSD(T)/CBS level.

Table 6 shows the errors (differences between the DFT and CCSD(T)/CBS energies) obtained for each proton transfer between every amino acid pair. If systematic errors are present in PA_{el}^{0K}, the accuracy of ΔE_{PT} will be greater than the accuracy of PA_{el}^{0K}. In the absence of systematic errors, the error should be larger in ΔE_{PT} than in PA_{el}^{0K} (due to error propagation). Table 7 shows the MUE in ΔE_{PT} for the 20 density functionals.

The present results can be used in many ways. To begin with, they clearly show that M06-2X is the most accurate density functional for proton transfers, in general, and B3LYP and B3LYP-D are the less accurate, in general, among the 20 best rated to predict PA_{el}^{0K} values. On the other hand, if one wants to be very specific, the ΔE_{PT} values are useful in identifying the best density functional to describe each specific enzymatic system. For example, for the pancreatic elastase catalytic mechanism[90] in which a proton transfer occurs from one serine to one histidine, we see that the three density functionals that provide the most accurate results are XYG3, BMK-D3, BMK, and M06-2X. Therefore, one of these four functionals might be used in proteins belonging to the serine protease family. Similarly, the results suggest that the M06-L, M06, XYG3, and M06-2X density functionals are the best to describe the proton transfer from carboxylic acids to histidines that occurs in the sulfotransferase enzymes.[91] However, we must keep in mind that the energetics of enzyme reactions depends on many more factors beyond proton transfers and that the choice of the density functional should not be only grounded in this criterion.

It is important to note that the pK_a values of the ionizable residues in folded proteins are influenced by the local environment. Therefore, the results presented in this work show only how well DFT describes the gas-phase proton affinities. The proton affinities in proteins will also depend on the description of

noncovalent interactions (among other factors), which are different for each functional.

Table 6. ΔE_{PT} for the 12 Density Functionals[a]

		MPW1B95						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
PBE1KCIS	Asp$^-$	0	0.81	1.33	−0.52	2.52	0.82	0.01
	Cys$^-$	1.16	0	0.52	−1.34	1.71	0.01	−0.81
	His	2.19	1.03	0	−1.85	1.20	−0.51	−1.32
	Ser$^-$	−1.08	−2.24	−3.27	0	3.05	1.34	0.53
	Arg	3.09	1.92	0.89	4.16	0	−1.70	−2.52
	Lys	1.54	0.38	−0.65	2.62	−1.55	0	−0.81
	Tyr$^-$	−0.33	−1.50	−2.53	0.74	−3.42	−1.88	0

		M062X						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
B1LYP	Asp$^-$	0	−1.62	−1.17	0.22	−0.34	−1.05	−0.31
	Cys$^-$	1.11	0	0.45	1.84	1.28	0.57	1.31
	His	2.44	1.33	0	1.39	0.83	0.12	0.86
	Ser$^-$	−0.75	−1.86	−3.19	0	−0.56	−1.27	−0.53
	Arg	3.05	1.94	0.61	3.80	0	−0.71	0.03
	Lys	1.51	0.40	−0.93	2.26	−1.53	0	0.74
	Tyr$^-$	0.12	−0.99	−2.32	0.87	−2.93	−1.39	0

		B1B95						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
PBE1PBE	Asp$^-$	0	1.00	1.37	−0.69	2.49	0.88	−0.10
	Cys$^-$	0.62	0	0.38	−1.69	1.49	−0.12	−1.10
	His	1.62	0.99	0	−2.06	1.11	−0.50	−1.47
	Ser$^-$	−0.86	−1.48	−2.48	0	3.18	1.57	0.59
	Arg	2.81	2.19	1.19	3.67	0	−1.61	−2.59
	Lys	1.22	0.59	−0.40	2.08	−1.59	0	−0.98
	Tyr$^-$	−0.58	−1.20	−2.19	0.28	−3.39	−1.79	0

		B97-1						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
B3P86	Asp$^-$	0	0.50	1.99	−1.20	2.67	1.57	−0.50
	Cys$^-$	1.18	0	1.50	−1.69	2.17	1.08	−0.99
	His	2.09	0.92	0	−3.19	0.68	−0.42	−2.49
	Ser$^-$	−0.99	−2.16	−3.08	0	3.87	2.77	0.70
	Arg	3.08	1.90	0.99	4.07	0	−1.10	−3.17
	Lys	1.46	0.28	−0.64	2.44	−1.62	0	−2.07
	Tyr$^-$	−0.35	−1.53	−2.44	0.64	−3.43	−1.81	0

		B3LYP						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
B98	Asp$^-$	0	1.43	2.81	−1.17	3.38	1.82	−0.01
	Cys$^-$	0.35	0	1.39	−2.60	1.96	0.40	−1.44
	His	2.03	1.68	0	−3.99	0.57	−0.99	−2.82
	Ser$^-$	−1.01	−1.36	−3.04	0	4.56	3.00	1.17
	Arg	2.73	2.38	0.70	3.74	0	−1.56	−3.39
	Lys	1.51	1.16	−0.52	2.52	−1.22	0	−1.83
	Tyr$^-$	−0.38	−0.73	−2.41	0.63	−3.11	−1.88	0

Table 6. Continued

		M06						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
M06-L	Asp$^-$	0	1.36	0.33	−1.53	1.74	−0.84	−0.88
	Cys$^-$	1.81	0	−1.03	−2.89	0.38	−2.20	−2.24
	His	0.13	−1.68	0	−1.86	1.41	−1.17	−1.21
	Ser$^-$	−2.33	−4.13	−2.46	0	3.27	0.69	0.65
	Arg	1.12	−0.69	0.99	3.45	0	−2.58	−2.62
	Lys	−0.58	−2.38	−0.71	1.75	−1.69	0	−0.04
	Tyr$^-$	−1.44	−3.25	−1.57	0.89	−2.56	−0.86	0

		MPW1B95-D3						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
B1B95-D3	Asp$^-$	0	0.75	1.29	−0.48	2.49	0.79	0.00
	Cys$^-$	0.80	0	0.55	−1.23	1.75	0.04	−0.74
	His	1.79	0.98	0	−1.78	1.20	−0.50	−1.29
	Ser$^-$	−0.60	−1.40	−2.38	0	2.98	1.27	0.49
	Arg	2.42	1.62	0.64	3.02	0	−1.71	−2.49
	Lys	0.79	−0.01	−1.00	1.38	−1.63	0	−0.79
	Tyr$^-$	−0.12	−0.92	−1.91	0.47	−2.54	−0.91	0

		BMK						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
BMK-D3	Asp$^-$	0	−0.94	1.78	0.82	2.55	0.97	−0.18
	Cys$^-$	−1.11	0	2.71	1.76	3.49	1.90	0.76
	His	1.70	2.81	0	−0.96	0.78	−0.81	−1.95
	Ser$^-$	0.91	2.01	−0.80	0	1.73	0.15	−1.00
	Arg	2.50	3.61	0.80	1.60	0	−1.59	−2.73
	Lys	0.90	2.01	−0.80	−0.01	−1.60	0	−1.14
	Tyr$^-$	−0.19	0.92	−1.89	−1.09	−2.69	−1.09	0

		CAM-B3LYP						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
CAM-B3LYP-D3	Asp$^-$	0	1.43	2.25	0.28	3.18	1.01	0.49
	Cys$^-$	1.35	0	0.83	−1.14	1.76	−0.42	−0.93
	His	2.20	0.85	0	−1.97	0.93	−1.25	−1.76
	Ser$^-$	0.34	−1.00	−1.85	0	2.90	0.73	0.21
	Arg	3.13	1.79	0.94	2.79	0	−2.18	−2.69
	Lys	0.95	−0.39	−1.24	0.61	−2.18	0	−0.51
	Tyr$^-$	0.49	−0.86	−1.71	0.14	−2.65	−0.47	0

		XYG3						
		Asp$^-$	Cys$^-$	His	Ser$^-$	Arg	Lys	Tyr$^-$
B3LYP-D	Asp$^-$	0	1.74	0.55	−0.24	1.19	1.33	−0.83
	Cys$^-$	1.37	0	−1.19	−1.98	−0.55	−0.41	−2.57
	His	2.95	1.58	0	−0.79	0.64	0.79	−1.37
	Ser$^-$	−1.05	−2.43	−4.01	0	1.43	1.57	−0.59
	Arg	4.08	2.71	1.13	5.14	0	0.14	−2.02
	Lys	2.12	0.75	−0.83	3.17	−1.96	0	−2.16
	Tyr$^-$	0.15	−1.22	−2.80	1.21	−3.93	−1.97	0

[a] All possible proton transfers between the seven side chains were included. All values in kcal/mol.

Table 7. MUE of the ΔE_{PT} Averaged over All Proton Transfers with a Given Density Functional (All Values in kcal/mol)

functional	MUE (ΔE_{PT})
M06-2X	0.82
MPW1B95-D3	1.13
XYG3	1.15
MPW1B95	1.20
B1B95	1.28
B1B95-D3	1.30
CAM-B3LYP-D3	1.33
CAM-B3LYP	1.37
BMK	1.46
M06	1.47
BMK-D3	1.48
PBE1PBE	1.58
B98	1.67
B1LYP	1.68
B97-1	1.73
M06-L	1.74
B3P86	1.77
PBE1KCIS	1.82
B3LYP	2.01
B3LYP-D	2.22

CONCLUSIONS

The results presented in this work are important for a conscientious characterization of the protonation states of amino acids and of proton transfers between different amino acids and will be helpful to further studies, e.g., on related acid/base catalytic mechanisms. A new, important and consistent DFT benchmarking database for PA_{el}^{0K} and for ΔPA_{el}^{0K} is provided. The PA_{el}^{0K} reference values were determined at the very accurate CCSD(T)/CBS level. These values were obtained by the sum of the complete basis set limit of the MP2 energies with a CCSD(T) correction term evaluated using the aug-cc-pVTZ basis set. The MP2/CBS energies were determined using both the Truhlar and Helgaker extrapolation schemes. The difference between the PA_{el}^{0K} values at the MP2/CBS level extrapolated with both schemes is very small (absolute average of 0.35 kcal/mol only). Analyzing the results of the DFT benchmarking, we conclude that the M06-2X, XYG3, MPW1B95, and B1B95 functionals are the more adequate to predict amino acid pK_as' and proton transfers between them. M06-2X is particularly attractive in this regard, as it has been shown to be accurate for intermolecular interactions (which will affect the pK_a in folded proteins) and themochemistry and kinetics (properties that are fundamental for most biological reactivity studies). It seems clearly to be one of the most attractive choices for studies of chemical reactivity in enzymes in which proton transfers play an important role.

ASSOCIATED CONTENT

Supporting Information. Absolute electronic energies calculated at the HF/aug-cc-pVXZ (X = 2−4), MP2/aug-cc-pVXZ (X = 2−4), and CCSD(T)/aug-cc-pVXZ (X = 2−3) theoretical levels are shown in Table S1. PA_{el}^{0K} calculated at the MP2/CBS level using the Truhlar extrapolation scheme as well as the separated HF and correlation energies are shown in Table S2. Mean Unsigned Errors (MUE) and Mean Signed Errors (MSE) for the absolute electronic energy values (total, HF, and Corr) of the difference between the MP2/CBS level using the Truhlar extrapolation scheme and the MP2/aug-cc-pVXZ (X = 2−4) basis sets, presented in Tables S3 and S4. Optimized geometries of all species are also presented. This information is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

*E-mail: mjramos@fc.up.pt.

Author Contributions

§These authors contributed equally to this work

ACKNOWLEDGMENT

N.F.B. and M.A.S.P. would like to thank the Fundação para a Ciência e a Tecnologia (FCT) for their Post-Doc and Ph.D. grants (SFRH/BPD/71000/2010 and SFRH/BD/43600/2008, respectively). The authors would like to acknowledge the Fundação para a Ciência e a Tecnologia for financial support (PTDC/QUI-QUI/102760/2008).

REFERENCES

(1) Jordan, I. K.; Kondrashov, F. A.; Adzhubei, I. A.; Wolf, Y. I.; Koonin, E. V.; Kondrashov, A. S.; Sunyaev, S. *Nature* **2005**, *433*, 633.
(2) Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. *J. Biol. Chem.* **2009**, *284*, 13285.
(3) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. *Protein Sci.* **2009**, *18*, 247.
(4) Harris, T. K.; Turner, G. J. *Iubmb Life* **2002**, *53*, 85.
(5) Harrison, A. G. *Mass Spectrom. Rev.* **1997**, *16*, 201.
(6) Matthew, J. B.; Gurd, F. R. N.; Garciamoreno, E. B.; Flanagan, M. A.; March, K. L.; Shire, S. J. *Crc Crit. Rev. Biochem.* **1985**, *18*, 91.
(7) Moser, A.; Range, K.; York, D. M. *J. Phys. Chem. B* **2010**, *114*, 13911.
(8) Range, K.; Lopez, C. S.; Moser, A.; York, D. M. *J. Phys. Chem. A* **2006**, *110*, 791.
(9) Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J. *J. Comput. Chem.* **2006**, *27*, 966.
(10) Ohno, K.; Kamiya, N.; Asakawa, N.; Inoue, Y.; Sakurai, M. *Chem. Phys. Lett.* **2001**, *341*, 387.
(11) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704.
(12) Mehler, E. L.; Guarnieri, F. *Biophys. J.* **1999**, *77*, 3.
(13) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 10478.
(14) Vessecchi, R.; Galembeck, S. E. *J. Phys. Chem. A* **2008**, *112*, 4060.
(15) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111*, 10439.
(16) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012.
(17) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
(18) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291.
(19) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2010**, *6*, 107.
(20) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 669.
(21) Miller, G. E.; Banerjee, N. C.; Stowe, C. M. *J. Pharmacol. Exp. Ther.* **1967**, *157*, 245–&.
(22) Butler, J.; Land, E. J.; Prutz, W. A.; Swallow, A. J. *J. Chem. Soc., Chem. Commun.* **1986**, 348.
(23) Ratilla, E. M. A.; Scott, B. K.; Moxness, M. S.; Kostic, N. M. *Inorg. Chem.* **1990**, *29*, 918.

(24) Schaefgen, J. R.; Newman, M. S.; Verhoek, F. H. *J. Am. Chem. Soc.* **1944**, *66*, 1847.
(25) Kristol, D. S.; Krautheim, P.; Stanley, S.; Parker, R. C. *Bioorg. Chem.* **1975**, *4*, 299.
(26) Bruice, T. C.; Schmir, G. L. *J. Am. Chem. Soc.* **1957**, *79*, 1663.
(27) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.
(28) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.
(29) Dunning, T. H.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244.
(30) Wilson, A. K.; Dunning, T. H. *J. Chem. Phys.* **1997**, *106*, 8718.
(31) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *294*, 45.
(32) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 6624.
(33) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *111*, 2921.
(34) Lee, J. S. *Phys. Rev. A* **2003**, 68.
(35) Schwartz, C. *Phys. Rev.* **1962**, *126*, 1015–.
(36) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639.
(37) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.
(38) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437.
(39) Jurecka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89.
(40) Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A. *J. Chem. Theory Comput.* **2010**, *6*, 2281.
(41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.
(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(43) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2010**, *6*, 395.
(44) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
(45) AK Grimme Homepage. http://www.uni-muenster.de/Chemie.oc/grimme/ (accessed October 2011).
(46) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.
(47) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 024101.
(48) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2006**, *124*, 174104.
(49) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.
(50) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.
(51) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.
(52) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.
(53) Zhang, I. Y.; Luo, Y.; Xu, X. *J. Chem. Phys.* **2010**, *132*, 104105.
(54) Zhang, Y.; Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4963.
(55) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(56) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.
(57) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
(58) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1997**, *274*, 242.
(59) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(60) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(61) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51.
(62) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264.
(63) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.
(64) Krieger, J. B.; Chen, J. Q.; Iafrate, G. J.; Savin, A. In *Electron Correlations and Materials Properties*; Gonis, A., Kioussis, N., Eds.; Plenum: New York, 1999; p 463.
(65) Schmider, H. L.; Becke, A. D. *J. Chem. Phys.* **1998**, *108*, 9624.
(66) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, 125.
(67) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(68) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.
(69) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
(70) Perdew, J. P. Unified Theory of Exchange and Correlation Beyond the Local Density Approximation. In *Electronic Structure of Solids '91*; Ziesche, P., Eschig, H., Eds.; Akademie Verlag: Berlin, 1991; p 11.
(71) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401/1−4.
(72) Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2701.
(73) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403.
(74) Hoe, W. M.; Cohen, A. J.; Handy, N. C. *Chem. Phys. Lett.* **2001**, *341*, 319.
(75) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.
(76) Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400.
(77) Gill, P. M. W. *Mol. Phys.* **1996**, *89*, 433.
(78) Boese, A. D.; Martin, J. M. L.; Handy, N. C. *J. Chem. Phys.* **2003**, *119*, 3005.
(79) Dahlke, E. E.; Truhlar, D. G. *J. Phys. Chem. B* **2005**, *109*, 15677.
(80) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715.
(81) Chai, J. D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
(82) Chai, J. D.; Head-Gordon, M. *J. Chem. Phys.* **2008**, 128.
(83) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
(84) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398.
(85) Kormos, B. L.; Cramer, C. J. *J. Phys. Org. Chem.* **2002**, *15*, 712.
(86) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.
(87) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.
(88) Slater, J. C. *The Self-Consistent Field for Molecules and Solids*; McGraw-Hill: New York, 1974; Vol. 4.
(89) Wei, Y.; Singer, T.; Mayr, H.; Sastry, G. N.; Zipse, H. *J. Comput. Chem.* **2008**, *29*, 291.
(90) Bras, N. F.; Goncalves, R.; Fernandes, P. A.; Mateus, N.; Ramos, M. J.; de Freitas, V. *Biochemistry* **2010**, *49*, 5097.
(91) Moon, A. F.; Edavettal, S. C.; Krahn, J. M.; Munoz, E. M.; Negishi, M.; Linhardt, R. J.; Liu, J.; Pedersen, L. C. *J. Biol. Chem.* **2004**, *279*, 45185.