JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

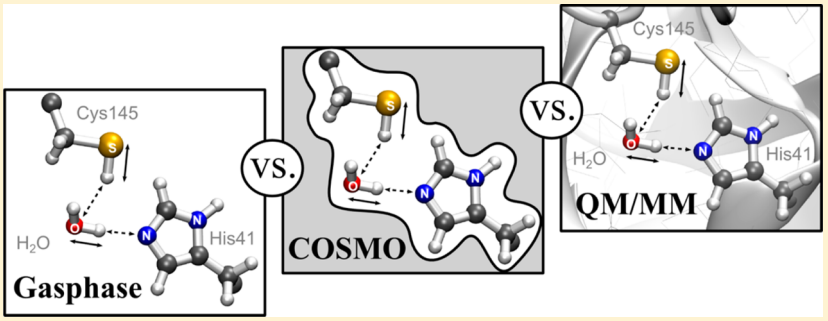
Benchmark Study for the Cysteine–Histidine Proton Transfer Reaction in a Protein Environment: Gas Phase, COSMO, QM/MM Approaches

Alexander Paasche,[†] Tanja Schirmeister,[‡] and Bernd Engels*[,†]

[†]Institut für Physikalische und Theoretische Chemie, Universität Würzburg, Emil-Fischer-Str. 42, 97074 Würzburg, Germany

[‡]Institut für Pharmazie und Biochemie, Johannes Gutenberg-Universität Mainz, Staudinger Weg 5, 55128 Mainz, Germany

Supporting Information



ABSTRACT: Proton transfer reactions are of crucial interest for the investigation of proteins. We have investigated the accuracy of commonly used quantum chemical methods for the description of proton transfer reactions in different environments (gas phase, COSMO, QM/MM) using the proton transfer between the catalytic dyad residues cysteine 145 and histidine 41 of SARS coronavirus main protease as a case study. The test includes thermodynamic, kinetic, and structural properties. The study comprises computationally demanding *ab initio* approaches (HF, CC2, MP2, SCS-CC2, SCS-MP2, CCSD(T)), popular density functional theories (BLYP, B3LYP, M06-2X), and semiempirical methods (MNDO/d, AM1, RM1, PM3, PM6). The approximated coupled cluster approach LCCSD(T) is taken as a reference method. We find that the robustness of the tested methods with respect to the environment correlates well with the level of theory. As an example HF, CC2, MP2, and their SCS variants show similar errors for gas phase, COSMO, or QM/MM computations. In contrast for semiempirical methods, the errors strongly diversify if one goes from gas phase to COSMO or QM/MM. Particular problems are observed for the recent semiempirical methods PM6 and RM1, which show the best performance for gas phase calculations but possess larger errors in conjunction with COSMO. Finally, a combination of SCS-MP2 and B3LYP or M06-2X allows reliable estimates about remaining errors.

INTRODUCTION

The determination of protonation states or proton transfers is relevant for nearly all theoretical studies dealing with enzymes and their actions in biological systems. The prediction of pK_a values for protein residues is one example for a routinely applied procedure. Moreover, proton transfers are involved in the crucial steps of peptide cleavage reactions or inhibition processes. Whenever these reactions are studied, the accurate description of the involved proton transfer reaction becomes important, which can be an extremely demanding task.

Whereas the determination of protein residue pK_a values can be efficiently tackled by using empirical pK_a prediction algorithms[1−3] or semimacroscopic protein dipole/Langevin dipole approaches,[4,5] theoretical studies of biochemical reactions mostly demand quantum chemical calculations. For whole enzymes that include thousands of atoms, the computational effort of calculations based on quantum mechanics is usually beyond the feasible limit. As a consequence, the system size treated with quantum chemical approaches needs to be restricted to the essential part of the investigated reaction while the environment has to be described on a more simple level of theory. Two very common approaches using this strategy are implicit solvent models[6−9] or combined quantum mechanical/molecular mechanical calculations (QM/MM).[10,11] For both concepts, recent reviews are available.[12−16]

Although the original intention of implicit solvent models is the mimicry of a solvent surrounding in an averaged manner, which makes it suitable for the prediction of pK_a values[17] or other thermodynamic properties,[18] their use has also been extended to biochemical reactions. Hereby, the character of the environment is mainly determined by the dielectric constant that can also be evaluated for a single case.[19] The spectrum of investigations reaches from the determination of protonation

Received: December 10, 2012

dx.doi.org/10.1021/ct301082y | *J. Chem. Theory Comput.* XXXX, XXX, XXX−XXX

states of catalytic dyads[20] to an investigation of hydrolysis reactions[21] as well as mechanistic studies of inhibition mechanisms.[22] Nowadays, the majority of studies that deal with proton transfers in a biochemical context employ QM/MM methods and rely on density functional theory in most cases.

A recent example for QM/MM investigation of this kind is the water mediated proton transfer between histidine and a hydrogen peroxide, which plays an important role in the catalytic cycle of the horseradish peroxidase.[23] Another reaction, where proton transfer occurs in a relayed fashion, is studied by comparing direct and hydroxyl group mediated hydrogen atom transfer in the peptide bond formation of ribosomes.[24] Besides enzyme specific studies, the fundamental understanding of neutral and zwitterionic resting states of cysteine/histidine catalytic dyads has been a frequently attended topic during recent years. Insights into this field have been provided by the work of Mladenovic et al. from the cysteine protease cathepsin B,[25] Kaukonen et al. from [Fe,Ni] hydrogenase,[26] or Ke et al., who investigated the catalytic mechanism of an arginine deiminase.[27,28] The KasA enzyme, a drug target against tuberculosis, gives further interesting insights since it contains one cysteine and two histidine residues that have been studied by Lee et al.[29] Nevertheless, such kinds of studies are not only conducted for cysteine/histidine residues but also for other types of catalytic protein sites, like those demonstrated for the Pili protein, which employs asparagin, lysin, and glutamate as catalytic residues.[30] Furthermore, proton transfer reactions are often involved in the inhibition process of an enzyme. Here, knowledge of the underlying mechanisms provides important hints for the rational drug design of inhibitor molecules. An example where a proton transfer is directly involved in an inhibition reaction is given by Cheng and co-workers,[31] who study the covalent inhibition of a serine protease that possesses a catalytic triad. Like most of the studies mentioned above, they use the often applied density functional B3LYP within a QM/MM approach. With an analogue methodology, Mladenovic et al. give insight to the importance of proton shifts for the inhibition of cysteine proteases by aziridines and epoxides as irreversible warheads.[32]

Recent work of Xie et al.[33] demonstrates a complementary use of an implicit solvation model on the CCSD(T) level of theory and QM/MM computations employing density functional theory to gain insight into the formation of carbamates from carbon dioxide. Whereas this study uses the two concepts of solvent models separately, Ryde and co-workers[34] apply a combination of explicit and implicit treatment of the environment in terms of the QM/MM-PBSA method, which is suited to estimate free energy differences of proton transfer reactions in a promising manner. Further interesting applications, where proton transfer reactions are involved, are the theoretical interpretation of kinetic isotope effects within a protein by Olsson et al.[35] or the investigation of the intramolecular proton transfer within malonaldehyde,[36] which both state examples for the calculation of nuclear quantum effects.

As mentioned above, the major part of studies employ density functional theory due to its good ratio between accuracy and computational cost, but there are also many examples where higher or lower levels of theory are used. Especially in the case of simulations with *ab initio* methods, the thereby associated computational costs require the use of computationally inexpensive methods,[37] like semiempirical Hamiltonians.[25,38] Although a lot of benchmarking studies are available throughout the literature that probe the accuracy of quantum chemical methods in the context of proton transfer reactions,[39−44] most focus on accuracy under gas phase conditions. There are only a few examples in the literature[45,46] that provide a comparison throughout different descriptions of the environment, like the work of Sharma et al.,[47] who assess the accuracy of different quantum chemistry models in the context of proton transfer tunneling and kinetic isotope effects.

Since implicit solvent models or QM/MM methods can have a large impact on the calculated properties, it is questionable how, or to what extent, benchmarks made under gas phase conditions are valid for these approaches.

To the best of our knowledge, there is actually no study available that systematically investigates the behavior of commonly used quantum chemical methods with respect to different modeling of the environment, like implicit solvation or the QM/MM approach. Therefore, we have designed this study to probe the "robustness" of quantum chemistry methods with regard to this aspect and to reveal a possible bias of parametrized methods toward gas phase models.

Hereby, we focus more on the direct comparison of the quantum chemical Hamiltonians rather than on methodologies for calculating pK_a values or free energies. For the latter topics, the reader is referred to the literature. A good overview about calculating pK_a values in proteins is given in the review of Gunner and co-workers,[48] whereas an overview about free energy methods has been recently addressed by two recommendable reviews that cover this field in a more general way[37] or with emphasis on the QM/MM approach.[49] Although we do not cover free energy calculations themselves, it should be emphasized that the reliability of methods, which derive Gibbs or Helmholtz free energies from QM or QM/MM potentials, strongly depends on the accuracy of the chosen quantum mechanical Hamiltonian. Therefore, the choice of the quantum chemical method can have a significant impact on the free energy results, which makes a critical assessment of its reliability necessary.

Our results should also help to reveal whether certain methods possibly possess a systematic error behavior with respect to the character of the environment. This would give a perspective toward the correcting of such kinds of errors by empirical factors.

For the selection of the probed methods, we have considered three important facts. The first one is how commonly used a method is. Therefore, we have included popular approaches like the Møller−Plesset perturbation theory MP2, density functional theory using B3LYP, or the classical semiempirical Hamiltonians MNDO, PM3, and AM1. As a second important issue, we wanted to take account of more recent developments like improvements due to spin component scaling techniques (SCS), meta-hybrid exchange-correlation functionals (e.g., M06-2X), or higher sophisticated parametrizations as found in the PM6 or RM1 method. The third and most decisive fact is whether a method is implemented for use in conjunction with an implicit solvent model or the QM/MM approach. Fortunately, this holds for the majority of methods that are applied for proton transfer reactions.

We have further included the approximated coupled cluster approach CC2, which is indeed rarely used for the modeling of proton transfer potentials but well suited for the calculation of excited states and their potential energy surfaces. For problems of this kind, often multireference methods are necessary,

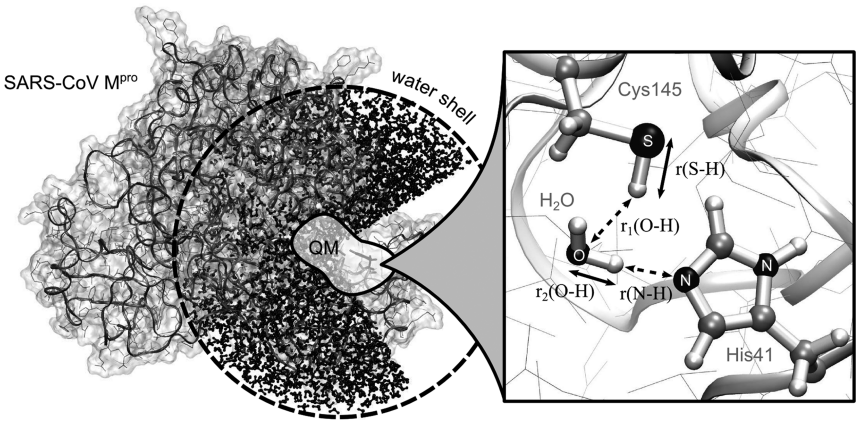


Figure 1. QM/MM scheme for the active site of SARS coronavirus main protease as a case study (left-hand side). Cysteine 145, histidine 41, and a bridging water molecule were taken as the QM part (right-hand side). Proton transfer reaction occurs along the coordinates r(S−H), r_1(OH), r_2(OH), and r(N−H). The r(S−H) coordinate (solid arrow) is used as the "driving" reaction coordinate; all other coordinates (dashed arrows) follow smoothly in forward and backward directions during the proton transfer reaction.

although DFT approaches can also deliver satisfying accuracies in some cases.[50−53] Thus, our results should give an estimate, which impact can be expected for the CC2 method when other environments than the gas phase are applied.

BENCHMARKING SETUP

To evaluate the robustness of different quantum chemical methods with respect to environment modeling, we compare gas phase calculations with an implicit solvent model and an explicit description of the environment in terms of a QM/MM approach.

Since we aim toward a consistent comparison of a broad spectrum of Hamiltonians, the major bottleneck of this study is the availability of all methods within the different environment models in the used program packages. To provide a complete and consistent comparison of all methods and schemes, it was necessary to employ six program packages. Especially the use of QM/MM methodology limits the choice of Hamiltonians, since an electrostatic embedding scheme, which is the most popular embedding scheme for biomolecular applications,[12] is not possible with all quantum chemical programs. Due to its flexible design and the possibility to interface multiple quantum chemistry codes, we employed the ChemShell program package.[54]

As an implicit solvent model, we have chosen the COSMO approach,[55] which is implemented in most program packages. An important parameter for the use of continuum models is the selection of the dielectric constant. We have considered two values, a low one of $\varepsilon = 4$ that represents an environment of low polarity and refers to a protein bulk,[56,57] whereas the second one of $\varepsilon = 78$ is usually chosen for very polar environments, comparable to water. The proton transfer reaction between the catalytic residues cysteine 145 and histidine 41 at the active site of the viral cysteine protease SARS coronavirus main protease serves as a test case, for which all methods are evaluated. This proton transfer represents the transition from the neutral resting state, in the following abbreviated with **N**, to the zwitterionic resting state, abbreviated with **ZW**, of the catalytic dyad. We further denote the transition state between the two minimum states as **TS**.

Experimental studies on the active-site residues of SARS-CoV main protease estimate pK_a values for cysteine 145 and histidine 41 to 7.7−8.3 and 6.2−6.4, thus finding the neutral state **N** as the predominant situation for the free enzyme.[58,59] The pK_a differences between the two groups can be employed to estimate the free solvation energy differences according to eq 1, which is adopted from the work of Warshel et al.[60,5,61] and discussed there in detail. For the given case, the free energy difference between **N** and **ZW** can expected to be in the range of 7−12 kJ/mol under standard conditions for temperature T.

$$\Delta G_{\mathrm{sol}}^{\mathrm{N}\rightarrow\mathrm{ZW}} \cong 2.3RT(\mathrm{p}K_{\mathrm{a}}^{\mathrm{Cys}} - \mathrm{p}K_{\mathrm{a}}^{\mathrm{His}}) \quad (1)$$

Geometric parameters from X-ray crystallography,[62−64] as well as from molecular modeling studies,[65,66] further reveal that proton transfer is most likely to occur in a water-mediated fashion. The main indicative factor for this finding is the measured distance between the cysteine sulfur and the histidine nitrogen. It is found to be in a range between 3.5 and 4 Å[62−64] and is therefore too large for a direct proton transfer that would require a direct hydrogen bond with a sulfur−nitrogen distance of about 3 Å. This is also consistent with our own findings from 10 ns MD simulations. The setup of the model system and the relevant reaction coordinates are shown in Figure 1.

Three properties have been evaluated for the proton transfer path: (1) thermodynamic and kinetic data in terms of relative energies for minima and transition states, (2) mean unsigned errors of the whole proton transfer potential, and (3) the location of minima and transition states along the reaction coordinate.

The assessment of accuracy for all methods used in this study is done relative to the LCCSD(T)|QZVP level of theory for gas phase and QM/MM computations. Due to its advantageous scaling behavior, accompanied by an insignificant loss of accuracy, density fitting approximation (also known as resolution of identity approximation) has been employed.[67,68] Whereas canonical CCSD(T) approaches are expected to have chemical accuracy for at least quadruple-ζ basis sets (mean error below 1 kcal/mol) with respect to experimental reaction energies,[69] local approximations can introduce an additional average error of 0.6 kcal/mol.[70]

For computations combining LCCSD(T) and COSMO, no code was available, so the SCS-CC2|QZVP level of theory is taken as a reference method, since it yields the least errors in comparison to LCCSD(T) results for our test case. Nevertheless, its performance is quite similar to SCS-MP2, which

dx.doi.org/10.1021/ct301082y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

should possess a mean absolute error below 1.8 kcal/mol for reaction energies in the gas phase.[71,72]

We have used two basis sets for each method, one of triple-ζ quality and one of quadruple-ζ quality, in order to roughly estimate the magnitude of the basis set effect for each method. It is notable that the description of anionic species can be further improved by introducing augmented basis sets, but since we experienced quite often convergence problems with these, their use has been omitted here.

The lowest minimum of the studied proton transfer path is given by cysteine 145 and histidine 41 in their neutral state **N**, which is taken as a reference state. Therefore, energy differences ΔU of transition states (**TS**) and zwitterionic states (**ZW**) are always taken relative to the respective neutral minimum state **N** on each potential energy curve. A detailed description of the evaluated energies is given in Figure 2.

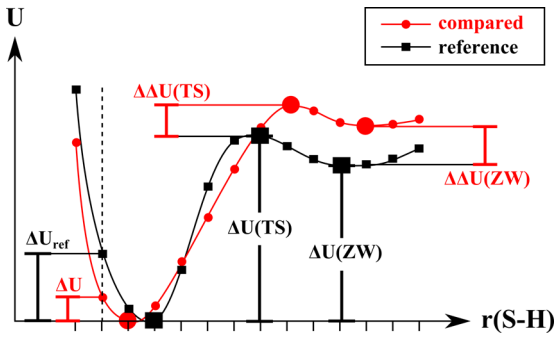


Figure 2. Explanation of the relative internal energies ΔU and their differences $\Delta\Delta U$, which are calculated from the respective grid points (solid horizontal lines) along the reaction coordinate r(S−H).

Errors for the relative energies of the minimum state of the zwitterion $\Delta\Delta U$(ZW) and the transition state $\Delta\Delta U$(TS) are calculated as differences between the obtained energy difference ΔU and the reference energy difference ΔU_{ref} of the respective single point structure. The groups of wave function based methods, density functional theory and semiempirical methods are discussed in an analogous manner: First, unsigned errors and their behavior with respect to the environment model are discussed. Second, tendencies of over- or underestimation are considered. Third, interesting trends are highlighted.

Calculations of mean unsigned errors (MUE) are done over the sum of all 14 data points of the proton transfer path according to eq 2. Energy differences ΔU of all points are taken relative to the neutral minimum state **N** on the respective potential energy curve, as described above.

$$\mathrm{MUE} = \frac{1}{N}\sum_{1}^{N} |\Delta U - \Delta U_{\mathrm{ref}}| \qquad (2)$$

A similar approach has been applied for deviations of minima and transition state locations along the reaction coordinate as shown in Figure 3. Errors for the distances of neutral states Δr(N), zwitterionic states Δr(ZW), and transition states Δr(TS) along the reaction coordinate r(S−H) are calculated as differences between the obtained distance r(N,TS,ZW) of the respective minimum/maximum point of the probed method and the reference distance value r_{ref} (N,TS,ZW).

The minima and maxima points along the reaction coordinate have been approximated by fits of harmonic functions (polynomial of second degree) that include two

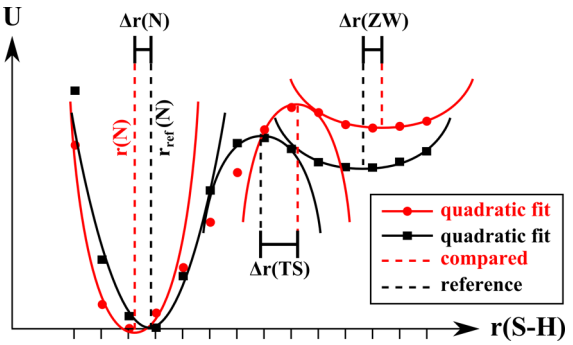


Figure 3. Quadratic fits of extremal values N, TS, and ZW. Fitted functions of the reference (black) and compared potentials (red) are depicted as solid lines. Deviations Δr (dashed vertical lines) are taken relative to fitted extremal values of the reference potential.

neighboring points in each direction around the closest incremental minimum point on the potential curve. The fits are schematically drawn as dashed lines in Figure 3. The location of the minimum or maximum point is hence obtained by determination of the extreme value of the fitted function. For **TS**, only one neighboring point in each direction has been used around the maximum value, since its location on the reaction coordinate is often very close to **ZW**. This fitting procedure should give more accurate estimates of the stationary points than simply taking the highest or lowest incremental value on the potential energy curve.

COMPUTATIONAL DETAILS

In order to get a selection of starting structures, molecular dynamic simulations were conducted with the NAMD 2.7 program package,[73,74] based on the crystal structure of SARS-CoV M^{pro} (PDB code 2DUC)[75] taken from RCSB protein data bank.[76] The initial protein structure was embedded in the center of a TIP3P[77] water shell with 110 Å diameter and energy minimized by keeping the protein structure fixed. Protonation states of titratable residues were determined by empirical pK_a prediction with the PROPKA methodology,[78−80] satisifying a pH value of 7. Since no residual charges were left in the whole system, it was not necessary to place counterions. To ensure a well equilibrated model system, it was slowly heated to 310 K in gradual steps of 10 K. After heating of the system, atom position constraints put on the protein structure were successively released, and the simulation was equilibrated for 1 ns, followed by a production run of 10 ns. Heating, equilibration, and the simulation procedure were conducted in an NVT ensemble with 1 fs time steps using the SHAKE algorithm[81] for all water molecules and the Verlet algorithm as an integration method.[82] Spherical harmonic boundary conditions were applied beyond a radius of 55 Å around the center of the sphere to prevent water molecules from evaporating. The root-mean-square deviation (RMSD) from the starting structure of the protein backbone plotted as a function of time (Figure 4) indicates a stable simulation system with an average RMSD of 1.6 Å.

All QM/MM calculations were performed within the ChemShell 3.4 computational chemistry suite,[54] using an electrostatic embedding scheme in conjunction with the charge shift method and link atom approach for the QM/MM boundary treatment.[11,83,84] The QM/MM model system was set up from the protein structure of the SARS coronavirus main protease (SARS-CoV M^{pro}). As a starting structure for all

dx.doi.org/10.1021/ct301082y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

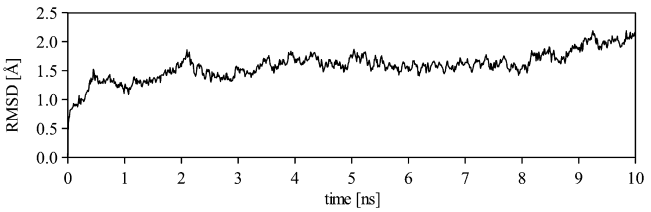


Figure 4. Root mean square deviation (RMSD) from the starting structure as a function of time for the molecular dynamic simulation of SARS coronavirus main protease.

calculations, we used one suitable snapshot from the previously performed molecular dynamics simulations (MD). The active site residues Cys145 and His41 and one bridging water molecule were selected as the quantum mechanical part (QM) of the model system, as depicted in Figure 1. All remaining atoms were treated on the molecular mechanic level of theory (MM). For the MM part, the all-atom force field CHARMM22 was used.[85]

To save computational resources in QM/MM calculations, the water shell was reduced to a radius of 30 Å around the active site residues His41 and Cys145. The outer water layer and protein structure at radii greater than 10 Å were kept fixed during structure optimizations. The minimum proton transfer path was obtained from a relaxed constrained minimization series along the reaction coordinate as shown in Figure 1, using increments of 0.1 Å for the r(S−H) distance. All of the other relevant reaction coordinates, r_1(O−H), $r2$(O−H), and r(N−H), adjusted smoothly to the stepwise changed r(S−H) distance during the constrained minimizations. The minimization series proceeded in both directions to ensure that the obtained minimum path comprises a meaningful connection between the neutral state and zwitterionic state. QM/MM structure minimizations were performed at the BLYP|TZVP level,[86,87] using a resolution of identity approximation.[88,89]

For the benchmarking, single point calculations of the obtained minimum path were performed for all methods. In detail, this included 14 grid points of the proton transfer potential which connect the neutral state **N** and the zwitterionic state **ZW**. It is noteworthy that instead of calculating single point energies, there is also the possibility to perform geometry optimizations on each level of theory, which might deliver slightly different results. Due to the fact that gradients are not available for every method/environment combination and further that higher level single point calculations "on top" of DFT geometries are a rather established and routinely applied approach for QM/MM computations, we have chosen the first possibility. Nevertheless, since it cannot be assured that GGA functionals like BLYP deliver reasonable structures for proton transfer reactions, as recently demonstrated by Adamo and coworkers,[90] we have furthermore performed structure optimizations on the MP2|TZVP and B3LYP |TZVP levels as well as with the semiempirical methods AM1 and PM3.

As an implicit solvent model, the COSMO approach of Klamt and Schüürmann[55] was used. Since default settings partly differ in MOLPRO, GAMESS-US, MOPAC, and TURBOMOLE, they have been set manually to keep results comparable between the different program packages. The values for the dielectric constant have been set to ε = 4 and ε = 78, which represent estimates of the protein environment[56,57] and water.[91] The multiplicative factor for cavity construction (RSOLV) was set to a radius of 1.2 times the respective van der Waals radius as recommended by Tomasi and coworkers.[92,93] Furthermore, the number of surface elements per atomic sphere was set to 92 and the distance threshold radius for segment−segment interactions to 10 Å in all COSMO calculations, which represent the most commonly used default values. Nevertheless, since the choice of the multiplicative factor might severely change the resulting proton transfer potentials, we have probed the influence of RSOLV on the relative transition state energies and zwitterionic minimum state energies at least for MP2|TZVP, M06-2X|TZVP, and PM3 by using a dielectric constant ε of 78. Changing RSOLV in the range from 0.5 to 3.0 influences the relative MP2 energies by a maximum deviation of 1 kJ/mol (**TS**) and 2 kJ/mol (**ZW**), relative M06-2X energies by 3 kJ/mol (**TS**) and 6 kJ/mol (**ZW**), and relative energies predicted by PM3 in the range of 2 kJ/mol (**TS**) and 6 kJ/mol (**ZW**). The respective energy plots are included in the Supporting Information.

The QM/MM calculations always included the whole system as described above, whereas COSMO and gas phase calculations were carried out only with the QM part. Reference calculations with the local coupled cluster method LCCSD-(T)[94] were conducted with the MOLPRO 2010 quantum chemistry package,[95] using a density fitting approximation.[68,96] Nonlocal CCSD(T) calculations, as well as Møller−Plesset perturbation theory MP2,[97] the approximated coupled cluster theory CC2,[98] Hartree−Fock (HF) calculations,[99] and density functional theory computations with the BLYP[86] and B3LYP[100] functional, were performed with the TURBOMOLE 6.2 program package.[101] Since spin component scaling (SCS) improves the accuracy of MP2 and CC2 calculations without causing further computational costs,[71,102] it has been applied for the reference calculation on the MP2 level and the CC2 calculation.

To estimate effects due to the extent of basis functions, polarized basis sets of single valence (SVP),[87] triple-ζ (TZVP),[103] and quadruple-ζ quality (QZVP)[104] were employed as implemented in TURBOMOLE and MOLPRO. The resolution of identity approximation[89] was used for MP2, CC2, and B-LYP calculations in conjunction with suitable auxiliary basis sets.[88] DFT calculations employing the meta-hybrid exchange-correlation functional M06-2X[105] were carried out with the GAMESS-US program package version OCT 2010 (R1).[106] Here, TZVP and QZVP basis set files were manually converted from TURBOMOLE, since these are not included in GAMESS-US. A cross check with B3LYP results between TURBOMOLE and GAMESS-US results confirmed a faultless conversion. QM/MM calculations with M06-2X could not be performed, since an electrostatic embedding scheme is not available for the GAMESS-US interface of ChemShell.

Gas phase and COSMO calculations that employ the semiempirical methods AM1,[107] RM1,[108] PM3,[109] PM6,[110] MNDO,[111] and MNDO including d orbitals[112,113] were conducted with the MOPAC2009 program.[114] For QM/MM calculations with semiempirical methods, the MNDO2005 program was used.[115] Since RM1 and PM6 are not implemented in MNDO2005, they had to be omitted.

RESULTS AND DISCUSSION

Impact of the Environment Model on the Proton Transfer Reaction. The proton transfer between cysteine and histidine connects the neutral state **N** with the zwitterionic state **ZW**. The kinetics of the reaction is determined by the height of transition state **TS**. The proton transfer potentials in Figure 5

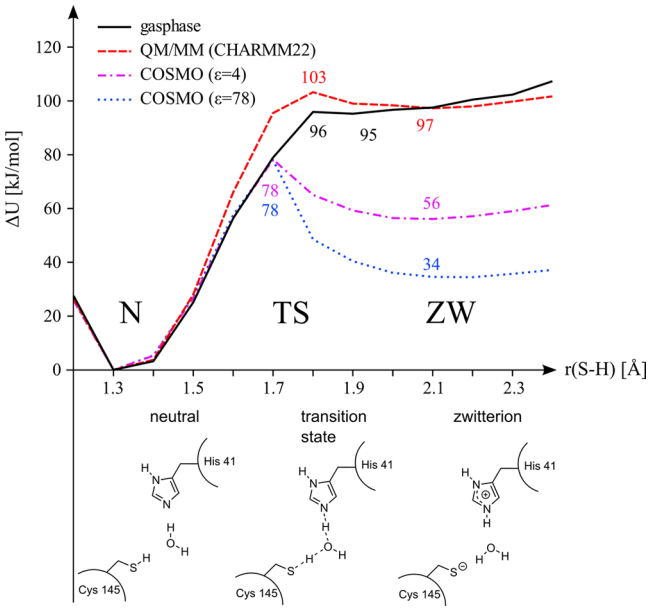


Figure 5. Comparison of proton transfer potential in the gas phase, continuum solvent model (COSMO), and QM/MM scheme. Gas phase and QM/MM potentials have been obtained on the LCCSD(T)|QZVP level of theory, COSMO potentials employing SCS-CC2|QZVP. Energies of transition states **TS** and zwitterionic states **ZW** are taken relative to the neutral state **N** of each potential curve in kJ/mol. The structures of the minimum energy path have been optimized within the QM/MM scheme on the RI-BLYP|TZVP level of theory.

compare the gas phase results with those obtained with the implicit environment model COSMO using two different dielectric constants and those computed with the QM/MM approach. The potentials are computed on the LCCSD(T)|QZVP (gas phase, QM/MM) and SCS-CC2|QZVP levels of theory (COSMO). As apparent from Figure 5, all environment models reproduce the experimental finding that **N** is more stable than **ZW**. However, the predictions show significant quantitative differences.

Proton transfer potential in the gas phase and QM/MM are in a similar range, predicting an energy difference of 95 and 97 kJ/mol for **ZW** over **N** by using LCCSD(T). A more distinctive difference can be observed for the barrier heights **TS**, where the QM/MM model shows a slightly higher value of 103 kJ/mol for **TS** with respect to 96 kJ/mol in the gas phase. The use of the COSMO model has a large impact on the proton transfer potential in terms of stabilizing **ZW**. Even with a low dielectric constant of $\varepsilon = 4$, which is thought to reflect the situation in a protein environment, the zwitterionic state is lowered by 41 kJ/mol with respect to the QM/MM description and moreover by 63 kJ/mol if a higher dielectric constant of $\varepsilon = 78$, comparable to water, is applied.

This result is counterintuitive on first sight, since the catalytic residues are located at the surface of the protein. Thus the dielectric constant of the surrounding should therefore be somewhere between 4 and 78, but even the COSMO description of a low $\varepsilon = 4$ delivers a significantly larger stabilization of **ZW** than the QM/MM model. On closer inspection, this difference is not surprising, since the implicit solvent model mimics an ideal and perfect solvent screening for the reaction in contrast to the QM/MM description, which models the environment in an explicit way, thus taking molecular effects into account that might arise at the water−protein interface. This is further underlined by the comparison to the experimentally derived free energy difference between 7 and 12 kJ/mol (eq 1), which is much better in agreement with the COSMO results. In summary, the qualitative result of having **N** as the lower state is correctly predicted by all proton transfer potentials in Figure 5, but only due to the fact that both states **N** and **ZW** are sufficiently distinctive in terms of their relative energies. In cases where the two states are very close together, a proper relaxation of the environment and further free energy calculations become important, as demonstrated by the work of Lee et al.,[29] who investigated similar proton transfers for another enzyme.

Thermodynamic and Kinetic Properties. Table 1 probes the accuracy of the wave function based *ab initio* methods, the density functional theory methods, and the semiempirical methods with respect to gas phase calculations, the COSMO approach, and the QM/MM scheme by a comparison of the relative energies ΔU(ZW) between the neutral state **N** and the zwitterionic state **ZW**, as well as the transition state energy ΔU(TS). The deviation of each method with respect to this reference is given in Table 2 and is obtained according to the energy differences defined in Figure 2. In the following, we only consider deviations >8 kJ/mol as significant. This error bar takes into account remaining uncertainties of non-gas-phase computations and is near the expected accuracy limit of the reference methods with respect to experimental gas phase values (1.6 kcal/mol for LCCSD(T)|QZVP[69,70] and 1.8 kcal/mol SCS-CC2|QZVP[71,72]), as discussed in the benchmarking setup paragraph.

Since the performance of a certain method to predict relative energies must not necessarily be reflected in the quality of structure predictions,[90] the BLYP|TZVP geometries were probed against B3LYP|TZVP and MP2|TZVP structure optimizations, as well as optimizations on the AM1 and PM3 semiempirical level. The comparison in Table 1 of MP2|TZVP (indicated by an asterisk) and MP2|TZVP//BLYP|TZVP and furthermore B3LYP|TZVP (indicated by an asterisk) and B3LYP|TZVP//BLYP|TZVP reveals that the proton transfer potentials are quite similar and differ in or below the range of significance for our test case. Interestingly, structure optimizations with AM1 and PM3 were not able to reproduce the minimum energy path properly and delivered quite different structures. Detailed plots of these comparisons can be found in the Supporting Information.

The inspection of the wave function based group of *ab initio* methods reveals a very consistent error behavior across the gas phase, COSMO, and the QM/MM approach (Table 2). In detail, the unsigned errors of the relative energy $\Delta\Delta$U(ZW) for the *post-HF* methods (CCSD(T), LCCSD(T), SCS-CC2, SCS-MC2, MP2, CC2) range in general below an absolute value of 14 kJ/mol and change only insignificantly when the environment model is switched (<5 kJ/mol). The errors of the relative transition state energy $\Delta\Delta U$(TS) behave similarly to this, since the largest unsigned deviation is 15 kJ/mol in case of the CC2|QZVP method within the QM/MM scheme. The error behavior with respect to the environment model is also in line with the $\Delta\Delta U$(ZW) values and range at the limit of significance with a maximum absolute difference of 9 kJ/mol (SCS-CC2|TZVP with COSMO versus QM/MM). The higher level methods CCSD(T), LCCSD(T), SCS-CC2, and SCS-MP2 reveal a consistent tendency of 12−15 kJ/mol overestimation of the energy difference between **N** and **ZW**, when the TZVP basis sets are used. By comparing the same basis set

dx.doi.org/10.1021/ct301082y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Evaluation of Thermodynamic and Kinetic Data from Different Quantum Chemical Methods in kJ/mol[a]

		gas phase		COSMO				QM/MM	
				$\varepsilon = 4$		$\varepsilon = 78$			
level of theory	basis set	ΔU(TS)	ΔU(ZW)	ΔU(TS)	ΔU(ZW)	ΔU(TS)	ΔU(ZW)	ΔU(TS)	ΔU(ZW)
LCCSD(T)[b]	QZVP	96	95	-[c]	-[c]	-[c]	-[c]	103	97
SCS-CC2[b]	QZVP	96	94	78	56	78	34	102	95
SCS-MP2[b]	QZVP	99	96	81	57	81	35	104	97
CC2[b]	QZVP	83[d]	83[d]	68	47	68	27	88	85
MP2[b]	QZVP	86[d]	86[d]	71	49	71	28	92	88
HF	QZVP	138	112	124	67	124	43	145	108
M06-2X	QZVP	86	84	71	37	76	6	-[c]	-[c]
B3LTP	QZVP	87	85	67	49	67	29	94	87
BLYP	QZVP	80	79	55	47	54	27	87	83
CCSD(T)[b]	TZVP	108	106	82	67	82	46	116	108
LCCSD(T)[b]	TZVP	108	105	-[c]	-[c]	-[c]	-[c]	115	106
SCS-CC2[b]	TZVP	107	106	81	69	81	47	115	110
SCS-MP2[b]	TZVP	111	109	85	70	85	49	118	112
CC2[b]	TZVP	95[d]	96[d]	73	61	72	41	103	101
MP2[b]	TZVP	100[d]	100[d]	77	63	77	43	107	104
MP2[b,e]	TZVP							96	94
HF	TZVP	131	106	16	62	117	39	140	104
M06-2X	TZVP	79	78	64	39	69	17	-[c]	-[c]
B3LYP	TZVP	81	80	60	44	59	25	87	80
B3LYP[e]	TZVP							85	79
BLYP[e]	TZVP	75[d]	75[d]	48	43	47	24	81	77
PM6		63[d]	69[d]	38	16	38	−14	-[c]	-[c]
PM3		141	129	91	58	66	22	141	119
RM1		64[d]	68[d]	57	9	56	−25	-[c]	-[c]
AM1		159	137	111	67	96	33	160	126
MNDO/d		324	257	273	185	247	149	318	239
MNDO		325	281	275	210	251	176	327	271

[a]Predicted values ΔU(TS) and ΔU(ZW) are taken relative to neutral state N on the respective level of theory. A detailed description of the relative energies is given in Figure 2. [b]Resolution of identity approximation or density fitting employed. [c]Method not available. [d]No local minimum or transition state found. Value represents the relative energy of the respective grid point. [e]Structure optimization on the respective level of theory.

qualities with each other, MP2 and CC2 show very similar performance and differ in their results only marginally (3−5 kJ/mol). Accuracy might be improved by spin component scaling but not consistently. Interestingly, spin component scaled methods tend to deliver better results with QZVP basis sets ($\Delta\Delta U$ < 3 kJ/mol vs $\Delta\Delta U$ < 15 kJ/mol for TZVP), whereas the unscaled CC2 and MP2 counterparts seem to give slightly but not significantly better results in conjunction with the smaller TZVP basis set ($\Delta\Delta U$ < 9 kJ/mol vs $\Delta\Delta U$ < 15 kJ/mol for TZVP). On the other hand, gas phase calculations employing MP2 and CC2 without spin component scaling do incorrectly reproduce the shape of the potential curve, since they predict a complete repulsive potential, without a minimum for **ZW** and the transition state.

Nevertheless, basis set effects remain significant and systematically as indicated by the comparison of TZVP and QZVP for each of the wave function based *ab initio* methods. Thus, applying a bigger basis set results in lowering the zwitterionic state by about 12−16 kJ/mol. One exception from this is the Hartree−Fock method, where a bigger basis set leads to an insignificantly higher transition state energy and minimum in the range of 4−7 kJ/mol. This special case can also be found in the comprehensive quantum chemistry textbook of Helgaker et al. who calibrate electronic structure methods systematically.[69] Anyway, the use of a "pure" HF method cannot be recommended as apparent from the calculated errors, which rise up to 46 kJ/mol for the predicted transition state energy $\Delta\Delta U$(TS). As a common trend of all wave function based methods, overestimation of the relative energy between **N** and **ZW** can be observed for the use of the TZVP basis set.

The comparison of thermodynamic and kinetic data predicted by density functional theory based methods reveals the often appreciated cost/benefit ratio provided by DFT methods. We have selected only a few exchange correlation functionals, since extensive benchmarks are available elsewhere,[72,116,117] in particular with a focus on proton affinities in amino acids.[44] Our selection is done according to the sophistication level on the "Jacobs Ladder" of density functional theory[118] and includes three of the most commonly used functionals, BLYP based on generalized gradient approximation (GGA), the popular hybrid-GGA functional B3LYP, and the more recent M06-2X functional as a meta-hybrid-GGA type. The use of empirical dispersion correction terms, like that from Grimme et al.,[119,120] is not necessary for this type of reaction, since proton transfers are primarily of an electrostatic nature and therefore marginally affected by errors due to dispersion. However, we have checked the impact of dispersion correction on our case study by B3LYP-D3 test calculations in the gas phase, delivering differences below 1 kJ/mol for the results with respect to the uncorrected B3LYP functional. Considering the robustness of DFT methods with respect to the environment model, it becomes apparent that they do not behave as consistently as the wave function based methods, but still in an

dx.doi.org/10.1021/ct301082y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Errors for Relative Transition State Energies ΔΔ*U*(TS), Relative Minimum Energies ΔΔ*U*(ZW), and Mean Unsigned Errors MUE for the Whole Set of 14 Single Point Energies along the Minimum Energy Path in kJ/mol[a]

		gas phase			COSMO						QM/MM		
level of theory	basis set				ε = 4			ε = 78					
		ΔΔ*U*(TS)	ΔΔ*U*(ZW)	MUE	ΔΔ*U*(TS)	ΔΔ*U*(ZW)	MUE	ΔΔ*U*(TS)	ΔΔ*U*(ZW)	MUE	ΔΔ*U*(TS)	ΔΔ*U*(TS)	MUE
LCCSD(T)[b]	QZVP	0	0	**0**	-[c]	-[c]	-[c]	-[c]	-[c]	-[c]	0	0	**0**
SCS-CC2[b]	QZVP	0	−1	**1**	0	0	**0**	0	0	**0**	−1	−2	**2**
SCS-MP2[b]	QZVP	3	1	**1**	3	1	**1**	3	1	**1**	1	−1	**1**
CC2[b]	QZVP	−13[d]	−12[d]	**7**	−10	−9	**6**	−10	−8	**6**	−15	−12	**8**
MP2[b]	QZVP	−10[d]	−9[d]	**5**	−7	−7	**4**	−7	−6	**4**	−12	−10	**6**
HF	QZVP	42	16	**17**	45	11	**18**	46	9	**18**	42	11	**18**
M06-2X	QZVP	−10	−11	**7**	−7	−19	**11**	−2	−29	**14**	-[c]	-[c]	-[c]
B3LYP	QZVP	−9	−10	**8**	−11	−7		−11	−5	**5**	−10	−10	**7**
BLYP	QZVP	−16	−16	**13**	−24	−9	**10**	−24	−7	**9**	−16	−14	**12**
CCSD(T)[b]	TZVP	12	11	**6**	4	11	**7**	4	12	**7**	12	11	**7**
LCCSD(T)[b]	TZVP	12	10	**6**	-[c]	-[c]	-[c]	-[c]	-[c]	-[c]	11	9	**6**
SCS-CC2[b]	TZVP	11	11	**7**	3	12	**7**	3	13	**8**	12	12	**7**
SCS-MP2[b]	TZVP	15	14	**8**	7	14	**9**	7	14	**9**	15	14	**9**
CC2[b]	TZVP	−1[d]	1[d]	**3**	−6	4	**4**	−6	6	**4**	0	3	**3**
MP2[b]	TZVP	4[d]	5[d]	**4**	−1	7	**5**	−1	9	**5**	4	6	**4**
HF	TZVP	35	11	**13**	38	6	**13**	39	4	**13**	37	7	**14**
M06-2X	TZVP	−17	−17	**12**	−15	−17	**12**	−9	−17	**10**	-[c]	-[c]	-[c]
B3LYP	TZVP	−15	−16	**12**	−19	−12	**10**	−19	−10	**9**	−17	−17	**13**
BLYP[b]	TZVP	−21[d]	−20[d]	**17**	−30	−13	**13**	−31	−10	**12**	−22	−20	**17**
PM6		−33[d]	−26[d]	**17**	−40	−40	**29**	−40	−48	**34**	-[c]	-[c]	-[c]
PM3		45	34	**24**	13	2	**13**	−12	−12	**12**	38	21	**21**
RM1		−32[d]	−27[d]	**17**	−22	−47	**29**	−22	−59	**37**	-[c]	-[c]	-[c]
AM1		64	42	**28**	32	11	**16**	18	−2	**12**	57	29	**24**
MNDO/d		228	162	**119**	195	128	**106**	169	114	**99**	215	142	**112**
MNDO		229	186	**131**	197	154	**119**	173	141	**113**	224	174	**128**

[a]The local coupled cluster approach LCCSD(T) in conjunction with the QZVP basis set is taken as a reference for gas phase and QM/MM values. Errors for the COSMO approach are taken relative to the SCS-CC2|QZVP level of theory. A detailed description of the relative energies is given in Figure 2. [b]Resolution of identity approximation or density fitting employed. [c]Method not available. [d]No local minimum or transition state found. Value represents the error of the respective grid point.

acceptable range. Whereas changes in the unsigned errors of B3LYP are below 6 kJ/mol, and therefore insignificant when the environment model is switched, BLYP and M06-2X can change up to 10 and 18 kJ/mol, respectively. Especially the M06-2X meta-hybrid functional shows an apparent error bias in conjunction with QZVP when COSMO is employed, which seems to increase as a function of the dielectric constant. The order of absolute errors ΔΔ*U*(ZW) is given by 11 kJ/mol < 19 kJ/mol < 29 kJ/mol for the gas phase, COSMO with $\varepsilon = 4$, and COSMO with $\varepsilon = 78$, respectively. Interestingly, this bias totally disappears when the smaller TZVP basis set is used, yielding a perfectly consistent error behavior. The analysis of errors of the relative energies ΔΔ*U*(ZW) shows that results for **ZW** in the case of M06-2X differ by 11−29 kJ/mol, B3LYP by 5−17 kJ/mol, and BLYP by 9−20 kJ/mol, depending on basis set size and environment model. The last one does not predict a minimum for the zwitterionic state in the gas phase, when the TZVP basis set is used. The deviations for the transition state energies ΔΔ*U*(TS) tend to be in the same range and comprise unsigned errors of 2−17 kJ/mol, 9−16 kJ/mol, and 23−48 respectively for M06-2X, B3LYP, and BLYP. The most apparent trend is the fact that all functionals tend to consistently underestimate barriers (2−31 kJ/mol) and relative energies (5−29 kJ/mol). This behavior has recently been discussed by Cohen et al.[121] and particularly by Sharma et al.[47] for proton transfer reactions in the gas phase. As shown in Table 2, this typical drawback of DFT methods also holds for our applied environment models COSMO and QM/MM. The effect of the basis set size is generally lower for the DFT methods than for the wave function based *ab initio* methods and ranges typically at the limit of significance (2−7 kJ/mol), when triple- and quadruple-ζ quality basis sets are compared for each functional. Due to their consistent trends to overestimate and underestimate the energy differences when triple-ζ basis sets are used, a combination of a wave function based method and DFT method should provide a good estimate of the error bar that can be expected for the proton transfer reaction. By taking SCS-MP2 and B3LYP as an example (Table 2), the expectable error moves within a range of 30 kJ/mol for the gas phase, 26−27 kJ/mol for COSMO, and 31−32 kJ/mol for QM/MM when the respective results of the two methods are taken as the upper and lower limits for ΔΔ*U*(TS) and ΔΔ*U*(ZW).

In contrast to the more or less consistent trends of *ab initio* approaches, semiempirical methods show a very erratic behavior for the accuracy of evaluated thermodynamic and kinetic parameters. The comparison of the proton transfer potentials in the gas phase predict only in the case of the

dx.doi.org/10.1021/ct301082y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 3. Deviations Δr of Minima N, Transition States TS, and Zwitterionic Minima ZW along the Reaction Coordinate r(S−H) in Picometers (pm)[a]

		gas phase			COSMO						QM/MM		
					ε = 4			ε = 78					
level of theory		Δr(N)	Δr(TS)	Δr(ZW)	Δr(N)	Δr(TS)	Δr(ZW)	Δr(N)	Δr(TS)	Δr(ZW)	Δr(N)	Δr(TS)	Δr(ZW)
LCCSD(T)[b]	QZVP	0	0	−7	[c]	[c]	[c]	[c]	[c]	[c]	0	0	2
SCS-CC2[b]	**QZVP**	**135**	**184**	**195**	**135**	**171**	**210**	**135**	**169**	**216**	**134**	**182**	**207**
SCS-MP2[b]	QZVP	0	0	2	0	0	1	0	0	1	0	0	2
CC2[b]	QZVP	0	[d]	[d]	0	−4	−8	0	−1	−7	0	0	−7
MP2[b]	QZVP	0	[d]	[d]	0	−1	−6	0	−1	−7	0	0	−5
HF	QZVP	−2	−4	37	−2	1	21	−2	1	16	−2	−3	32
M06-2X	QZVP	1	0	−1	1	−2	1	1	−2	18	[c]	[c]	[c]
B3LYP	QZVP	1	0	1	1	0	−2	1	0	0	1	1	2
BLYP	QZVP	3	1	−6	3	−1	−6	3	1	−6	3	2	−2
CCSD(T)[b]	TZVP	1	0	0	1	2	−2	1	1	−4	31	1	2
LCCSD(T)[b]	TZVP	1	0	3	[c]	[c]	[c]	[c]	[c]	[c]	1	1	5
SCS-CC2[b]	TZVP	1	1	−3	1	2	−3	1	1	−5	1	2	−1
SCS-MP2[b]	TZVP	0	0	−1	0	2	−2	0	1	−4	1	1	1
CC2[b]	TZVP	1	[d]	[d]	1	1	−9	1	0	−10	1	2	−8
MP2[b]	TZVP	1	[d]	[d]	1	1	−8	1	0	−8	1	2	−6
HF	TZVP	−1	−4	35	−1	1	20	−1	1	15	−1	−3	32
M06-2X	TZVP	2	0	−4	2	−2	0	1	−2	12	[c]	[c]	[c]
B3LYP	TZVP	3	0	1	2	1	−2	2	0	0	2	0	2
BLYP[b]	TZVP	−5	[d]	[d]	4	3	−7	4	1	−6	4	2	−2
PM6		3	[d]	[d]	3	−6	−14	4	−6	3	[c]	[c]	[c]
PM3		−2	0	26	−1	11	26	−1	11	32	−3	2	25
RM1		−1	[d]	[d]	−1	−8	25	−1	−6	47	[c]	[c]	[c]
AM1		−2	−1	33	−1	8	29	−1	4	19	−1	1	30
MNDO/d		−2	0	76	−2	11	50	−2	12	40	−2	2	64
MNDO		−6	0	64	−5	11	41	−5	11	32	−5	2	56

[a]Extremal values have been obtained from quadratic fits around the incremental points closest to minima or the transistion state. Absolute values of minima/maxima obtained from the fitted potential curve on the SCS-CC2|QZVP level of theory are given as bold numbers and are taken as a reference. A detailed description of the relative values Δr is given in Figure 3. [b]Resolution of identity approximation or density fitting employed. [c]Method not available. [d]No minima or maxima found.

"classical" semiempiric Hamiltonians MNDO(/d), AM1, and PM3 a transition state and a zwitterionic minimum along the reaction path. The more recent PM6 and further the reparameterized version of the Austin model 1, RM1, deliver a continuous repulsive energy curve along the minimum path without a local minimum for **ZW**. Considering robustness with respect to the environment model, PM6 and RM1 seem to be slightly superior, since maximum changes in energy are observed in the range of 7−32 kJ/mol, whereas all older models comprise maximum changes by 44−59 kJ/mol when the gas phase, COSMO, and QM/MM are compared for the same method. Unsigned errors for the relative energies $\Delta\Delta U$(ZW) and $\Delta\Delta U$(TS) predicted by PM6, PM3, AM1, and RM1 are settled between 2 and 64 kJ/mol and further exceed 112 kJ/mol for the two MNDO variants. Except for this difference, no clear accuracy order can be derived for the semiempirical methods, if all environment models are taken into account. Considering MNDO and its extended implementation MNDO/d, including d functions, **ZW** is overestimated by more than 112 kJ/mol, and the barrier heights $\Delta\Delta U$(TS) are beyond 169 kJ/mol, which make them practically unusable for the proton transfer reaction. This underlines the deficiencies of these methods with respect to hydrogen bond description. The more often employed AM1 and PM3 Hamiltonians range in a more acceptable region for $\Delta\Delta U$(ZW) between −12 and 42 kJ/mol and a slightly broader range for $\Delta\Delta U$(TS) between an underestimation of 12 kJ/mol and an overestimation of 64 kJ/mol. With respect to the environment model, they provide surprisingly good results in conjunction with the COSMO model, where the unsigned errors for $\Delta\Delta U$(ZW) are below 12 kJ/mol. Most interesting are the results obtained with the recent semiempirical methods PM6 and RM1. Although no minima for the gas phase are found and QM/MM results are not available, the results obtained from the two COSMO cases are rather disappointing. Particularly, when a high dielectric constant of ε = 78 is applied, RM1 and PM6 are the only methods throughout this benchmarking that predict the energy of **ZW** below that of **N**. In detail, the error of the relative energy $\Delta\Delta U$(ZW) reaches −48 kJ/mol for PM6 and −59 kJ/mol for RM1, featuring a massive overstabilization of the zwitterion. In the case of PM6, this problem has also been mentioned by Stewart[122] in terms of the applicability of their semiempirical Hamiltonian to protein modeling. The overstabilization seems further to be a general trend for the Recife Model RM1, since findings from Lee et al.[29] indicate similar problems, when zwitterions are involved.

Mean Absolute Errors. The mean absolute errors (MUE, defined in eq 1), given in Table 2, are well suited to get an overall estimate of the accuracies for the whole proton transfer potential and further compare the different environment models. Considering the wave function based methods HF, CC2, MP2, SCS-MP2, SCS-CC2, CCSD(T), and LCCSD(T), the respective MUEs obtained for the different environment models do not change much (∼1−2 kJ/mol). This reflects the

dx.doi.org/10.1021/ct301082y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

robust behavior of the wave function based methods, already found for the thermodynamic properties. The basis set size (TZVP vs QZVP) on average changes the computed potential systematically by 1−8 kJ/mol, whereas the spin-component scaled counterparts are slightly more affected (5−8 kJ/mol) than the other wave function based methods (1−6 kJ/mol). Therefore, SCS-CC2 and SCS-MP2 seem to benefit most from a bigger basis set.

The density functional theories show also a robust behavior toward the different environment models in most cases and do not exceed 7 kJ/mol difference in the MUEs by going from the gas phase over COSMO to the QM/MM approach. Differences between the accuracy of the three functionals are too small to be significant, but nevertheless there is a consistent improvement in accuracy of 1−6 kJ/mol by going from a triple-ζ basis set to quadruple-ζ quality.

The MUE values for the spectrum of semiempirical methods emphasize the error behavior, found for the thermodynamic and kinetic data. By comparing the trends, it becomes apparent that the older semiempirical methods MNDO(/d), AM1, and PM3 show increased accuracy by going from the gas phase toward the COSMO model, whereas the most recent Hamiltonians PM6 and RM1 descend from their superior performance in the gas phase. In the gas phase, the latter ones predict the proton transfer potential in a quality (MUE = 17 kJ/mol) comparable to the density functional BLYP. However, if the environment is modeled by the COSMO approach (ε = 78), the mean unsigned error rises up to 34 kJ/mol. MUEs of the older semiempirical ones decrease to appealing values of 12 kJ/mol for PM3 and AM1. Taking also into account that no zwitterionic minimum is predicted by PM6 and RM1, both approaches are unsuitable for calculating the proton transfer potential for our case study. The most robust behavior with respect to the environment model provides PM3, where the maximum difference of MUEs is 12 kJ/mol, followed by AM1 with 16 and 17 kJ/mol for RM1 and PM6. The detailed results of this comparison might also be applied in terms of error scaling, to reduce the systematic errors by going from gas phase models to the COSMO or QM/MM approach. The direct comparison of AM1 and RM1 gives further interesting insights into the bias of the semiempirical methods. Since RM1 represents only a reparameterization of AM1 and therefore does not differ in its functional form or implementation, it shows to what extent the choice of parameters influences the robustness of the semiempirical method with respect to the environment model. For our model reaction, best performance in the gas phase is achieved by RM1 with a MUE of 17 kJ/mol, and worse values are obtained by AM1 (28 kJ/mol). In conjunction with COSMO, mimicking a very polar environment of ε = 78, the AM1 method can even exceed these accuracies with a mean unsigned error of 12 kJ/mol in contrast to a significant dropped performance of 34 kJ/mol for RM1. This change of trend represents a new aspect in terms of method accuracy. It underlines previous conclusions[39] that, despite the recent developments of RM1 and PM6, semiempirical methods are still unreliable for proton transfer predictions. It further indicates that the progress of achieving higher accuracy, like recently demonstrated by RM1 or PM6, seems to be bought on the cost of robustness, thus introducing a stronger bias toward gas phase data, which are typically used for parameter optimizations.

Location of Minima and Transition States. The proton transfer potential between cysteine and histidine reveals the neutral state **N**, the zwitterionic state **ZW**, and the transition state **TS**. The r(S−H) distance shown in Figure 1 serves as a reaction coordinate for our case study. It represents a covalent sulfur−hydrogen bond for low values and a noncovalent hydrogen bond between the anionic sulfur atom and the oxygen-bound hydrogen atom for larger distances. Table 3 gives the offsets from the locations, when other methods are compared, which were computed as described in the benchmarking setup section and illustrated in Figure 3. Further, the reference values (SCS-CC2|QZVP) are given in Table 3.

The analysis of the reference values in the gas phase and COSMO reveal the tendency to predict a zwitterionic state **ZW** that is shifted further away for increased polarity of the environment, whereas the predicted **TS** moves slightly closer to **N**. The QM/MM approach delivers **TS** at 182 pm, which is similar to the gas phase description, and **ZW** at 207 pm, comparable to the COSMO prediction for a low dielectric constant of ε = 4. The lowest state **N** of the proton transfer potential is well described by nearly all methods. Only BLYP and MNDO without d functions overestimate or underestimate the location of the minimum by up to 6 pm. The rest of the methods show absolute deviations below 4 pm, whereas the DFT approaches possess a consistent tendency of marginal overestimations. By taking a look toward **TS** and **ZW** in the gas phase, it becomes obvious that only half of the tested methods predict a transition state, and thus a local minimum for the zwitterionic state on the proton transfer potential. However, we have to admit that our case study is challenging, since the shape of the potential is very flat in that region. The comparison of location of the **TS** among the various methods reveals differences between the accuracy of *ab initio* methods that in general do not deviate by more than 4 pm from the reference and the semiempirical Hamiltonians. For the latter ones, an interesting observation with respect to the error behavior within the gas phase, COSMO, and QM/MM can be made. Closer inspection of the semiempirical methods shows that the recent methods RM1 and PM6 behave differently than the older Hamiltonians. Although the absolute deviations of the transition state location range in the same magnitude, namely up to 12 pm beyond the reference, PM6 and RM1 both tend to underestimate the r(S−H) of **TS**, whereas the "classical" methods PM3, AM1, and MNDO(/d) consistently overestimate it. The location of the zwitterionic state **ZW** shows in general larger deviations than both of the other stationary points. Besides the shortcomings of the different methods, this is also due to the fact that the potential is much flatter in this region (Figure 5). Comparing the gas phase, COSMO, and QM/MM, no systematic behavior can be observed. Nevertheless, the relative comparison of all Hamiltonians reveals some trends that mainly confirm the observations made for the MUEs and the thermodynamic data. Accordingly, HF shows large deviations from the reference values, but also the unscaled versions of MP2 and CC2 can deviate by more than 10 pm. The DFT methods deliver quite reasonable predictions, in particular B3LYP, which is in line with previous findings from Adamo and co-workers.[123] They comprise errors below 7 pm with the exception of the M06-2X functional. The lineup of the gas phase offset of −1 pm and the COSMO (ε = 78) offset of 18 pm results in a maximum difference of 19 pm for the zwitterionic state along the same proton transfer potential when a QZVP basis set is used. Compared to the B3LYP functional that differs by less than 2 pm from the reference value across all environment models, M06-2X seems to be even less robust

than BLYP (changes < 5 pm), which ranks two rungs below on the "Jacobs ladder" of density functional theory. In contrast to the observations made for the thermodynamic properties, where this "COSMO problem" was only observed for the QZVP basis set, this deficiency does not vanish here for the smaller basis set. For TZVP, the error range still spans across 16 pm, which is apparently more than the deviations of the other functionals.

Largest deviations and mostly unsystematic error behavior are observed for the semiempirical methods. Besides the expected bad performance of MNDO(/d) for the location of **ZW**, the recently developed RM1 method delivers a rather disappointing performance with a maximum offset of 47 above the reference in the case of the COSMO description with ε = 78. In contrast to this, PM6 makes a surprisingly good prediction that deviates only 3 pm from the reference. Unfortunately, this encouraging result is accompanied by considerable errors in the relative energy of **ZW**, as already discussed. At the end, PM3 and AM1 Hamiltonians remain with deviations of 22−24 pm and 19−33 pm, respectively, depending on the environment model.

In summary, only the B3LYP functional provides excellent predictions of all stationary points in our test case, by keeping sufficiently consistent throughout the different environment models and the two basis sets. This finding is also supported by reports from Clyaessens et al., who compared QM/MM transition state geometries predicted by B3LYP with LCCSD-(T) results.[124] Furthermore, the wave function based approach SCS-MP2 behaves similarly robustly.

CONCLUSIONS

In summary, we have evaluated some critical aspects for the description of the cysteine−histidine proton transfer that might also hold for other proton transfer reactions in the protein environment. Our case study shows the impact of implicit and explicit environment modeling on the proton transfer potential and demonstrates that the application of a continuum solvent model with a low ε value of 4, corresponding to protein bulk, still tends to overestimate stabilization of the ionic state and transition state. Therefore, its use for the modeling of biochemical proton transfer reactions seems to be limited. By evaluating the robustness of the probed quantum chemical methods with respect to the environment model, we find a quite good correlation with the applied level of theory. Hence, the wave function based *ab initio* methods show consistent error behavior and do not change their accuracy significantly by switching between the gas phase, COSMO, and QM/MM approach. For the TZVP basis set, a consistent trend of overestimation (12−16 kJ/mol) of **ZW** can be observed for the *post-HF* methods. As expected, the CC2 approach, which is usually applied for excited states, delivers similar results as MP2 for our ground state problem.

The density functional theory methods show marginally higher deviations for thermodynamic and kinetic predictions and MUEs, depending on the functional. The recent M06-2X functional reveals weaknesses in combination with the COSMO model, which seems to introduce larger errors (up to 29 kJ/mol) for the relative energies and location of **ZW**, when higher dielectric constants are applied. Therefore, it can be assumed that its superior accuracy in the gas phase is not necessarily represented within the COSMO description and possibly in QM/MM approaches. By taking the performance of thermodynamic data, MUEs, and location of stationary points into account, B3LYP appears to be the most robust functional with respect to the change of the environment model and can, in agreement with other studies,[40] deliver proton transfer predictions comparable to wave-function-based approaches. As a general trend, consistent underestimation of **TS** and **ZW** (2−31 kJ/mol) is found for all DFT methods.

The semiempirical methods PM6, PM3, RM1, AM1, and MNDO(/d), which are often used for the computation of free energies, show a very erratic and unsystematic error behavior throughout all evaluated properties. Although the recent methods PM6 and RM1 provide significant advances in gas phase accuracy, in line with assessments in other studies,[108,110,122,125,126] their application within the COSMO model reveals a dramatic change in performance. For our case study, they wrongly predict the stability order of **N** and **ZW**. Whereas this "zwitterion problem" is already described for PM6,[122] it seems also to hold for RM1. Besides this deficiency, none of the semiempirical methods deliver a satisfactory robustness, since they possess changes up to 59 kJ/mol for the thermodynamic data by going from one environment model to another. Taking these facts together, the fewest problems can only be expected for PM3 and AM1. However, PM3 has the drawback of absolute errors in the maximum range of 45 kJ/mol for the thermodynamic and kinetic predictions. AM1 also behaves similar and comprises a maximum error of 64 kJ/mol for our test case.

In summary, all methods possess strengths and weaknesses for our challenging case study. Therefore, we can recommend applying a combination of two methods to obtain a trustworthy range for the prediction of proton transfer potentials. Due to their consistent trends to overestimate and underestimate barriers and energy differences, when triple-ζ basis sets are used, a combination of a wave function based method and a DFT method should deliver a good estimate of the error bar that can be expected for the proton transfer reaction, by keeping the computational costs feasible.

ASSOCIATED CONTENT

Supporting Information

Additional plots of the QM/MM proton transfer potentials obtained by full optimizations on the MP2|TZVP and B3LYP|TZVP level, as well as on the semiempirical AM1 and PM3 level, also including direct comparisons of the respective potentials when calculated with relaxed and unrelaxed geometries. Plots showing the influence of the RSOLV multiplicative factor, which are used for cavity construction in the COSMO model, on the relative energies of **TS** and **ZW**. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bernd.engels@mail.uni-wuerzburg.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support by the Deutsche Forschungsgemeinschaft within the framework of the SFB 630 "Recognition, Preparation and Functional Analysis of Agents against Infectious Diseases" and of the Volkswagen Stiftung is gratefully acknowledged.

REFERENCES

(1) Stanton, C. L.; Houk, K. N. *J. Chem. Theory Comput.* **2008**, *4*, 951−966.
(2) Davies, M. N.; Toseland, C. P.; Moss, D. S.; Flower, D. R. *BMC Biochem.* **2006**, *7*, 18.
(3) Liao, C.; Nicklaus, M. C. *J. Chem. Inf. Model.* **2009**, *49*, 2801−2812.
(4) Borštnar, R.; Repin, M.; Kamerlin, S. C. L.; Vianello, R.; Mavri, J. *J. Chem. Theory Comput.* **2012**, *8*, 3864−3870.
(5) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458−4472.
(6) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.
(7) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1−20.
(8) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187−4226.
(9) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.
(10) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(11) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(12) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(13) Schmidt, T. C.; Paasche, A.; Grebner, C.; Ansorg, K.; Becker, J.; Lee, W.; Engels, B. In *Topics in Current Chemistry*; Springer Berlin Heidelberg: Berlin, 2012; pp 1−77.
(14) Hu, H.; Yang, W. *J. Mol. Struct.: THEOCHEM* **2009**, *898*, 17−30.
(15) Zhang, R.; Lev, B.; Cuervo, J. E.; Noskov, S. Y.; Salahub, D. R. In *Advances in Quantum Chemistry*; Elsevier: New York, 2010; Vol. *59*, pp 353−400.
(16) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3094.
(17) Zhang, S. *J. Comput. Chem.* **2012**, *33*, 517−526.
(18) Klamt, A.; Eckert, F.; Arlt, W. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 101−122.
(19) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Bioinf.* **2001**, *44*, 400−417.
(20) Shokhen, M.; Khazanov, N.; Albeck, A. *Proteins* **2009**, *77*, 916−926.
(21) Štrajbl, M.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 4471−4484.
(22) Paasche, A.; Schiller, M.; Schirmeister, T.; Engels, B. *ChemMedChem* **2010**, *5*, 869−880.
(23) Derat, E.; Shaik, S.; Rovira, C.; Vidossich, P.; Alfonso-Prieto, M. *J. Am. Chem. Soc.* **2007**, *129*, 6346−6347.
(24) Kästner, J.; Sherwood, P. *Mol. Phys.* **2010**, *108*, 293.
(25) Mladenovic, M.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. *J. Am. Chem. Soc.* **2008**, *130*, 8696−8705.
(26) Kaukonen, M.; Söderhjelm, P.; Heimdal, J.; Ryde, U. *J. Chem. Theory Comput.* **2008**, *4*, 985−1001.
(27) Ke, Z.; Guo, H.; Xie, D.; Wang, S.; Zhang, Y. *J. Phys. Chem. B* **2011**, *115*, 3725−3733.
(28) Ke, Z.; Zhou, Y.; Hu, P.; Wang, S.; Xie, D.; Zhang, Y. *J. Phys. Chem. B* **2009**, *113*, 12750−12758.
(29) Lee, W.; Luckner, S. R.; Kisker, C.; Tonge, P. J.; Engels, B. *Biochemistry* **2011**, *50*, 5743−5756.
(30) Hu, X.; Hu, H.; Melvin, J. A.; Clancy, K. W.; McCafferty, D. G.; Yang, W. *J. Am. Chem. Soc.* **2011**, *133*, 478−485.
(31) Cheng, Y.; Cheng, X.; Radić, Z.; McCammon, J. A. *J. Am. Chem. Soc.* **2007**, *129*, 6562−6570.
(32) Mladenovic, M.; Schirmeister, T.; Thiel, S.; Thiel, W.; Engels, B. *ChemMedChem* **2007**, *2*, 120−128.
(33) Xie, H.-B.; Zhou, Y.; Zhang, Y.; Johnson, J. K. *J. Phys. Chem. A* **2010**, *114*, 11844−11852.
(34) Kaukonen, M.; Söderhjelm, P.; Heimdal, J.; Ryde, U. *J. Phys. Chem. B* **2008**, *112*, 12537−12548.
(35) Olsson, M. H. M.; Siegbahn, P. E. M.; Warshel, A. *J. Am. Chem. Soc.* **2004**, *126*, 2820−2828.
(36) Wong, K. F.; Sonnenberg, J. L.; Paesani, F.; Yamamoto, T.; Vaníček, J.; Zhang, W.; Schlegel, H. B.; Case, D. A.; Cheatham, T. E.; Miller, W. H.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 2566−2580.
(37) Christ, C. D.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2010**, *31*, 1569−1582.
(38) Li, G.-S.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F. *Chem. Phys. Lett.* **1998**, *297*, 38−44.
(39) Range, K.; Riccardi, D.; Cui, Q.; Elstner, M.; York, D. M. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3070.
(40) Anick, D. J. *J. Phys. Chem. A* **2003**, *107*, 1348−1358.
(41) Sadhukhan, S.; Muñoz, D.; Adamo, C.; Scuseria, G. E. *Chem. Phys. Lett.* **1999**, *306*, 83−87.
(42) Jursic, B. S. *J. Mol. Struct.* **1998**, *427*, 137−142.
(43) Frey, J. A.; Müller, A.; Losada, M.; Leutwyler, S. *J. Phys. Chem. B* **2007**, *111*, 3534−3542.
(44) Brás, N. F.; Perez, M. A. S.; Fernandes, P. A.; Silva, P. J.; Ramos, M. J. *J. Chem. Theory Comput.* **2011**, *7*, 3898−3908.
(45) Li, G.-S.; Maigret, B.; Rinaldi, D.; Ruiz-Lopez, M. F. *J. Comput. Chem.* **1998**, *19*, 1675−1688.
(46) Harrison, M. J.; Burton, N. A.; Hillier, I. H. *J. Am. Chem. Soc.* **1997**, *119*, 12285−12291.
(47) Sharma, R.; Thorley, M.; McNamara, J. P.; Watt, C. I. F.; Burton, N. A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2475.
(48) Gunner, M. R.; Mao, J.; Song, Y.; Kim, J. *Biochchim. Biophys. Acta* **2006**, *1757*, 942−968.
(49) Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573−601.
(50) Suter, H. U.; Pleß, V.; Ernzerhof, M.; Engels, B. *Chem. Phys. Lett.* **1994**, *230*, 398−404.
(51) Engels, B.; Peyerimhoff, S. D. *J. Phys. Chem.* **1989**, *93*, 4462−4470.
(52) Zhao, H.-M.; Pfister, J.; Settels, V.; Renz, M.; Kaupp, M.; Dehm, V. C.; Würthner, F.; Fink, R. F.; Engels, B. *J. Am. Chem. Soc.* **2009**, *131*, 15660−15668.
(53) Helten, H.; Schirmeister, T.; Engels, B. *J. Phys. Chem. A* **2004**, *108*, 7691−7701.
(54) Sherwood, P.; De Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct.* **2003**, *632*, 1−28.
(55) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.
(56) Simonson, T.; Brooks, C. L. *J. Am. Chem. Soc.* **1996**, *118*, 8452−8458.
(57) Gilson, M. K.; Honig, B. H. *Biopolymers* **1986**, *25*, 2097−2119.
(58) Huang, C.; Wei, P.; Fan, K.; Liu, Y.; Lai, L. *Biochemistry* **2004**, *43*, 4568−4574.
(59) Solowiej, J.; Thomson, J. A.; Ryan, K.; Luo, C.; He, M.; Lou, J.; Murray, B. W. *Biochemistry* **2008**, *47*, 2617−2630.
(60) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, 1253−1272.
(61) Warshel, A. *Biochemistry* **1981**, *20*, 3167−3177.
(62) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13190−13195.
(63) Xue, X.; Yang, H.; Shen, W.; Zhao, Q.; Li, J.; Yang, K.; Chen, C.; Jin, Y.; Bartlam, M.; Rao, Z. *J. Mol. Biol.* **2007**, *366*, 965−975.
(64) Lee, T.-W.; Cherney, M. M.; Liu, J.; James, K. E.; Powers, J. C.; Eltis, L. D.; James, M. N. G. *J. Mol. Biol.* **2007**, *366*, 916−932.
(65) Tan, J.; Verschueren, K. H. G.; Anand, K.; Shen, J.; Yang, M.; Xu, Y.; Rao, Z.; Bigalke, J.; Heisen, B.; Mesters, J. R.; Chen, K.; Shen, X.; Jiang, H.; Hilgenfeld, R. *J. Mol. Biol.* **2005**, *354*, 25−40.
(66) Zheng, K.; Ma, G.; Zhou, J.; Zen, M.; Zhao, W.; Jiang, Y.; Yu, Q.; Feng, J. *Proteins* **2007**, *66*, 467−479.
(67) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175−3183.
(68) Schütz, M.; Manby, F. R. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349.
(69) Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*, 1st ed.; Wiley: New York, 2000.

(70) Werner, H.-J.; Pflüger, K. In *Annual Reports in Computational Chemistry*; Spellmeyer, D. C., Ed.; Elsevier: New York, 2006; Vol. *2*, pp 53−80.
(71) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
(72) Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670.
(73) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283−312.
(74) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(75) Wang, H.; Kim, Y. T.; Muramatsu, T.; Takemoto, C.; Shirouzu, M.; Yokoyama, S. Crystal structure of SARS coronavirus main proteinase (3CLPRO). 2007.
(76) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(77) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(78) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704−721.
(79) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins* **2008**, *73*, 765−783.
(80) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Am. Chem. Soc.* **2011**, *7*, 525−537.
(81) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327−341.
(82) Verlet, L. *Phys. Rev.* **1967**, *159*, 98−103.
(83) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.
(84) Sinclair, P. E.; De Vries, A.; Sherwood, P.; Catlow, R. A.; Van Santen, R. A. *J. Am. Chem. Soc., Faraday Trans.* **1998**, *94*, 3401−3408.
(85) MacKerell, A. D., Jr.; Bashford, D.; Bellott; Dunbrack; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(86) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
(87) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571−2577.
(88) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119−124.
(89) Kendall, R. A.; Früchtl, H. A. *Theor. Chim. Acta* **1997**, *97*, 158−163.
(90) Mangiatordi, G. F.; Brémond, E.; Adamo, C. *J. Chem. Theory Comput.* **2012**, *8*, 3082−3088.
(91) Akerlof, G. C.; Oshry, H. I. *J. Am. Chem. Soc.* **1950**, *72*, 2844−2847.
(92) Bonaccorsi, R.; Palla, P.; Tomasi, J. *J. Am. Chem. Soc.* **1984**, *106*, 1945−1950.
(93) Bonaccorsi, R.; Cammi, R.; Tomasi, J. *J. Comput. Chem.* **1991**, *12*, 301−309.
(94) Werner, H.-J.; Schütz, M. *J. Chem. Phys.* **2011**, *135*, 144116.
(95) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. *MOLPRO*, version 2010.1; Cardiff University: Cardiff, U. K.; Universität Stuttgart: Stuttgart, Germany, 2010.
(96) Werner, H.-J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*, 8149−8160.
(97) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331−340.
(98) Hättig, C.; Weigend, F. *J. Chem. Phys.* **2000**, *113*, 5154−5161.
(99) Häser, M.; Ahlrichs, R. *J. Comput. Chem.* **2004**, *10*, 104−111.
(100) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(101) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165−169.
(102) Hellweg, A.; Grün, S. A.; Hättig, C. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4119.
(103) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.
(104) Weigend, F.; Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *119*, 12753−12762.
(105) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *120*, 215−241.
(106) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(107) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(108) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J. Comput. Chem.* **2006**, *27*, 1101−1111.
(109) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.
(110) Stewart, J. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(111) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(112) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391−404.
(113) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1996**, *93*, 315−315.
(114) Stewart, J. J. P. *MOPAC2009*; Stewart Computational Chemistry: Colorado Springs, CO, 2009.
(115) Thiel, W. *MNDO2005*; Max-Planck-Institut für Kohlenforschung: Mülheim, Germany, 2005.
(116) Xu, X.; Alecu, I. M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 1667−1676.
(117) Korth, M.; Grimme, S. *J. Chem. Theory Comput.* **2009**, *5*, 993−1003.
(118) Perdew, J. P.; Schmidt, K. *AIP Conf. Proc.* **2001**, *577*, 1−20.
(119) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
(120) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463−1473.
(121) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. *Science* **2008**, *321*, 792−794.
(122) Stewart, J. *J. Mol. Model.* **2009**, *15*, 765−805.
(123) Barone, V.; Orlandini, L.; Adamo, C. *Chem. Phys. Lett.* **1994**, *231*, 295−300.
(124) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. *Angew. Chem.* **2006**, *118*, 7010−7013.
(125) Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1749−1760.
(126) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. A* **2006**, *110*, 13551−13559.