Does the DFT Self-Interaction Error Affect Energies Calculated in Proteins with Large QM Systems?

Fouda, Adam; Ryde, Ulf

Published in:
Journal of Chemical Theory and Computation

DOI:
10.1021/acs.jctc.6b00903

2016

Document Version:
Publisher's PDF, also known as Version of record

JCTC Journal of Chemical Theory and Computation

Article

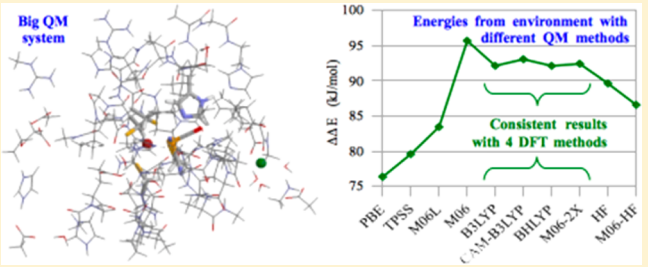
pubs.acs.org/JCTC

Does the DFT Self-Interaction Error Affect Energies Calculated in Proteins with Large QM Systems?

Adam Fouda and Ulf Ryde*

Department of Theoretical Chemistry, Chemical Centre, Lund University, P. O. Box 124, SE-221 00 Lund, Sweden

ABSTRACT: We have examined how the self-interaction error in density-functional theory (DFT) calculations affects energies calculated on large systems (600−1000 atoms) involving several charged groups. We employ 18 different quantum mechanical (QM) methods, including Hartree−Fock, as well as pure, hybrid, and range-separated DFT methods. They are used to calculate reaction and activation energies for three different protein models in vacuum, in a point-charge surrounding, or with a continuum-solvent model. We show that pure DFT functionals give rise to a significant delocalization of the charges in charged groups in the protein, typically by ∼0.1 *e*, as evidenced from the Mulliken charges. This has a clear effect on how the surroundings affect calculated reaction and activation energies, indicating that these methods should be avoided for DFT calculations on large systems. Fortunately, methods such as CAM-B3LYP, BHLYP, and M06-2X give results that agree within a few kilojoules per mole, especially when the calculations are performed in a point-charge surrounding. Therefore, we recommend these methods to estimate the effect of the surroundings with large QM systems (but other QM methods may be used to study the intrinsic reaction and activation energies).



INTRODUCTION

During the latest two decades, quantum mechanical (QM) calculations using density-functional theory (DFT) have become a standard tool in the investigation of the structure and function of biomolecules, owing to their favorable combination of a high accuracy and a low computational cost.[1,2] DFT is a powerful approach, and the best functionals often give an accuracy approaching that of advanced wave function methods. However, it must be remembered that DFT is only an approximate approach.

Arguably, the most severe shortcoming of the DFT methods is the self-interaction error (SIE).[3−5] It arises from the spurious interaction of an electron with itself in the Coulomb term in the DFT Hamiltonian, which is not exactly canceled by the exchange contribution, e.g., as in the Hartree−Fock (HF)[6,7] approach. As a consequence, the eigenvalues of the highest occupied orbitals do not correspond to the ionization potential, the gap between the highest occupied and the lowest unoccupied orbitals becomes too small for ionic systems, charge tends to be overly delocalized, and charge-transfer transitions get too low energies. The effect may be further enhanced by an incorrect asymptotic decay of the exchange potential.

The SIE can be partly reduced by employing a portion of the exact HF exchange in the DFT calculations, giving so-called hybrid functionals (in contrast to pure functionals that do not contain such contributions).[4,8] Alternatively, the standard DFT functional is used only for the short-ranged exchange interactions, whereas the HF method is used for long-ranged interactions, employing a switching function for intermediate distances.[9] Thereby, an asymptotically correct decay is also obtained. DFT functionals employing such an approach are called range-separated functionals. Perdew and Zunger have suggested methods to correct for the SIE[3], and a variational approach with complex orbitals has been implemented.[10] Unfortunately, this approach is not invariant to unitary transformation of the orbitals and it does not always improve the accuracy of the results.[5,11]

With the continuous improvement of computer hardware and QM software performance, increasingly large systems can be treated by DFT methods. Currently, single-point energies can be calculated for over 1000 atoms in one or a few days with standard QM software and several thousands of atoms can be treated with more specialized software.[12−14] This opens up many interesting applications, e.g., calculating accurate reaction or ligand-binding energies in protein reactions, NMR chemical shifts, reduction potentials, or polarized charges for entire proteins, automatically including all important interaction energies, e.g., electrostatics, polarization, charge transfer, charge penetration, and exchange-repulsion, i.e., effects that are hard to treat accurately with molecular-mechanics methods.[15−27]

However, recently Jakobsen et al. showed that the charge distribution obtained for large protein models, involving groups with net positive and negative charges, are incorrect when calculated with pure DFT methods, owing to the SIE.[8] For example, in zwitterionic polyglycine chains, pure DFT methods tended to reduce the charge of the charged terminal groups, so that approximately half of an electron was transferred from the negatively charged carboxy terminal to the positively charged

Received: September 14, 2016
Published: October 17, 2016

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

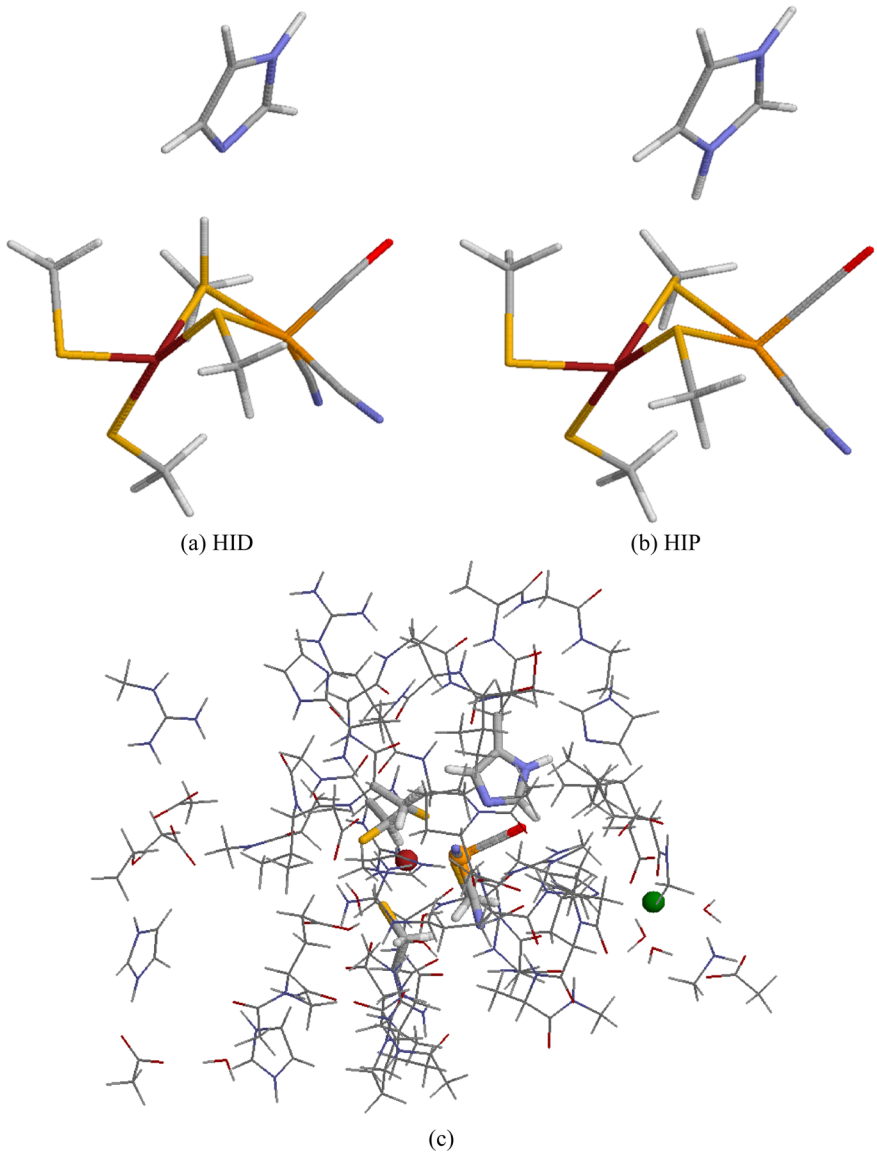


Figure 1. Minimal QM system for [NiFe] hydrogenase in the (a) HID and (b) HIP states. The corresponding big-QM system is shown in panel c with the minimal QM system highlighted in thick stick and the three ions shown as balls (including a Mg^{2+} ion in green).

amino terminal. The error increased with the distance of the two charges, but it could be essentially removed by using hybrid functionals with 50% or more exact exchange or range-separated DFT functionals. Similar problems were also observed for the insulin monomer, a 51-residue polypeptide with five positively and six negatively charged residues.[8]

Naturally, this may be a disaster for all approaches that employ large protein models with charged sites to calculate energies and properties in proteins. For example, we have suggested the big-QM approach to calculate accurate and stable reaction and activation energies in proteins.[12,26] It suggests that all chemical groups within 4.5−6 Å of a minimal active-site model should be included in the calculations, as well as two capped residues around each protein residue in the active site and all buried charged groups in the protein, typically 600−1000 atoms and several charged groups. This approach has been employed for several proteins,[25,28−31] and similar approaches have been suggested by other groups.[27]

In this work, we investigate how serious the SIE is for the calculation of big-QM energies for three different proteins, viz., [NiFe] hydrogenase,[32] glyoxalase I,[33] and sulfite oxidase.[34] We employ 17 different DFT methods, pure, hybrid, or range-separated, together with HF calculations. We also investigate whether the problem can be reduced by performing the calculations in a point-charge or continuum-solvation model of the surroundings.

METHODS

Systems. Three different enzyme models were employed, all containing metals. For each enzyme, QM models of two sizes (one minimal and one big) were employed to separate the intrinsic reaction energy of the active site and the effect of the surroundings. The first studied system was a model of a simple proton-transfer reaction between a Cys and a His residue in the active site of [NiFe] hydrogenase.[25,26,32,35] This system was used as a test system when the big-QM method was developed.[12] The small QM model contained the Ni and Fe ions, the first-sphere four Cys, one CO, and two CN^- metal ligands, as well as the second-sphere His model. This 38-atom system is shown in Figure 1a,b. The corresponding big-QM

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

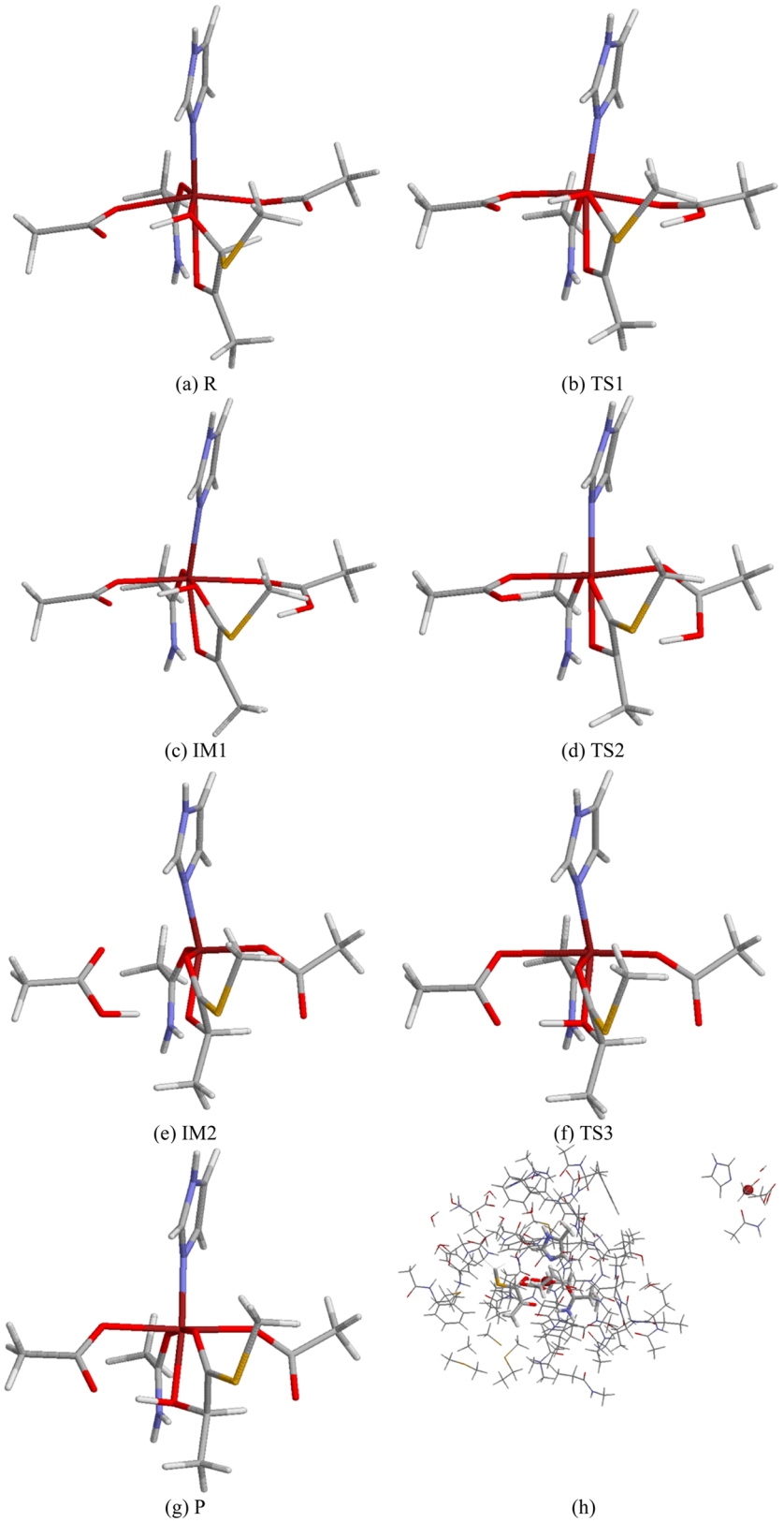


Figure 2. Minimal QM system for glyoxalase I in the (a) R, (b) TS1, (c) IM1, (d) TS2, (e) IM2, (f) TS3, and (g) P states. The corresponding big-QM system is shown in panel h with the minimal QM system highlighted in thick sticks and the Zn site in the other subunit to the right.

system was developed in ref 12 (system C). It contained 12 cationic, 15 anionic, and 38 neutral groups (including a Mg^{2+} ion), in total 675 atoms (4.5 Å surrounding). It is shown in Figure 1c. The net charge was +1. For this system, we calculated the energy difference between two states, one with the proton on the Cys residue (called the HID state; Figure 1a) and the other with the proton on His (called the HIP state; Figure 1b). The two systems differ mainly in the ∼0.64 Å

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

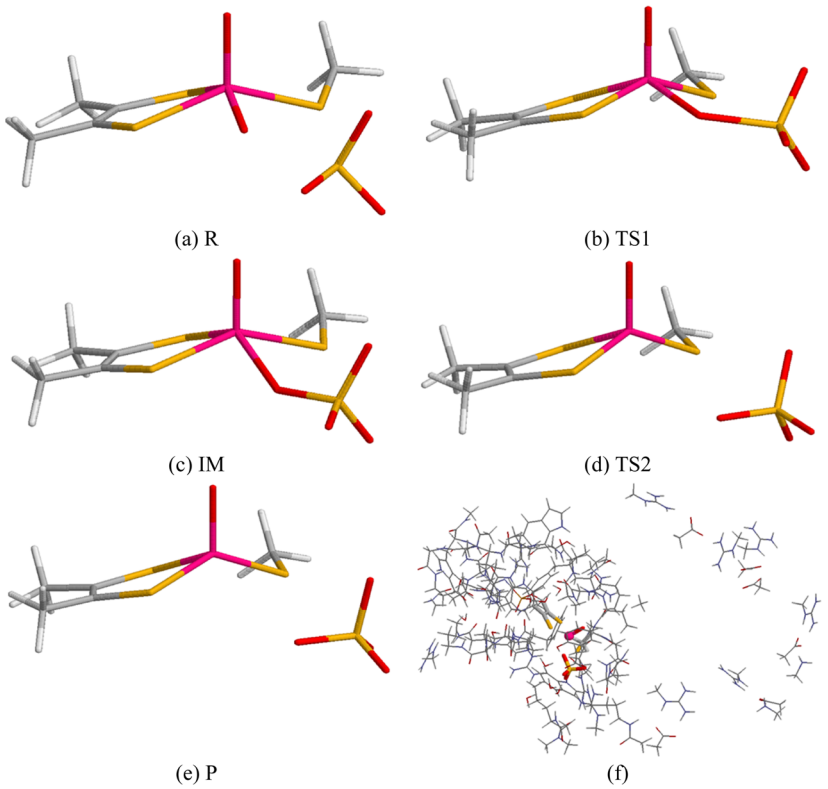


Figure 3. Minimal QM system for sulfite oxidase in the (a) R, (b) TS1, (c) IM, (d) TS2, and (e) P states. The corresponding big-QM system is shown in panel f with the minimal QM system highlighted in thick sticks.

movement of the proton (the S−H and H−N distances are 1.41 and 1.64 Å in the HID state and 2.01 and 1.09 Å in the HIP state).

Ten different structures of each state were studied, taken from a molecular dynamics simulation of the HIP state.[12] Five of the snapshots were then optimized by QM/MM for the HIP state, using QM for the minimal QM system (enhanced with a second-sphere protonated Glu group) and allowing all residues within 6 Å of the small-QM system to move, followed by QM/MM optimization of the HID state with all atoms outside the small-QM system fixed. For the other five snapshots, the HID state was instead first optimized, followed by the HIP state. The HID state was obtained by restraining the S−H distance to 1.405 Å, which is the optimum distance for the isolated small QM system, where the HID state is a local minimum. The coordinates of all atoms have changed between the various snapshots (by 1.1 Å on average for the protein atoms). However, the active sites are less affected, with 0.2 Å average movement of the atoms in the minimal QM system and with a maximum variation of the metal−ligand bond lengths of less than 0.05 Å. Both the Fe and Ni ions were in their low-spin +II oxidation states (giving closed-shell calculations).[28,36]

The second system was an active-site model of glyoxalase I.[33] The minimal QM system contained a Zn^{2+} ion; four coordinating amino acids, Gln33, Glu99, His126, and Glu172; and a substrate model ($CH_3COCH(OH)SCH_3$, *S* isomer).[37] This 48-atom system is neutral and is shown in Figure 2a−g. The corresponding big-QM system involved 732 atoms (all atoms within 6 Å of the minimal QM system). The protein is a homodimer and two of the zinc-coordinating residues come from each subunit. The big-QM system included also the Zn site in the other subunit (but coordinated to two water molecules instead of the substrate). Besides the Zn ligands, it contained two cationic and 36 neutral residues, as well as 10 water molecules. The substrate had one positive and two negative charges outside the minimal QM model. This gave a net charge of +2 in the big-QM model, which is shown in Figure 2h. Seven states in the Richter and Krauss mechanism of the *S* substrate were studied, as is shown in Figure 2a−g.[37,38] The structures were taken from a QM/MM investigation of this enzyme, employing the small QM system.[31]

The third system was an active-site model of sulfite oxidase.[34] The minimal QM system contained a Mo ion, two O^{2-} ions, the SO_3^{2-} substrate, a Cys ligand, and the bidentate molybdopterin ligand, modeled as $(CH_3CS)_2^{2-}$. This 24-atom system had a net charge of −3 and is shown in Figure 3a−e. The corresponding big-QM system contained 805 atoms (6 Å surrounding), including 16 cationic, nine anionic, and 39 neutral groups, as well as six water molecules. It is shown in Figure 3f, and it had a net charge of +3. Five states were studied in the preferred S → OMo mechanism,[39] as is shown in Figure 3a−e. The Mo ion was always in the low-spin and closed-shell states. The structures were taken from a QM/MM investigation of this enzyme, using an intermediate-QM system of 165 atoms.[30]

QM Methods. Most QM calculations were performed with the Turbomole 7.1 software.[40,41] We employed 12 DFT methods available in this software, TPSS,[42] PBE,[43] BP,[44,45] B97D,[46] PBE0,[47] TPSSH,[48] B3LYP,[44,49,50] BHLYP,[51] M06,[52] M06L,[53] M06-2X,[52] and M06-HF.[54] In addition, calculations with the HF approach were also performed.[6,7] All calculations employed the def2-SV(P) basis set.[55] The calculations were sped up by expanding the Coulomb interactions in an auxiliary basis set, the resolution-of-identity (RI) approximation.[56,57]

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

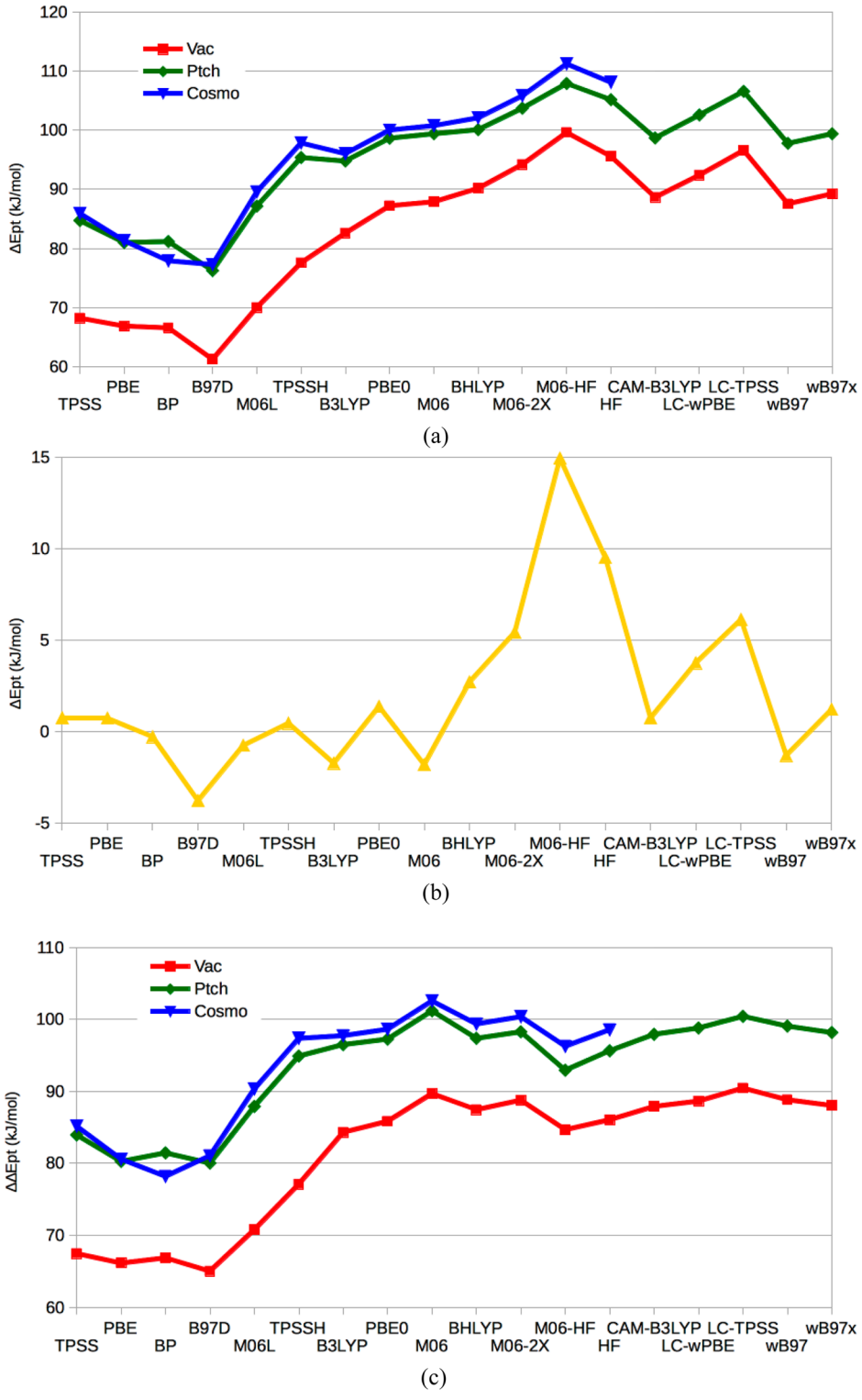


Figure 4. Results for the proton-transfer energy calculated with the three different embeddings and the 18 different QM methods. (a) Big-QM energies (ΔE_{pt}), (b) ΔE_{pt} for the minimal QM system, and (c) $\Delta\Delta E_{pt}$ energy difference between the big-QM and minimal QM system.

The calculations on the big-QM systems also employed the multipole-accelerated resolution-of-identity J approach (marij keyword). TPSS, PBE, BP, B97D, and M06L are pure functionals, with no admixture of HF exchange. The other DFT methods are hybrid functionals with 10% (TPSSH), 20% (B3LYP), 25% (PBE0), 27% (M06), 50% (BHLYP), 54% (M06-2X), and 100% (M06-HF) HF exchange. Interestingly, we did not note any major difference in the time consumption between the pure and hybrid DFT functionals with the Turbomole software. Instead, each big-QM calculation in this work could typically be finished within 2 (glyoxalase) to 15 h (hydrogenase) on a single computer core. In addition, calculations with five range-separated DFT methods, CAM-B3LYP,[58] LC-TPSS,[9] LC-ωPBE,[59] ωB97, and ωB97X,[60] were performed with the same basis set, using the Gaussian-09 software.[61] All QM calculations were single-point energy

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

calculations on QM/MM structures optimized at the BP/def2-SV(P) ([NiFe] hydrogenase) or TPSS/def2-SV(P) (glyoxalase I and sulfite oxidase) levels of theory.

Three different schemes were employed to describe the surroundings. In the first, the big-QM systems were simply studied in vacuum (Vac). In the second approach (Ptch), we included a point-charge model of all surrounding atoms in the protein, as well as a sphere of water molecules with radii of 50, 40, and 30 Å for [NiFe] hydrogenase, glyoxalase I, and sulfite oxidase, respectively. The charges were taken from the Amber 99SB force field.[62] All charges were included, except the carbon link atoms (i.e., the carbon atoms that were converted to hydrogen atoms in the QM calculations).[63] The number of point charges were 50797, 23655, and 10826 for the three systems, respectively. In the third type of calculations, the big-QM system was immersed into a continuum solvent, employing the conductor-like screening model (COSMO), implemented in Turbomole.[64,65] The default optimized the COSMO radii (and a water solvent radius of 1.3 Å),[66] whereas a radius of 2 Å was used for the metals.[67] The dielectric constant was 4, whereas we used default values for all other parameters. Since this method is not implemented in Gaussian-09, this embedding was not used for the range-separated functionals.

RESULT AND DISCUSSION

In this study, we have investigated how the SIE affects the results of big-QM calculations for three different enzyme systems, [NiFe] hydrogenase, glyoxalase I, and sulfite oxidase. The calculations employed QM systems with 675−805 atoms and 5−27 charged groups outside the minimal QM system. We employed HF and 12 different DFT methods with a varying amount of HF exchange (0−100%). We also used five range-separated DFT methods. The big-QM approach is intended as a postprocessing to improve the calculated energies. Therefore, all calculations in this article are single-point energy calculations on QM/MM structures obtained with smaller QM systems.[12,30,31]

A problem with such an investigation is that both reaction energies and the effect of the surroundings depend on the QM method. Therefore, differences between the various QM methods for big-QM calculations cannot directly be assigned to the SIE, but may also be caused by differences in correlation and other effects. Since we are interested primarily in the effect of the surroundings on the reaction energies, we performed calculations with all QM methods also on a minimal active-site model (24−48 atoms). By forming the difference in energy between the small and big-QM models ($\Delta\Delta E$), we separate the intrinsic effect of the QM method on the reactive site and the (mainly electrostatic) effect of the surroundings on the reaction energies.

In the following we will discuss the results for the three enzymes separately. For [NiFe] hydrogenase, we will make the most thorough comparison, including all 18 QM methods, performing calculations on 10 different structures of each state, and testing different treatments of the surroundings outside the big-QM system (no surroundings, a point-charge model, or a continuum solvent). For the other two systems, only 10 QM methods were employed for a single structure of each state, using a point-charge model of the surroundings.

It should be noted that all the big-QM calculations are performed with a rather small basis set (def2-SV(P)). Therefore, the reported energies should not be considered as final. They represent a rather realistic estimate of the influence of the surroundings on the energies. However, to obtain accurate reaction and activation energies, calculations with larger basis sets, as well as including corrections for dispersion, van der Waals interactions, entropy, and thermal effects are needed.[2,68] Such effects are discussed in the original literature documents.[12,30,31] Consequently, we will not discuss biochemical implications of the results in this work.

[NiFe] Hydrogenase. We have studied the transfer of a proton between a His residue and the Cys Ni ligand in the active site of [NiFe] hydrogenase. We calculate the energy difference (ΔE_{pt}) between the HIP state, in which the proton resides on the His residue and the HID state, in which it resides on the Cys residue (Figure 1a,b). Previous investigations have shown that this reaction energy is very sensitive to the surroundings.[12,26,35,63] It is close to thermoneutral for the isolated minimal QM system in Figure 1a,b, whereas it changes to over 110 kJ/mol if the 25 closest groups are included in the QM system, and it becomes ∼70 kJ/mol with a balanced account of the surroundings, e.g., by the big-QM approach, calculated at the BP/def2-SV(P) level.[26,35] In this study, we investigate how this reaction energy and its dependence on the surroundings are affected by the SIE, employing 18 different QM methods.

The first calculations were performed on a single structure, viz., the last snapshot from a MD simulation, followed by a QM/MM optimization for both states. Then, single-point big-QM energies were calculated for a 675-atoms system, running the calculations either in vacuum, with a point-charge model of the surrounding protein and solvent, or in a COSMO continuum solvent with a dielectric constant of 4. The results are gathered in Figure 4.

In general, the COSMO calculations gave the largest ΔE_{pt} energy difference between the HIP and HID states (HIP was always the most stable state) and the Vac calculations gave the smallest difference. The difference between the results from the COSMO and Ptch calculations was less than 3 kJ/mol, and for one method (BP), the Ptch result was actually larger. The difference between the Vac and Ptch calculations was somewhat larger, 8−18 kJ/mol. Moreover, this difference depended on the amount HF exchange in the DFT functional: The pure functionals gave a slightly larger difference, 14−17 kJ/mol, than the hybrid and range-separated functionals (and also HF), 8−12 kJ/mol, except for TPSSH, which gave the largest difference, 18 kJ/mol. The difference between the Vac and Ptch results comes partly from electrostatic interactions between the big-QM system and the surroundings, partly from the polarization of the big-QM system by the point charges, both of which are included in the Ptch calculations, but not in the Vac calculations. The latter effect is also partly included in the COSMO calculations, although estimated in a different way. It is the dominating effect, as can be confirmed by explicitly calculating the former term (2 kJ/mol difference between Vac and Ptch). In the following, we will concentrate our discussion mainly on the Ptch results, because the three embedding schemes gave similar trends.

Comparing the Ptch results of the 18 different QM methods, the ΔE_{pt} energy difference varied from 76 to 108 kJ/mol. ΔE_{pt} increased with the amount of HF exchange: It was 76−87 kJ/mol for the pure functionals, 95−108 kJ/mol for the hybrid functionals, 98−107 kJ/mol for the range-separated functionals, and 105 kJ/mol for HF. In fact, there was a good correlation between ΔE_{pt} and the amount of HF exchange for the seven hybrid functionals, R^2 = 0.9 in all three embeddings.

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

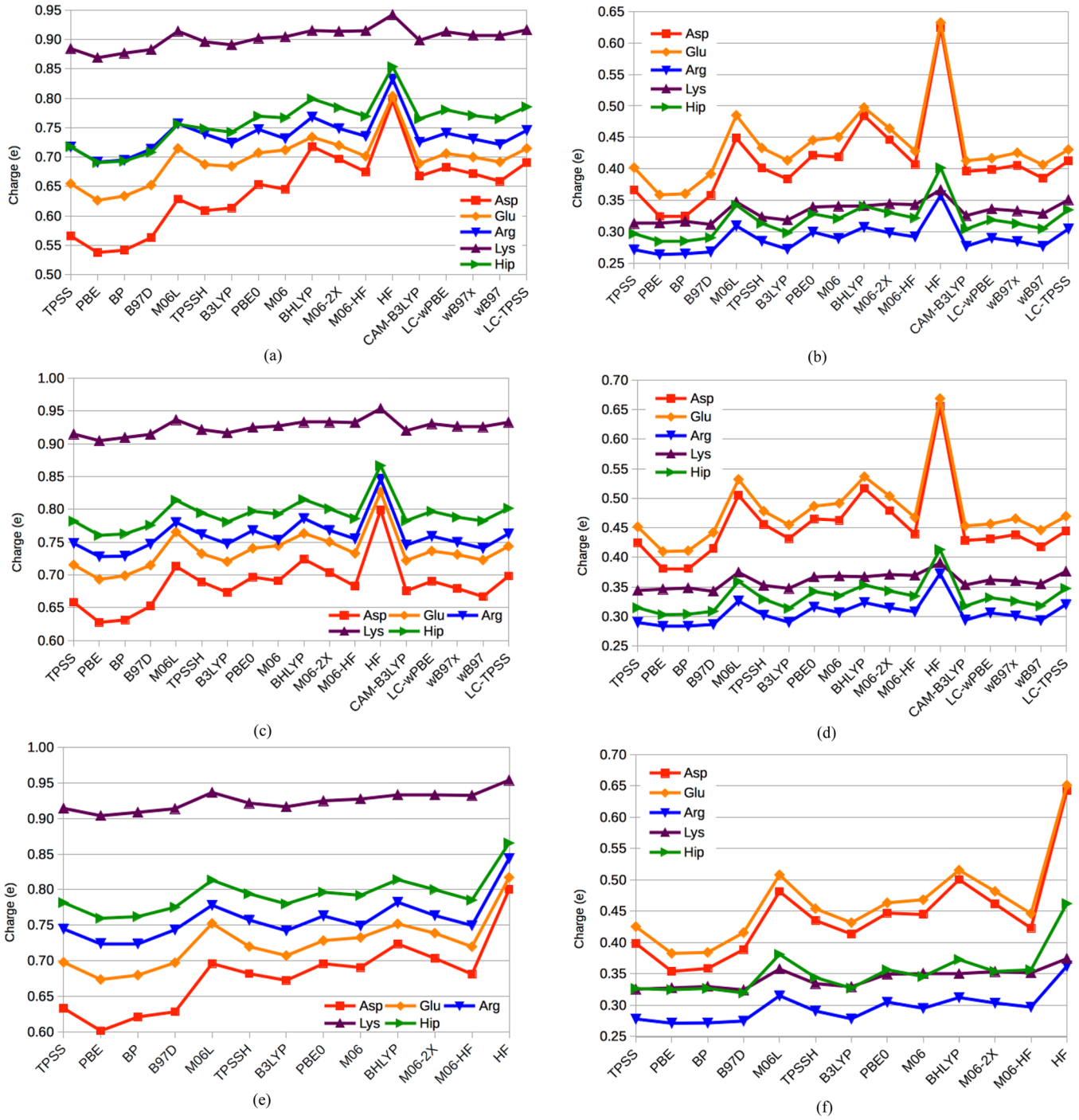


Figure 5. Variation of the Mulliken charges of the charged (a, c, e) residues or the O and H atom of the charged residues (b, d, f) in [NiFe] hydrogenase for the various methods, calculated in the Vac (a, b), Ptch (c, d), or COSMO (e, f) surroundings.

These results indicate the ΔE_{pt} depends on the DFT functionals and especially the amount of exchange, which may be related to the SIE. However, it is also possible that the differences reflect the intrinsic differences of the various methods for this particular proton-transfer reaction. Therefore, we also calculated the vacuum energy difference between the HIP and HID states for a minimal QM system, consisting of only 38 atoms (Figure 1a,b). The results are shown in Figure 4b, and they showed that there indeed were some intrinsic differences between the various QM methods, giving ΔE_{pt} energies ranging from −4 (B97D) to +15 kJ/mol (M06-HF). The pure functionals still gave lower values (−4 to +1 kJ/mol) than the hybrid (−2 to +15 kJ/mol) and the range-separated functionals (−1 to +6 kJ/mol), but the differences were less pronounced. There was a strong correlation between ΔE_{pt} and the amount of HF exchange in the hybrid functionals (R^2 = 0.9), but for the four methods with up to 27% no such correlation could be seen.

To emphasize the effect of the surroundings, we subtracted ΔE_{pt} of the minimal system from that of the big-QM system, calculated with the same method, giving $\Delta\Delta E_{pt}$, which represents the influence of the big-QM surroundings on the minimal QM system. These results are shown in Figure 4c. It can be seen that there still was a clear difference between the

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

results obtained with the pure functionals ($\Delta\Delta E_{pt}$ = 80−88 kJ/mol for the Ptch embedding) and with the hybrid or range-separated functionals (93−101 kJ/mol). However, there was no correlation between $\Delta\Delta E_{pt}$ and the amount of HF exchange for the hybrid functionals (R^2 < 0.2, often showing anticorrelation). Instead, all the range-separated functionals, as well as B3LYP, PBE0, BHLYP, and M06-2X, gave results that agreed within 3 kJ/mol (97−100 kJ/mol). TPSSH gave slightly lower results (95 kJ/mol), indicating some influence for the SIE. Interestingly, M06-HF gave even lower results (93 kJ/mol), whereas M06 gave the largest value of $\Delta\Delta E_{pt}$ (101 kJ/mol), indicating that these heavily parametrized methods give slightly deviating results. The HF method gave $\Delta\Delta E_{pt}$ = 96 kJ/mol.

Without the point-charge model (Vac), the TPSSH and B3LYP methods showed larger deviations from the other hybrid functionals, indicating that the point-charge models reduce somewhat the effect of the SIE. Thus, we can conclude that for the net effect of the surroundings ($\Delta\Delta E_{pt}$), all hybrid functionals gave results that agree within 8 kJ/mol (8%) and nine methods gave results that agreed within 3 kJ/mol (3%) and therefore can be considered as the consensus results for the influence of the surroundings on the proton-transfer energy. It is also clear that the pure functionals underestimate $\Delta\Delta E_{pt}$ by almost 20 kJ/mol.

Charges. To investigate whether this underestimation comes from the delocalization of charges, caused by the SIE,[8] we next examined the Mulliken charges in the various calculations. Mulliken charges often show a large basis-set dependence and do not ideally reproduce electrostatic properties, but when calculated with a split-valence basis set (as in this study) they provide an excellent account of chemical trends for related molecules.[67,69] We looked at the charges on the charged groups in the big-QM system, three Asp, six Glu, six Arg, one Lys, and three (HID state) or four (HIP state) doubly protonated His groups. We examined both the sum of the charges of all atoms in each residue and the average charge of the atoms bearing the main net charge in each residue (the two OD atoms in Asp, the two OE atoms in Glu, HE and the four HH atoms in Arg, the three HZ atoms in Lys, and the HD1 and HE2 atoms in His). To facilitate the discussion, we will in the following only discuss the absolute value of the charge (i.e., we will ignore the negative sign of the charges on Asp and Glu, as well as on the OD and OE atoms in these residues). The results of the calculations on the HID and HIP states were averaged because they were nearly identical (within 0.01 *e*) for all residues.

Indeed, there were clear trends in these charges, which were especially pronounced in the vacuum calculations, as can be seen in Figure 5a,b. For example, the charge on the Asp residues was lowest for the pure functionals (0.54−0.57 *e*, except M06L, which gave 0.63 *e*) and highest for HF (0.80 *e*). The hybrid and range-separated functionals gave intermediate values (0.61−0.72 *e* and 0.66−0.69 *e*, respectively. PBE gave always the smallest charges, but BP gave in most cases the same charge within 0.01 *e*. The largest charges were obtained with HF for all residues. BHLYP typically gave the largest charge for the DFT methods, except for the H atoms of the positively charged residues, for which M06L gave slightly larger charges. Consequently, there was only a weak (but always positive) correlation between the charges and the amount of HF exchange among the hybrid functionals, R^2 < 0.4.

The variation in the charges was largest for the Asp residues (0.26 *e* difference between the largest and smallest average charges obtained for the various methods) and for the corresponding OD atoms (0.30 *e*). It was 0.18 *e* for the Glu residues (but 0.27 *e* for the OE atoms). For the three positively charged residues the variation was 0.07−0.16 *e* and 0.06−0.12 *e* for the corresponding H atoms. Excluding the results of the HF method, the variation was 0.11−0.18 and 0.04−0.10 *e* for the negatively and positively charged residues and atoms, respectively.

In the point-charge surrounding, the variation of the charges was suppressed (shown in Figure 5c,d): The variation of the negative and positive charged residues was decreased to 0.13−0.17 and 0.05−0.12 *e* (0.07−0.10 and 0.03−0.07 *e* without HF), respectively, although the variation for the OD and OE atoms in Asp and Glu still was 0.26−0.27 *e* (0.13−0.14 *e* without HF). This decrease came mainly from an increase in the charges for the pure functionals (by up to 0.09 *e* for Asp), whereas the HF charges changed by less than 0.02 *e* for the residues (up to 0.04 *e* for the atoms). The results were similar (within 0.03 *e*), but often slightly smaller in the COSMO continuum solvent (Figure 5e,f). Thus, the point-charge model (and to a somewhat smaller extent a continuum solvent) is a good way to partly reduce the effect of the SIE, as also was seen for the energies.

We can conclude that there is a significant delocalization of the charges in the calculations with the pure functionals. However, the effect is appreciably smaller (∼0.1 *e* in a point-charge surrounding) than in the studies of the Gly_{10} zwitterion by Jakobsen et al., for which a transfer of half an electron was reported.[8] Moreover, there was a variation of up to 0.06 *e* for the charges of the charged residues (up to 0.1 *e* for the charged atoms) also for the nine DFT methods giving the same consensus energies for $\Delta\Delta E_{pt}$ discussed in the previous section, showing that there is no exact connection between the group charges and $\Delta\Delta E_{pt}$ energies.

Geometry Dependence. To check whether the energy results depend on the geometry of the studied system, we repeated the big-QM calculations on nine additional structures, started from different snapshots from a MD simulation of [NiFe] hydrogenase and then optimized by QM/MM before the big-QM calculation. To reduce the computational effort, the calculations were restricted to 10 QM methods that gave typical or interesting results in the first tests, viz., TPSS, PBE, M06L, B3LYP, M06, BHLYP, M06-2X, M06-HF, HF, and CAM-B3LYP. The calculations were still performed in the three environments (Vac, Ptch, and COSMO), and calculations on the corresponding minimal QM system were also performed in vacuum with all ten methods.

The results in Figure 6a show that the results obtained with the various methods follow the same trends for the various geometries. There was a pronounced variation in ΔE_{pt} for the various snapshots, ranging from 15 to 39 kJ/mol, depending on the QM method and the embedding scheme. It was lowest for the pure functionals and largest for HF, with a difference of 10−14 kJ/mol. It was also lower in the Vac calculations and largest in the Ptch calculations, with a difference of 8−14 kJ/mol. The variation was similar if instead $\Delta\Delta E_{pt}$ was considered (within 2 kJ/mol for the DFT methods and 5 kJ/mol for HF and M06L), showing that it comes mainly from the surroundings.

Consequently, we can consider the average over the 10 snapshots for the seven methods and the three embedding schemes. If this was done for ΔE_{pt}, there was still some dependence on the functionals and the amount of HF

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

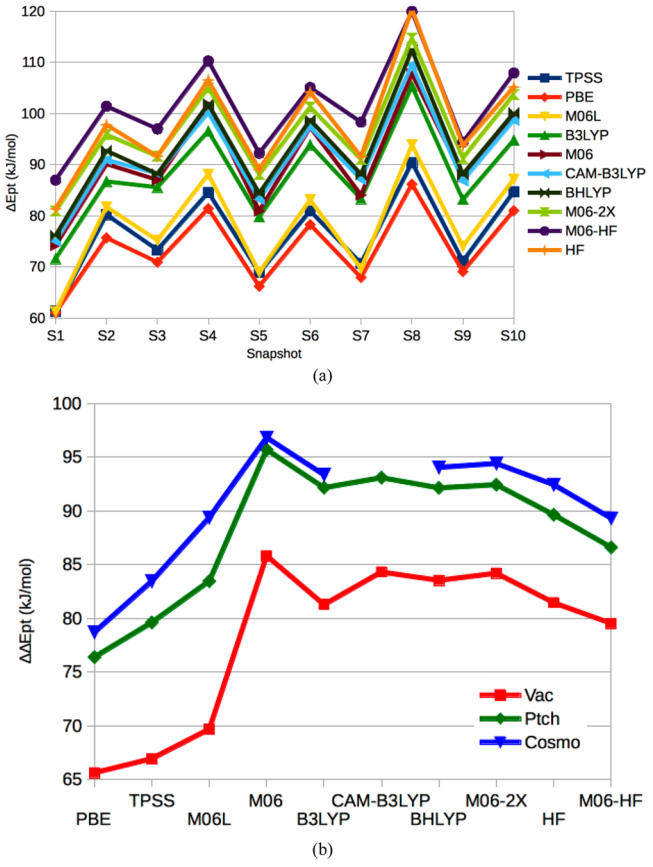


Figure 6. (a) ΔE_{pt} for the 10 snapshots from the MD simulations, calculated with 10 different QM methods for [NiFe] hydrogenase with the point-charge surrounding. (b) $\Delta\Delta E_{pt}$ for the average values over the 10 snapshots for the 10 QM methods in the three surroundings.

exchange, with CAM-B3LYP giving a results between M06 and BHLYP, and a difference of 28−31 kJ/mol between the M06-HF and PBE results. However, if we instead considered $\Delta\Delta E_{pt}$, the results were more even, as can be seen in Figure 6b. Now the B3LYP, CAM-B3LYP, BHLYP, and M06-2X methods all gave the same results within 1 kJ/mol, except B3LYP in vacuum (3 kJ/mol lower). The HF results were also similar, but 1−3 kJ/mol lower. The three pure functionals still gave 5−18 kJ/mol lower values of $\Delta\Delta E_{pt}$. M06 gave the largest and M06-HF the smallest $\Delta\Delta E_{pt}$ among the hybrid functionals.

Thus, we can conclude that there is a significant effect of the SIE for the big-QM $\Delta\Delta E_{pt}$ energies for [NiFe] hydrogenase, affecting energies calculated by pure functionals. However, all tested hybrid functionals gave results that agreed to within 6 kJ/mol, and four hybrid functionals gave results agreeing within 1 kJ/mol, at least when calculated in a continuum solvent or with a point-charge model.

Glyoxalase I. Next, we studied whether these results apply also to other systems by examining glyoxalase I with the same big-QM approach. We studied seven states in the Richter and Krauss mechanism of the enzyme (shown in Figure 2a−g):[33,37,38] The substrate bidentately bound to the active site (R); the enediolate intermediate (IM1), formed by the transfer of a proton from a C atom of the substrate to Glu172; the deprotonated product (IM2) formed by the transfer of the proton from Glu172 to the other C atom of the substrate, with the concurrent transfer of the hydroxyl proton to Glu99; and the product (P) formed by the transfer of the proton from Glu99 to the other hydroxyl oxygen atom as well as three transition states (TS1−TS3) between these states. The structures were optimized with QM/MM,[31] followed by single-point big-QM calculations using 732 atoms in the QM system (Figure 2h). The calculations were performed with the same 10 QM methods that were used for the hydrogenase snapshots, but with only the point-charge embedding.

Figure 7a shows the big-QM energies for the seven states in the reaction mechanism. It can be seen that the reaction and

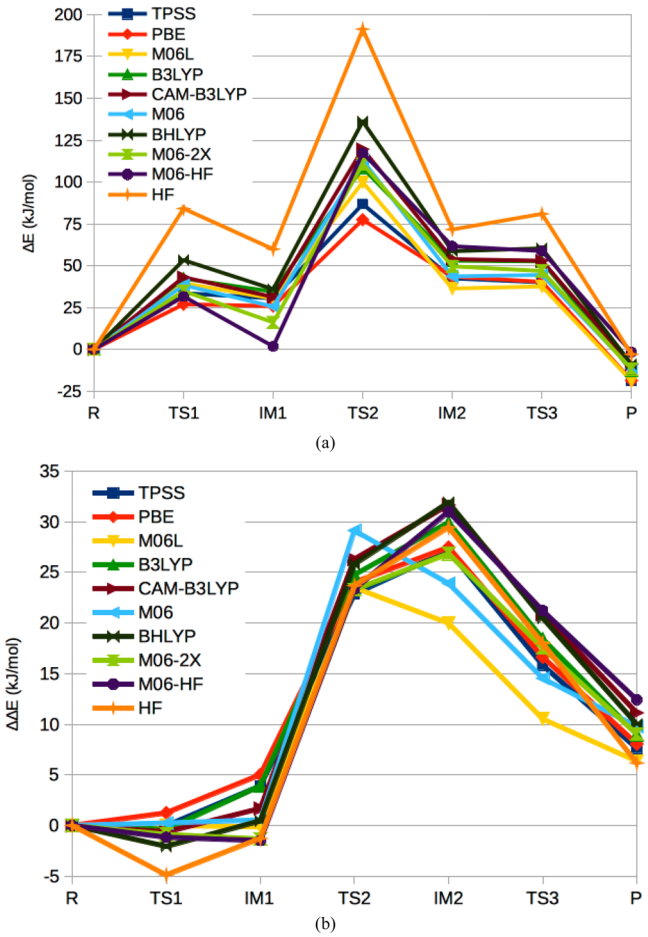


Figure 7. ΔE (a) and $\Delta\Delta E$ (b) for the seven states in the reaction mechanism of glyoxalase I, calculated with 10 different QM methods with a point-chare surrounding.

activation energies varied quite extensively among the seven QM methods (up to 114 kJ/mol), with HF giving the most deviating energies (the variation was up to 59 kJ/mol for the DFT methods). In general, the pure functionals (especially PBE) gave the lowest activation energies and BHLYP the highest barriers (besides HF). However, all methods agree that TS2 is the rate-limiting step for the mechanism with a barrier of 77−191 kJ/mol.

However, the same effects were seen also for the minimal (48-atom) QM system, so that if the $\Delta\Delta E$ energy differences were considered, the effect of the surroundings was actually rather small, only −5 to 32 kJ/mol, as can be seen in Figure 7b. The effect was minimal for the first two states (within 6 kJ/mol for all QM methods), and it was largest for the TS2, IM2, and TS3 states. As for the hydrogenase reaction, the PBE and TPSS functionals gave nearly identical results (within 1 kJ/mol for all states). The CAM-B3LYP, BHLYP, and M06-HF methods also

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

gave results that agreed within 3 kJ/mol, but they differed from that of the pure functionals by up to 5 kJ/mol (3 kJ/mol on average). B3LYP gave results that were intermediate between the latter three functionals (maximum and average differences of 4 and 2 kJ/mol) and the two pure functionals. M06-2X gave results that were identical to that of M06-HF for the first four states in the mechanism, but it showed differences of up to 5 kJ/mol toward the results of the pure functionals for the last three states. HF also gave somewhat varying results that differed from the CAM-B3LYP, BHLYP, and M06-HF results by up to 5 kJ/mol (3 kJ/mol on average). M06 and M06L showed the largest differences from the latter three methods by up to 8 and 12 kJ/mol, especially for the TS2, IM2, and TS3 states. If these methods are excluded, actually all the other QM methods gave results that agreed within 6 kJ/mol for all seven states in the mechanism.

In conclusion, this glyoxalase I test case illustrates effects expected in a simple system with a Zn ion in the active site and only rather small effects from the surrounding protein. For such a system, all methods gave $\Delta\Delta E$ energies within 12 kJ/mol (6 kJ/mol if M06 and M06L are excluded). However, we could still see a significant effect of the SIE, leading the pure PBE and TPSS methods to give somewhat diverging results, but by only 3 kJ/mol on average. CAM-B3LYP, BHLYP, and M06-HF gave results that agreed within 1 kJ/mol on average, whereas those of B3LYP and M06-2X differed by 2 kJ/mol on average.

Sulfite Oxidase. Finally, we studied also sulfite oxidase. We examined the S → OMo mechanism[34,39] in which the sulfur atom of the sulfite substrate attacks one of the Mo-bound oxo groups, directly forming a Mo-bound sulfate ion (IM) via a first transition state (TS1). In the second step, the sulfate ion dissociates, via a second transition state (TS2). Thus, five states were studied with the big-QM approach, including a reactant complex (R) and a product complex (P) with either sulfite or sulfate binding in the second sphere of the Mo ion; cf. Figure 2a–e. All structures were optimized with QM/MM before the big-QM energies were calculated, using an 805-atom system.[30] The reaction and activation energies were compared to energies calculated with a minimal QM system, consisting of 24 atoms. The same 10 QM methods as for glyoxalase I were employed, still with the point-charge embedding.

The raw big-QM energies showed large variation between the various QM methods (up to 360 kJ/mol), except for the PBE and TPSS pure functionals, as can be seen in Figure 8a. As usual, the HF gave the most differing energies, with the largest difference toward the pure functionals. In fact, there was a nearly perfect anticorrelation between the energies and the amount of HF exchange for the hybrid functionals ($R^2 > 0.9$ for all states).

However, this variation was partly paralleled in the calculations for the minimal QM system. Therefore, the $\Delta\Delta E$ energies were more stable, showing trends similar to those for hydrogenase and glyoxalase I, although the energies were much larger, as can be seen in Figure 8b. In particular, CAM-B3LYP, BHLYP, and M06-2X gave results that agreed within 7 kJ/mol (always with BHLYP in the middle, within 1 kJ/mol of the average). The results of the pure functionals were 32–37 kJ/mol smaller on average. B3LYP and M06 gave intermediate results, with a difference from the three hybrid functionals of 15–22 kJ/mol on average. HF and especially M06-HF gave the largest values, with average differences of 28 and 36 kJ/mol from the three hybrid functionals.

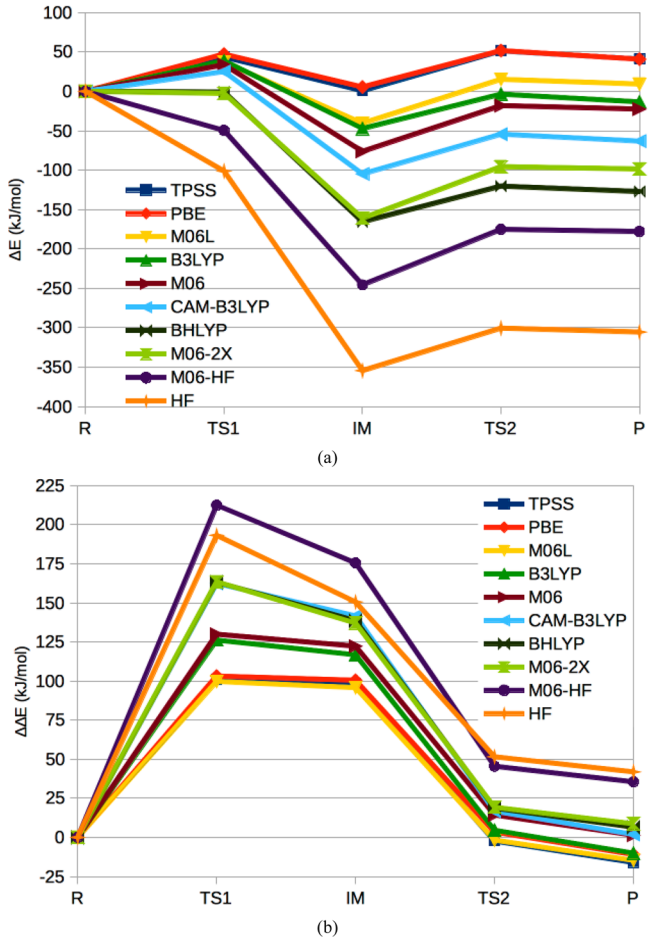


Figure 8. ΔE (a) and $\Delta\Delta E$ (b) for the five states in the reaction mechanism of sulfite oxidase, calculated with 10 different QM methods with a point-charge surrounding.

To check that this enhanced effect of the surroundings did not come from increased differences in the charges obtained for the various QM methods, we calculated the Mulliken charges for the charged residues also in this protein. However, the results showed that the variation in the charges was similar to what was found for [NiFe] hydrogenase: It was 0.10–0.14 *e* for the Asp and Glu residues, 0.25–0.26 *e* for the OD and OE atoms in these residues, and 0.08–0.12 *e* for the Arg, Lys, and His residues, as well as the H atoms in these residues (and approximately half of this variation if the HF method was excluded).

Thus, we can conclude that for the sulfite reductase reaction, which is extremely sensitive to both the surroundings and the correlation effects, the effect of the SIE for the big-QM energies is enhanced. In particular, 27% HF exchange (as in M06) is not enough to avoid the SIE. Instead, 50% HF exchange or a range-separated functional seems to be needed. Fortunately, BHLYP, M06-2X, and CAM-B3LYP all gave results that agreed within 2 kJ/mol on average.

CONCLUSIONS

In this work we have examined how the SIE affects reaction and activation energies calculated by the big-QM approach, i.e., for QM calculations involving all atoms within 4.5–6.0 Å of a minimal active-site model, as well as all buried charges in the protein, and with no border between the QM and MM systems

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667–5679

within three residues from a minimal active-site model.[12,26] Such models typically involve 600−1000 atoms, including several groups with a net charge. Recently, Jakobsen and co-workers pointed out that the SIE in pure DFT methods may lead to a delocalization of such charges by as much as 0.5 *e*, which would give rise to an incorrect charge distribution and electrostatic potential.[8] The problem could be avoided by employing hybrid DFT methods with a high amount of HF exchange or range-separated functionals. Of course, this could potentially be a disaster for both big-QM and similar approaches in obtaining energies and electrostatic or spectroscopic properties from QM calculations with large QM systems.[15−23,27]

Therefore, we have in this study investigated how reaction and activation energies calculated with the big-QM approach for three typical enzyme systems change when the QM method is varied, employing 17 different DFT methods, including pure, hybrid, and range-separated functionals, as well as the HF approach, and also using three different embedding schemes. By looking at the Mulliken charges of charged groups in the large QM systems, we confirm that there is a significant delocalization of charges for DFT methods with no or little HF exchange. However, the effect was smaller than in the studies of Jakobsen et al.:[8] The variation in net charge of the charged group was 0.07−0.26 *e* among the tested methods and less than 0.18 *e* if HF was excluded, which is known to overestimate charges and dipole moments.[70] The variation could be further suppressed by performing the calculations with a point-charge model of the surroundings (to less than 0.17 and 0.10 *e* with and without HF).

For energies, the comparison is complicated by the fact that different QM methods give varying results for reaction and activation energies. We have solved this problem by looking mainly at the energy difference between calculations on the big-QM system and a minimal model of the QM system ($\Delta\Delta E$), representing the effect from the surroundings, which can be expected to depend less on the QM method (besides the SIE effect on the charges). We have shown that there is a systematic variation of $\Delta\Delta E$ for all three enzymes studied. Fortunately, a number of DFT methods could be identified that gave nearly identical results. These always included BHLYP with 50% HF exchange and the range-separated CAM-B3LYP method. M06-2X also belonged to this group, although for glyoxalase I, it showed a somewhat larger difference (up to 5 kJ/mol). For the two enzymes with modest effects of the surroundings, B3LYP, PBE0, and M06-HF also gave similar results, at least with a point-charge or COSMO surrounding, but for the third enzyme with very large effects from both correlation and from the surroundings, these methods gave diverging results. The other methods gave less reliable results: The pure functionals typically gave a too small effect from the surroundings, probably reflecting the underestimated charges. Conversely, HF typically gave a too large effect, reflecting the overestimated charges. The M06 and M06L methods often gave results diverging from the trends obtained with the other methods.

The absolute effect of the SIE on the big-QM energies differed between the three studied systems, owing to differences in the number of charged groups and the sensitivity to the surroundings. For glyoxalase I, the influence of the surrounding was quite small, less than 32 kJ/mol. Therefore, all methods gave $\Delta\Delta E$ energies that agreed within 12 kJ/mol and actually within 6 kJ/mol if the M06 and M06L methods were excluded. For this enzyme, the choice of the method for the big-QM calculations is not very crucial. For [NiFe] hydrogenase, the influence of the surroundings was quite large, 76−96 kJ/mol, and therefore, the variation in $\Delta\Delta E$ among the various methods was larger, 21−25 kJ/mol. Pure functionals, clearly gave too low values by ∼15 kJ/mol. However, if pure functionals TPSSH (with only 10% HF exchange) and M06 were excluded, the remaining hybrid and range-separated functionals gave results that agreed within 4−7 kJ/mol. For sulfite oxidase, the effect of the surroundings (up to 212 kJ/mol) and of correlation effects were extremely large with variations of up to 113 kJ/mol among the various methods. Therefore, only three methods gave results that agreed within 2 kJ/mol on average (CAM-B3LYP, BHLYP, and M06-2X). For the other methods, the differences were large, 15−37 kJ/mol on average, with the pure functionals, but also M06-HF giving the largest differences. However, it should be noted that glyoxalase I is probably closer to a general enzyme—most enzymes have no or only a few buried charges (and then often close to the active site).

Consequently, for the estimation of the effect of the surroundings on reaction energies using large QM systems, we tend to recommend the CAM-B3LYP, BHLYP, and M06-2X methods. The calculations should be performed in a point-charge surrounding, which decreases the charge delocalization and also improves the convergence of the calculations.[12] However, such functionals typically do not give accurate reaction and activation energies for metal-containing systems; instead hybrid DFT functionals with a smaller amount of HF exchange (10−15%) typically give the best results.[1,2,71,72] Therefore, it is normally necessary to estimate the intrinsic reaction energy with one DFT functional (or a more accurate wave function method) on a small QM system, and then estimate the effect of the surroundings with calculations using one of the CAM-B3LYP, BHLYP, or M06-2X functionals on both the small and big QM systems, estimating $\Delta\Delta E$ energy differences, as in this work. Naturally, dispersion effects should be included in the big-QM energies, as has been done in our previous studies,[12,25,26,28−31] and the intrinsic reaction energies should include proper corrections for large basis set, relativistic, zero-point, thermal, and entropy effects.[2,68]

AUTHOR INFORMATION

Corresponding Author

*E-mail: Ulf.Ryde@teokem.lu.se. Tel.: +46−46 2224502. Fax: +46−46 2228648.

Funding

This investigation has been supported by grants from the Swedish Research Council (Project 2014-5540) and from the Knut and Alice Wallenberg Foundation (Grant KAW 2013.0022).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The computations were performed on computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc at Lund University.

REFERENCES

(1) Neese, F. *JBIC, J. Biol. Inorg. Chem.* **2006**, *11*, 702−711.
(2) Blomberg, M. R. A.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. M. *Chem. Rev.* **2014**, *114*, 3601−3658.

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

(3) Perdew, J. P.; Zunger, A. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1981**, *23*, 5048−5079.
(4) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
(5) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. *J. Chem. Phys.* **2005**, *123*, 062201.
(6) Hartree, D. R. *Math. Proc. Cambridge Philos. Soc.* **1928**, *24*, 89−132.
(7) Fock, V. *Eur. Phys. J. A* **1930**, *62*, 795−805.
(8) Jakobsen, S.; Kristensen, K.; Jensen, F. *J. Chem. Theory Comput.* **2013**, *9*, 3978−3985.
(9) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 3540−3544.
(10) Lehtola, S.; Jónsson, H. *J. Chem. Theory Comput.* **2014**, *10*, 5324−5337.
(11) Klüpfel, S.; Klüpfel, P.; Jónsson, H. *J. Chem. Phys.* **2012**, *137*, 124102.
(12) Hu, L.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2013**, *9*, 640−649.
(13) Kussmann, J.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2015**, *11*, 918−922.
(14) Dziedzic, J.; Fox, S. J.; Fox, T.; Tautermann, C. S.; Skylaris, C. K. *Int. J. Quantum Chem.* **2013**, *113*, 771−785.
(15) Fox, S.; Wallnoefer, H. G.; Fox, T.; Tautermann, C. S.; Skylaris, C. K. *J. Chem. Theory Comput.* **2011**, *7*, 1102−1108.
(16) Cave-Ayland, C.; Skylaris, C. K.; Essex, J. W. *J. Phys. Chem. B* **2015**, *119*, 1017−1025.
(17) Lee, L. P.; Cole, D. J.; Skylaris, C.; Jorgensen, W. L.; Payne, M. C. *J. Chem. Theory Comput.* **2013**, *9*, 2981−2991.
(18) Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O. J.; Essex, J. W.; Skylaris, C.-K. *J. Phys. Chem. B* **2013**, *117*, 9478−9485.
(19) Brown, S. P.; Schaller, T.; Seelbach, U. P.; Koziol, F.; Ochsenfeld, C.; Klärner, F.-G.; Spiess, H. W. *Angew. Chem., Int. Ed.* **2001**, *40*, 717−720.
(20) Sumowski, C. V.; Ochsenfeld, C. *J. Phys. Chem. A* **2009**, *113*, 11734−11741.
(21) Flaig, D.; Beer, M.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2012**, *8*, 2260−2271.
(22) Sumowski, C. V.; Schmitt, B. B. T.; Schweizer, S.; Ochsenfeld, C. *Angew. Chem., Int. Ed.* **2010**, *49*, 9951−9955.
(23) Sulpizi, M.; Raugei, S.; VandeVondele, J.; Carloni, P.; Sprik, M. *J. Phys. Chem. B* **2007**, *111*, 3969−3976.
(24) Genheden, S.; Ryde, U.; Söderhjelm, P. *J. Comput. Chem.* **2015**, *36*, 2114−2124.
(25) Dong, G.; Ryde, U. *JBIC, J. Biol. Inorg. Chem.* **2016**, *21*, 383−394.
(26) Sumner, S.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2013**, *9*, 4205−4214.
(27) Liao, R.-Z.; Thiel, W. *J. Comput. Chem.* **2013**, *34*, 2389−2397.
(28) Delcey, M. G.; Pierloot, K.; Phung, Q. M.; Vancoillie, S.; Lindh, R.; Ryde, U. *Phys. Chem. Chem. Phys.* **2014**, *16*, 7927−7938.
(29) Hedegård, E. D.; Kongsted, J.; Ryde, U. *Angew. Chem., Int. Ed.* **2015**, *54*, 6246−6250.
(30) Caldararu, O.; Feldt, M.; Cioloboc, D.; Ryde, U. *J. Biol. Inorg. Chem.* **2017**, submitted for publication.
(31) Dong, G.; Sadat Alavi, F.; Fouda, A.; Irani, M.; Ryde, U. *J. Biol. Chem.* **2017**, submitted for publication.
(32) Lubitz, W.; Ogata, H.; Rüdiger, O.; Reijerse, E. *Chem. Rev.* **2014**, *114*, 4081−4148.
(33) Creighton, D. J.; Hamilton, D. S. *Arch. Biochem. Biophys.* **2001**, *387*, 1−10.
(34) Kappler, U.; Enemark, J. H. *JBIC, J. Biol. Inorg. Chem.* **2015**, *20*, 253−264.
(35) Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. *J. Phys. Chem. A* **2009**, *113*, 11793−11800.
(36) Ogata, H.; Krämer, T.; Wang, H.; Schilter, D.; Pelmenschikov, V.; van Gastel, M.; Neese, F.; Rauchfuss, T. B.; Gee, L. B.; Scott, A. D.; Yoda, Y.; Tanaka, Y.; Lubitz, W.; Cramer, S. P. *Nat. Commun.* **2015**, *6*, 7890.
(37) Jafari, S.; Ryde, U.; Irani, M. *J. Mol. Catal. B: Enzym.* **2016**, *131*, 18−30.
(38) Richter, U.; Krauss, M. *J. Am. Chem. Soc.* **2001**, *123*, 6973−6982.
(39) van Severen, M. C.; Andrejić, M.; Li, J. L.; Starke, K.; Mata, R. A.; Nordlander, E.; Ryde, U. *JBIC, J. Biol. Inorg. Chem.* **2014**, *19*, 1165−1179.
(40) *TURBOMOLE*, version 7.1; University of Karlsruhe and Forschungszentrum Karlsruhe GmbH: Karlsruhe, Germany, 1989.
(41) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 91−100.
(42) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(43) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(44) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.
(45) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822−8824.
(46) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(47) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982.
(48) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129−12137.
(49) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(50) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(51) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
(52) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(53) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.
(54) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126−13130.
(55) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(56) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119−124.
(57) Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283−289.
(58) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(59) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 234109.
(60) Chai, J.-D.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106.
(61) 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, J. M.; 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*, Revision E.01; Gaussian: Wallingford, CT, USA, 2009.
(62) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(63) Hu, L.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2011**, *7*, 761−777.
(64) Klamt, A.; Schüütmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.
(65) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187−2193.

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679

(66) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.
(67) Sigfridsson, E.; Ryde, U. *J. Comput. Chem.* **1998**, *19*, 377−395.
(68) Ryde, U. *Methods Enzymol.* **2016**, *577*, 119−158.
(69) Cramer, C. J. *Essentials of Computational Chemistry*; John Wiley & Sons: Chichester, U.K., 2002.
(70) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27−85.
(71) Reiher, M.; Salomon, O.; Hess, B. A. *Theor. Chem. Acc.* **2001**, *107*, 48−55.
(72) Jensen, K. P. *J. Phys. Chem. A* **2009**, *113*, 10133−10141.

DOI: 10.1021/acs.jctc.6b00903
J. Chem. Theory Comput. 2016, 12, 5667−5679