



# LUND UNIVERSITY

## Effect of geometry optimisations on QM-cluster and QM/MM studies of reaction energies in proteins

Sumner, Sophie; Söderhjelm, Pär; Ryde, Ulf

*Published in:*  
Journal of Chemical Theory and Computation

*DOI:*  
[10.1021/ct400339c](https://doi.org/10.1021/ct400339c)

2013

[Link to publication](#)

*Citation for published version (APA):*  
Sumner, S., Söderhjelm, P., & Ryde, U. (2013). Effect of geometry optimisations on QM-cluster and QM/MM studies of reaction energies in proteins. *Journal of Chemical Theory and Computation*, 9(9), 4205-4214.  
<https://doi.org/10.1021/ct400339c>

*Total number of authors:*  
3

### General rights

Unless other specific re-use rights are stated the following general rights apply:  
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00



# **Effect of geometry optimisations on QM-cluster and QM/MM studies of reaction energies in proteins**

**Sophie Sumner, Pär Söderhjelm, Ulf Ryde \***

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

Correspondence to Ulf Ryde, E-mail: [Ulf.Ryde@teokem.lu.se](mailto:Ulf.Ryde@teokem.lu.se),

Tel: +46 – 46 2224502, Fax: +46 – 46 2228648

2013-04-25

## Abstract

We have examined the effect of geometry optimisation on energies calculated with the quantum-mechanical (QM) cluster, the combined QM and molecular-mechanics (QM/MM), the big-QM approaches (very large single-point QM calculations taken from QM/MM-optimised structures, including all atoms within 4.5 Å of the minimal active site, all buried charged groups in the protein, and truncations moved at least three residues away from the active site). We study a simple proton-transfer reaction between His-79 and Cys-546 in the active site of [Ni,Fe] hydrogenase and optimise QM systems of 50 different sizes (56–362 atoms). Geometries optimised with QM/MM are stable and reliable, whereas QM-cluster optimisations give larger changes in the structures and sometimes lead to large distortions in the active site if some hydrogen-bond partners to the metal ligands are omitted. Keeping 2–3 atoms for each truncated residue (rather than one) fixed in the optimisation improves the results, but does not solve all problems for the QM-cluster optimisations. QM-cluster energies in vacuum and a continuum solvent are insensitive to the geometry optimisations with a mean absolute change upon the optimisations of only 4–7 kJ/mol. This shows that geometry optimisations do not improve the convergence of QM-cluster calculations – there is still a ~60 kJ/mol difference between calculations in which groups have been added to the QM system according to their distance to the active site or based on QM/MM free-energy components. QM/MM energies do not show such a difference, but they converge rather slowly with respect to the size of the QM system, although the convergence is improved by moving away truncations from the active site. The big-QM energies are stable over the 50 different optimised structures,  $57 \pm 1$  kJ/mol, although some smaller trends can be discerned. This shows that both QM-cluster geometries and energies should be interpreted with caution. Instead, we recommend QM/MM for geometry optimisations and energies calculated by the big-QM approach.

**Key Words:** Quantum mechanical cluster calculations, QM/MM, geometry optimisation, density-functional theory, [Ni,Fe] hydrogenase.

## Introduction

During the latest two decades, quantum mechanical (QM) calculations have been established as an important and powerful complement to experiments for the study of biochemical reactions [1,2,3,4,5,6,7,8]. In particular, it has been shown that QM calculations can give structures of active sites with an accuracy that is better than what is obtained in low- and medium-resolution crystal structures [9], and they can provide structures of transition states of putative enzyme reactions, which are hard to study experimentally. The calculations also give energies of the transition states and intermediates, which can be used to compare different putative reaction mechanisms.

Two approaches are used for such calculations. In the QM cluster approach, the most important residues of the active site (typically 50–200 atoms) are cut out of the enzyme [1,2,3,4]. The geometries are normally optimised with a split-valence basis set and energies are calculated on these geometries using polarised triple-zeta basis sets. Typically, hybrid density-functional QM methods are employed and recently the importance of dispersion corrections has been emphasized [3,10,11,12]. Zero-point energies are estimated from vibrational frequencies and sometimes also entropic and thermal effects. To account for the excluded remainder of the protein, the QM cluster is immersed in a polarisable continuum-solvation model with a dielectric constant of  $\sim 4$ . Although this choice is quite arbitrary, it has been shown that as the size of the QM system is increased to 150–200 atoms, the results obtained with different values of the dielectric constant typically converge [3,13,14,15,16]. Moreover, some atoms are often fixed in the geometry optimisations to ensure that the geometry remains close to that in protein crystal structures. On the other hand, this means that entropies can no longer be obtained [3].

The other approach is to use QM for a central active-site model of a similar size (50–200 atoms) but include the rest of the enzyme, as well as a number of explicit water molecules, at the molecular mechanics (MM) level of theory. In this QM/MM approach, QM and MM energies and forces are added, ensuring that no terms are double-counted [5,6,7,8]. Several different approaches are available to treat the electrostatic interactions between the QM and MM systems and the positions where covalent bonds in the MM system are broken to form the truncated QM system (junctions) [6,7]. In mechanical-embedding QM/MM, the electrostatic interactions between the QM and MM systems are treated purely by MM, so that the QM calculation is performed in vacuum, whereas in electrostatic-embedding QM/MM, the electrostatic interactions are treated in the QM calculation, including a point-charge model of the surroundings.

Naturally, the two approaches have different strengths. The advantage of the QM-cluster approach is the small size of the considered systems, allowing for accurate QM methods and full control of the movement of the atoms. The disadvantage is that most of the enzyme is ignored, making it hard to judge the effect of the surroundings and leading to the risk that important residues are left out in the calculations. The latter problem is solved by the QM/MM approach, in which all groups are explicitly included in the calculations. On the other hand, this makes the size of the modelled system so large that it becomes hard to ensure that all states along a reaction path reside in the same local minimum. This can be addressed by running the optimisations forth and back several times along the reaction path, by studying many structures of the surroundings, e.g. taken as snapshots from molecular dynamics simulations [5,7,17], or by calculating free energies [7,18,19], rather than pure energies from minimised structures.

A problem with both approaches is that reaction energies often depend on the size of the QM system. Convergence studies by Himo and coworkers have shown that the energies are often stable. For example, they observe a maximum difference of 12 kJ/mol for activation energies calculated with cluster models of 83, 112, and 161 atoms in haloalcohol dehalogenase [14]. However, in other cases the variation is larger [15], e.g. up to 56 kJ/mol between a minimal 27-atom model and a 220-atom model for aspartate decarboxylase (and

still 27 kJ/mol difference for a 135-atom model) [20] or 38 kJ/mol between models of 77 and 177 atoms in 4-oxalocrotonate tautomerase [13]. This shows that the QM-cluster approach requires careful selection of the QM system and that results obtained with less than ~150 atoms should be considered with caution.

Even worse, we have shown that QM-cluster energies can differ by over 60 kJ/mol even for QM systems of over 400 atoms, depending on whether atoms have been selected for the QM system with respect to their distance to the active site or with respect to their energetic influence in a QM/MM free-energy perturbation technique [16]. Moreover, Sumowski and Ochsenfeld have shown that QM-cluster energies can differ by 45 kJ/mol between calculations with 300 and 1637 atoms in the QM system [21]. Both these studies were performed without changing the geometry of the QM systems in order to avoid effects of the local minima problem. However, it has been suggested that the lack of geometry optimisation will artificially slow down the convergence of the energies with respect to the size of the QM system and that these results therefore are not relevant for the standard QM-cluster approach [3,20].

Even if the QM/MM approach includes all atoms of the enzyme, there is still a division between atoms treated by QM or by MM. Intuitively, the effect of converting a group from QM to MM is expected to be less drastic than converting it from QM to nothing or to a continuum solvent. Unfortunately, this is not necessarily the case. In a direct comparison of convergence of QM and QM/MM methods with respect to the size of the QM system, we showed that standard (electrostatic-embedding) QM/MM showed only a slightly better mean absolute error than a pure QM calculation (17 compared to 21 kJ/mol) and actually a slightly larger maximum error (57 compared to 56 kJ/mol) over 40 systems with 46–446 atoms in the QM system [22]. The problem seems to be the junctions and electrostatic interactions between the QM and MM systems. However, with a carefully designed mechanical-embedded QM/MM method the mean absolute error and maximum errors could be decreased to 7 and 28 kJ/mol, respectively. Also in these studies, all geometries were kept fixed. Eriksson and coworkers have shown a variation of QM/MM reaction and activation energies of up to 40 kJ/mol with QM systems of 55–96 atoms (with optimised geometries) [23]. Ochsenfeld and coworkers have also studied the convergence of QM/MM calculations (for both energies and NMR shifts) with respect to the size of the QM and suggested that systems of 200–1000 atoms are needed for convergence [21,24,25]. They have also concluded that the convergence of QM/MM is typically better than of QM-cluster calculations.

Based on these and related studies [26,27,28], we have recently suggested a method (called big-QM in the following) that combines features of the two approaches [17]. It employs the QM/MM method to optimise geometries, using a small QM system. Then, accurate energies are calculated with a very large QM system, consisting of all chemical groups within 4.5 Å of the minimal QM system, as well as all buried charged groups in the protein, and with junctions moved three residues away from the minimal QM system. This gives QM systems of 600–1000 atoms, which can be treated with standard QM software. With such a size of the QM system, the treatment of the surroundings becomes less important, and we showed that three approaches (a continuum solvent, mechanical embedding, and electrostatic embedding) gave results that agreed within 13 kJ/mol for a reaction that preserves the net charge of the active site.

However, such an approach still depends on the size of the QM system used in the geometry optimisation. Recently, Liao and Thiel have observed changes of up to 105 kJ/mol in QM/MM energies when the geometries are optimised with different sizes of the QM system [29]. In this paper, we examine in a systematic way how QM-cluster, QM/MM, and big-QM energies vary when the geometry is optimised either by the QM-cluster or QM/MM approach using different sizes of the QM system. The results show that geometry optimisations do not solve the convergence problem of QM-cluster energies, but that big-QM energies are reasonably independent on the QM systems used in the QM geometry optimisations. In

addition, we also show that the QM-cluster method is much more sensitive to the size of the QM system used in the geometry optimisation than QM/MM methods and that different approaches to fix atoms give different results. However, quite satisfactorily, for QM systems of >200 atoms QM-cluster and QM/MM geometry optimisations give similar results.

## Methods

### *The protein*

The calculations were started from our previous QM/MM structure of [Ni,Fe] hydrogenase [16,26,30]. It is based on the 1.81-Å crystal structure of the Ser499Ala mutant from *Desulfovibrio fructosovorans* [31]. Residues more than 27 Å from the active-site Ni ion were deleted and solvation water molecules were added to the protein, forming a sphere with radius 33 Å. This gave 602 protein residues, 1045 water molecules, and a total of 12178 atoms. All Lys, Arg, Asp, and Glu residues were assumed to be charged, except Glu-25, which shares a hydrogen atom with the Ni ligand Cys-543, and Glu-S16 (an initial S refer to the small subunit, whereas residue numbers without S refer to the large subunit), which is involved in the proton-transfer path from Cys-543 to the protein surface. Cys ligands coordinating to metals were deprotonated. Among the His residues, His-S92, S160, 481, and 549 were protonated on the ND1 atom, His-S13, 27, 66, 113, 118, 121, 123, 210, 228, 349, 367, and 419 were protonated on the NE2 atom, and His-79, 115, 204, 305, and 538 were protonated on both these atoms. The added protons and water molecules were equilibrated, whereas the heavy atoms were kept at the crystal structure [30].

### *QM/MM calculations*

The QM/MM calculations were performed with the ComQUM software [32,33]. In this approach, the protein and solvent are split into two subsystems: System 1 was the QM system and it was relaxed by QM methods. System 2 consisted of the remaining part of the protein and the water molecules. It was kept fixed at the original coordinates (the crystal structure for the heavy atoms).

In the QM calculations, system 1 was represented by a wavefunction, whereas all the other atoms were represented by an array of partial point charges, one for each atom, taken from MM libraries (i.e. electrostatic embedding). When there is a bond between systems 1 and 2 (a junction) the hydrogen link-atom approach was employed: The QM system was capped with hydrogen atoms (H link atoms, HL), the positions of which are linearly related to the corresponding carbon atoms (C link atoms, CL) in the full system [32,34]. All atoms were included in the point-charge model, except the CL atoms [22].

The total QM/MM energy in ComQUM is calculated as [32,33]:

$$E_{\text{QM/MM}} = E_{\text{QM1+ptch2}}^{\text{HL}} + E_{\text{MM12}, q_i=0}^{\text{CL}} - E_{\text{MM1}, q_i=0}^{\text{HL}} \quad (1),$$

where  $E_{\text{QM1+ptch2}}^{\text{HL}}$  is the QM energy of the QM system truncated by HL atoms and embedded in the point charges modelling system 2 (but excluding the self-energy of the point charges).

$E_{\text{MM1}, q_i=0}^{\text{HL}}$  is the MM energy of the QM system, still truncated by HL atoms, but without any electrostatic interactions. Finally,  $E_{\text{MM12}, q_i=0}^{\text{CL}}$  is the classical energy of all atoms in the system with CL atoms and with the charges of the QM system set to zero (to avoid double-counting of the electrostatic interactions). By this approach, which is similar to the one used in the Oniom method [35], errors caused by the truncation of the QM system should cancel.

The geometry optimisations were continued until the energy change between two iterations was less than 2.6 J/mol ( $10^{-6}$  a.u.) and the maximum norm of the Cartesian gradients was below  $10^{-3}$  a.u. The MM calculations were performed with the Amber software [36],

using the Amber 1999 force field [37,38].

### QM calculations

All QM calculations were run with the Turbomole 6.1 software [39] using the Becke 1988–Perdew 1986 (BP86) density functional [40,41] together with the def2-SV(P) basis set [42]. The calculations were sped up by expanding the Coulomb interactions in auxiliary basis sets, the resolution-of-identity approximation using the auxiliary basis sets corresponding to the def2-SV(P) basis set [43,44]. The big-QM calculations used also the multipole-accelerated resolution-of-identity *J* approach (marij keyword).

In some cases, solvation effects were estimated by single-point calculations using the continuum conductor-like screening model (COSMO) [45,46]. These calculations were performed with default values for all parameters (implying a water-like probe molecule) and a dielectric constant of 4. For the generation of the cavity, we used the optimised COSMO radii in Turbomole (1.30, 2.00, 1.83, 1.72, and 2.16 Å for H, C, N, O, and S, respectively [47] and 2.0 Å for the metals [48]).

The geometry optimisations employed 50 different sizes of the QM system, taken from our previous study [16]. We started from a minimal QM system (Q0, shown in Figure 1) consisting of the Fe and Ni ions, the CO and the two CN<sup>−</sup> ligands, four CH<sub>3</sub>S<sup>−</sup> groups as models of the Cys ligands, imidazole as a model of His-79, and CH<sub>3</sub>COOH as a model of Glu-25, which forms a hydrogen bond to Cys-543. This system was then successively enlarged by adding 20 or 30 chemical groups (not full residues) from the protein, either according to their distance to Q0 (D1–D20) or according to their energy contribution to the QM/MM free-energy difference in a previous study [26] (E1–E30) [16]. The added groups are listed in the tables below. The D2–D15 QM systems differed slightly from those previously used [16] in that two atoms were added, the back-bone O atom of Gly-73 and the H atom of Val-74, thereby avoiding two connected junctions, which gave problems in QM/MM optimisations. In QM systems D5–D8, atoms O of Ile-544 and H of Ala-545 were also added for the same reason. With these changes, D8 became identical to D9.

In addition, big-QM calculations [22] were performed, using QM system C in ref. no. 22, i.e. all groups within 4.5 Å of the minimal QM system Q0 without the Glu-25 model, with junctions moved three residues away, and including all buried charged groups in the protein that are directly connected to the central QM system. To reduce the number of junctions, we also added the side-chain methyl group of Thr-210, as well as the back-bone O atom of Val-496 and the H atom of Val-497. In addition, we added some groups that are optimised in the D1–D20 and E1–E30 systems but not included in the big QM system: the side-chains of Thr S18, Arg-23, Asp-63, Arg-85, Asp-88, Arg-103, Asp-126, and His-538, as well as the N, HA, and C atoms of Leu-479, and two water molecules close to the active site. As the present calculations were based on the protein set-up in ref. no. 16, this big-QM system differed in the protonation state of His-S13 and 27, doubly protonated in ref. no. 22 but protonated only on the NE2 atom in the present calculations. Therefore, the big-QM was neutral in the present calculations and it contained 764 atoms (5368 basis functions and 2930 electrons). This big-QM system is shown in Figure 2.

The big-QM calculations were performed with a point-charge model of the surroundings, because this gave the fastest convergence in our previous study [17]. To the big-QM energy, we added a MM correction,  $E_{\text{MM12}, q_i=0}^{\text{CL}} - E_{\text{MM1}, q_i=0}^{\text{HL}}$ , although it turned out to almost negligible for this large QM system (0.0±0.1 kJ/mol for the 50 different optimised QM systems with a maximum of 1.3 kJ/mol).

The geometry optimisations were performed either using the QM-cluster or the QM/MM approaches. The former approach was used only for QM systems D1–D20 and with three different sets of fixed atoms. In the first set, the HL atoms were fixed (H-fix). In the second

set, we fixed instead the heavy atom to which HL is bound (C-fix). In the third set, both these atoms were fixed, as was recently recommended by Siegbahn (S-fix) [49,50]. All geometry optimisations employed the DFT-D2 dispersion correction [51] and the QM-cluster optimisations were performed in the COSMO continuum solvent.

For accurate reaction energies, basis-set extrapolations, zero-point vibrational energy, thermal corrections, and tests with different DFT functionals should be performed [17]. However, for these convergence studies such calculations are not needed (and they would be extremely costly for the large QM systems). Moreover, our previous studies show that these corrections are small and partly cancelling, giving a net lowering of the HID–HIP reaction energy of only  $\sim 4$  kJ/mol.

## Result and Discussion

In this paper, we study the effect of geometry optimisations on QM-cluster, QM/MM, and big-QM calculations. We use the same test system as in our previous studies [16,17,22,26,30], viz. a simple proton-transfer reaction in [Ni,Fe] hydrogenase that has been shown to be sensitive to the surroundings. The active site of this enzyme is a binuclear Ni–Fe cluster, in which the Fe ion is coordinated to one CO and two CN<sup>-</sup> ligands, as well as two Cys residues. These two Cys residues are also ligands of the Ni ion, which is coordinated by two additional Cys residues. One of the bridging Cys ligands, Cys-546 forms a hydrogen bond to the ND1 atom of His-79. The proton involved in this hydrogen bond can reside either on the Cys residue (called the HID state) or on the His residue (called the HIP state). The proton moves by only  $\sim 0.6$  Å in this reaction, with minimal movements of the other atoms in the active site as can be seen in Figure 1.

Our previous studies have shown that the two states have the same energy when calculated for the small Q0 QM system in vacuum. However, if the closest groups are added to the QM system, the HIP state becomes strongly favoured and the reaction energy stabilises at 110–120 kJ/mol after the addition of  $\sim 13$  groups (i.e. the QM system D13), both if the calculations are performed in a vacuum or in the COSMO continuum solvent [16]. Likewise, the reaction energy converges if groups are instead added according to their importance in a QM/MM free-energy study [26] after the addition of about  $\sim 20$  groups (i.e. the QM system E20). Unfortunately, in this case, the reaction energy is lower, 40–60 kJ/mol [16], showing that even with QM systems of  $>400$  atoms, there may still be energy differences of  $\sim 60$  kJ/mol depending on how the added residues are selected. QM/MM free-energy perturbations and big-QM calculations have indicated that that the latter estimate is probably more reliable [17,26].

Of course, this is problematic for the QM-cluster approach, because it usually employs QM systems of 100–200 atoms. However, it has recently been argued that this poor convergence of the cluster energies is caused by the fact that the QM systems were not optimised [3,20]. In this paper, we test this suggestion by recalculating the energies after geometry optimisations. These calculations have been done for 20 systems, for which the groups were added according to their distance to the active site (D1–D20) and for 30 systems, for which the groups were added according to their QM/MM free energies (E1–E30). Unfortunately, it is not evident how to perform geometry optimisations for large QM systems, especially when not all groups are connected. Therefore, we have tested four different approaches. First, we optimised all structures by QM/MM calculations, for which the MM system ensures that the structure does not diverge significantly from the crystal structure. After the QM/MM geometry optimisation, the MM system was removed and QM-cluster energies were calculated in vacuum or in the COSMO continuum solvent with a dielectric constant of 4. To test the stability of our recently suggested big-QM approach [17], we also calculated big-QM energies with a 764-atom QM system (i.e. taking some of the coordinates from the MM system), shown in Figure 2.

Alternatively, we performed QM-cluster optimisations of the isolated QM systems. In accordance with the QM-cluster approach, some atoms close to the positions where the QM system is truncated were kept fixed in these calculations, in order to avoid unphysical movements of the atoms when the surroundings are removed. We tested three different schemes to select the fixed atoms:

- H-fix: The HL atoms were fixed, i.e. the atoms that were converted to H atoms during the truncation.
- C-fix: We fixed the heavy QM atom, to which the HL atoms are bound (XL). This has been the standard approach to fix atoms previously [1,52].
- S-fix: Both the XL and HL atoms were fixed. This is the approach recently recommended by Siegbahn [49,50]. It typically gives three fixed atoms for each truncated amino acid, because they normally have two junctions, e.g. from the back-bone N and C atoms, when cut at the CA atom.

These approaches were tested for the D1–D20 QM systems, but not for the E1–E30 systems, because the latter contain some groups that are not connected (by covalent or hydrogen bonds) to the other residues and therefore could move much during the geometry optimisations, even if one end of the group is fixed (cf. Figure 3). In some cases, the HID state turned out not to be a stable state. Therefore, it was obtained by restraining the S–H distance of the moving proton to 1.405 Å (the distance in the optimised Q0 structure in vacuum).

#### *QM-cluster optimisations*

Table 1 shows the results obtained for the D1–D20 QM systems optimised with the QM-cluster method with the S-fix scheme. It can be seen that during the geometry optimisation, the atoms move by 0.25–0.53 Å on average (excluding the fixed atoms), with a slightly decreasing trend as the optimised system grows larger (from 63 to 274 atoms). However, a closer look shows in many cases quite extensive changes in the active-site geometry. In particular, for most of the D2–D10 systems, one of the CN<sup>−</sup> ligands moves to a position where it bridges the two metal ions, with nearly equal Ni–C and Ni–N distances for the HID states (~2.1 Å), cf. Figure 3. In five HID cases, it is also connected with a dissociation of the Cys-546 ligand from the Ni ion.

However, once a model of the guanidine side chain of Arg-476 is included in the QM system in D11, the active-site structure is intact, with variations of the Fe– and Ni–ligand distances of less than 0.03 Å and less than 0.07 Å for the Ni–Fe distance (cf. Table S1 in the supplementary material). The covalent S–H and N–H bond lengths of the groups involved in the hydrogen bond are constant within 0.01 Å, whereas the hydrogen-bond distances involving the moving hydrogen atom show variations of up to 0.16 Å.

The energy difference between the HID and HIP state varies between 15 and 120 kJ/mol. For the D2–D10 systems, with their large changes in the active-site geometry, it differs by up to 67 kJ/mol from the calculations without any geometry optimisation. However, for D11–D20 the difference is less than 9 kJ/mol and the energy still converges towards ~120 kJ/mol. This shows that the geometry optimisations have a quite restricted influence on the studied energy as long as the active-site geometry is not distorted.

We also tested QM-cluster optimisations with the H-fix and C-fix schemes. However, the results in Tables 1 and S1 show even larger changes in the active-site geometry. In particular, the His-79 group rotated away from the Cys-546 group in the HIP state of systems D2–D5 to instead form a hydrogen bond to the back-bone carbonyl group of Cys-75, cf. Figure 5. Therefore, these calculations were not pursued beyond the D6 state and the results agree with Siegbahn's suggestion that the S-fix scheme is preferred for QM-cluster optimisations [49,50].

### *QM/MM optimisation on the D1–D20 systems*

Finally, we optimised systems D1–D20 also with the QM/MM approach. The results of these calculations are collected in Table 2. It can be seen that the geometries change appreciably less than in the QM-cluster optimisations, by 0.10–0.17 Å, although no atoms in the QM system were kept fixed in these calculations. This shows that QM-cluster calculations with fixed atoms do not make the systems too rigid, as is often assumed [50], but rather too flexible, owing to the omission of the surroundings. For the QM/MM optimisations, the average movement of the QM atoms slowly increases as the number of optimised atoms is increased.

The active-site geometry is very stable in the QM/MM optimisation (cf. Table S2): The metal–ligand distances show variations of 0.01–0.04 Å over the 20 systems, whereas the Ni–Fe distance varies by up to 0.08 Å. The hydrogen-bond distances for the moving hydrogen show a much larger variation of up to 0.48 Å for the N–H distance in the HID state and 0.74 Å for the S–H distance in the HIP state. However, this variation is mainly caused by the D3–D5 HID structures and the D16 HIP structure. Excluding these deviating states, the variations are only 0.17 and 0.27 Å, respectively. For the larger systems, the obtained active-site geometries are quite similar to those obtained with the QM-cluster approach. For example, for the D11–D20 systems, the metal–ligand distances differ by up to 0.04 Å, the Ni–Fe distance by up to 0.09 Å, whereas the flexible hydrogen-bond distance differs by up to 0.50 Å in the HID state and by up to 0.42 Å in the HIP state.

For the QM/MM-optimised structures, we calculated reaction energies with five different methods. First, we calculated vacuum and COSMO energies for the isolated QM systems. They are directly comparable to the old calculations with fixed geometries [16] and the results in Table 2 show that they are closely similar. The mean absolute difference (MAD) between the results obtained with and without geometry optimisations is 6 and 7 kJ/mol for calculations in a vacuum and with the COSMO continuum solvent, respectively, with a maximum of 15 and 11 kJ/mol. For both calculations, the deviations are positive for the small QM systems (D1–D6) and negative for the large QM systems (D10–D20), and for the vacuum calculations, the deviations are largest (in absolute terms) for the smallest systems (D1–D5).

The COSMO energies are also directly comparable to the energies obtained with the QM-cluster optimisations. Quite satisfactorily, it can be seen that for the D11–D20 systems (for which the cluster optimisations did not distort the active-site structure), the two methods give similar results with a MAD of 6 kJ/mol and maximum deviations of up to 8 kJ/mol. This shows that as soon as the QM system is big enough to not give distorted structures, the two approaches give similar results. In particular, this means that also for the QM/MM optimisations, the energy difference converges to ~118 kJ/mol for D14–D20.

The raw QM/MM energies also give estimates of the energy difference between the HID and HIP states. For a perfect QM/MM method, all QM systems would give the same result. For the five smallest QM systems this is not the case. Instead, the calculations give a low QM/MM energy of 1–15 kJ/mol. However, as soon as the junctions of the His-79 residue is moved to the neighbouring residues in D6, the QM/MM energy goes up to 42 kJ/mol and gets rather stable, although it still shows a slowly increasing trend. Yet, the variation (up to 20 kJ/mol) is smaller than for the QM-cluster vacuum or COSMO energies (variations of up to 57 and 26 kJ/mol for the same systems). The  $E_{\text{QM1+ptch2}}^{\text{HL}}$  term gives similar results (actually with a slightly smaller variation), showing that the MM contribution is small (less than 6 kJ/mol for all systems, and less than 3 kJ/mol for D11–D20).

Finally, we calculated also big-QM energies for the same geometries with the 678-atom QM system in Figure 2. From the results in Table 2, it can be seen that these energies are much more stable, 47–61 kJ/mol with an average of  $56 \pm 1$  kJ/mol over all 20 optimised systems. The largest variation is found when the junctions of the first-sphere ligands are moved to the neighbouring residues (systems D1–D6) and when His-481 is included in the optimised system (D10). For QM systems D10–D20, the variation is less than 6 kJ/mol. This

shows that the big-QM approach gives very stable energies even if the optimised QM system is varied from 46 to 274 atoms.

#### *QM/MM optimisations of the E1–E30 systems*

We have also performed geometry optimisations of QM systems E1–E30. In this case, we used only the QM/MM method, because some charged groups are not connected to the active site, as can be seen in Figure 3. This is also reasonable considering the results from the D1–D20 systems, showing that QM/MM gives the most reliable structures, which are similar to those obtained with QM-cluster optimisations for the largest QM systems.

The results of the geometry optimisations are shown in Tables 3. From these, it can be seen that the atoms move slightly less in the E1–E30 systems than in the D1–D20 systems, 0.08–0.16 Å on average, with a similar increasing tendency with the size of the QM system.

The active-site geometries are very stable (collected in Table S3): The metal–ligand distances vary by 0.01–0.05 Å for the various QM systems and the Ni–Fe distance vary by up to 0.07 Å. As usual, the N–H hydrogen-bond distance in the HID state shows a quite large variation, 1.77–2.41 Å, but the S–H hydrogen-bond distance in the HIP state shows a much smaller variation, 2.08–2.20 Å. The average metal–ligand distances over the E1–E30 systems differ by 0–0.02 Å from the average distances over the D1–D20 systems (optimised by QM/MM), whereas the Ni–Fe distance differs by 0.01–0.02 Å, the N–H hydrogen-bond distance by 0.05 Å, and the S–H hydrogen-bond distance by 0.12 Å, indicating that the results are essentially identical.

We have studied the same energy differences between the HID and HIP states as for the D1–D20 states (Table 3). The QM vacuum energies show quite large variations, –4 to 56 kJ/mol, but for E13–E30, it is reasonably converged to 24–56 kJ/mol. The COSMO calculations show a somewhat smaller variation, 36–79 kJ/mol with the largest and smallest values obtained for the E19 and E26 systems.

If the results are compared to the corresponding calculations without geometry optimisations, it can be seen that the two results are closely similar: The MAD for both the vacuum and COSMO calculations is only 4 kJ/mol. The maximum deviation is 14 kJ/mol, but the deviation is above 9 kJ/mol only for the E13 system, both in vacuum and in the COSMO continuum solvent. Thus, the present calculations clearly show that geometry optimisations have a very small effect on the HID–HIP energy difference. In particular, the large difference between the converged results of the D (~110 kJ/mol) and E (64 kJ/mol with COSMO) approaches remains and it is certainly not affected by the geometry optimisations, as has been recently suggested [3,20].

The QM/MM energies show a larger variation than the vacuum energies, –17 to 51 kJ/mol. However, at E13, convergence has been reached and for all the larger systems, the energy difference has stabilised to 37–51 kJ/mol, i.e. appreciably better than for both the vacuum (24–56 kJ/mol) and the COSMO calculations (36–79 kJ/mol). Again, it seems to be crucial to include the back-bone of His-79 to stabilise the energies (E12). The  $E_{\text{QM1+ptch2}}^{\text{HL}}$  energy gives similar results (–14 to 64 kJ/mol), but they differ somewhat more from the QM/MM energies than for the D systems, because the MM term is rather large for some systems, up to 16 kJ/mol for E21.

The big-QM calculations give stable results, 43–69 kJ/mol and  $57 \pm 1$  kJ/mol on average. They show a slight increase to E7 (when the back-bone around Cys-72 is added) and then a decrease to E13 (when the back-bone around His-79 and a water molecule that forms hydrogen bonds to the backbone of three of the active-site residues, including His-79, are added).

## Conclusions

In this paper, we have studied the effect of geometry optimisations for energies calculated with QM-cluster, QM/MM, and big-QM methods. We have used the same test case as in our previous convergence studies of these methods and QM/MM free energy perturbations [16,17,22,26], the proton transfer between His-79 and the Ni ligand Cys-546 in [Ni,Fe] hydrogenase, which has been shown to be very sensitive to the surroundings. In most previous studies, we have kept the surroundings outside the 46-atom Q0 QM system fixed at the crystal structure, whereas we here optimise an increasing number of atoms, from 56 to 362, testing 50 different optimised QM systems. Several important results are obtained.

- We have tested four different methods for the geometry optimisations. QM/MM optimisations give the most stable and reliable results. QM-cluster optimisations always give larger changes in the structures and for the smaller QM systems, large changes in the active-site structure are observed (when not all groups forming hydrogen bonds to the metal ligands are included). Fixing both the XL and HL atoms [49,50] improves the results somewhat, but does not fully avoid major changes in the structure. However, for the largest QM systems, both energies and active-site structures obtained from optimisations with the QM/MM and the QM-cluster methods are similar.
- The energy difference between the HID and HIP states, calculated with QM-cluster calculations in vacuum or in a continuum solvent with a dielectric constant of 4, is amazingly insensitive to the geometry optimisations: The energy changes by less than 10–14 kJ/mol, with a MAD of 4–7 kJ/mol. This should be compared to the reaction energies themselves, which change between –4 and 108 kJ/mol for the vacuum energies and between 36 and 113 kJ/mol for the COSMO energies for the same variation of the size of the QM systems.
- In particular, the geometry optimisations do not improve the convergence of the QM cluster calculations, as has been recently suggested [3,20]: The two approaches to add residues to the QM system, by their distance (D1–D20) or by their QM/MM free-energy contribution (E1–E30), still converge to results that differ by ~60 kJ/mol as can be seen in Figure 6. The reason for this is that the convergence of the D calculations is only apparent, because important residues are missing even in D20, as is evident from the big-QM calculations.
- The QM/MM energy differences show an almost as slow convergence as the QM-cluster vacuum energies and a worse convergence than the COSMO energies. However, they do not show the large discrepancy between the D1–D20 and E1–E30 systems as the QM-cluster calculations. This indicates that most of the effect of the missing residues in D20 can be recovered by including them as point charges. Moreover, the convergence is strongly improved as soon as the junctions are moved away from the active-site residues and with the largest QM systems, QM/MM actually converges towards the big-QM results for both the D and E systems.
- Single-point big-QM energies, which are obtained for QM systems where all residues within 4.5 Å of the minimal active site, junctions are moved three residues away from the active site, and all buried charged groups are included (764 atoms) are amazingly stable to the size of the optimised QM system, with variations of 14–26 kJ/mol for the 50 different optimised QM systems and almost the same average for the two approaches (D and E), 56–57±1 kJ/mol.

In conclusion, the present results show that great caution is needed to obtain reliable results with the QM-cluster approach, requiring a careful selection of the residues to include in the QM system. Clearly, all groups forming hydrogen-bonds to the active site need to be included. Fixing 2–3 atoms for each truncated residue is highly recommended. Still, there is a great risk that the QM system is too flexible and that important interactions with buried charges inside the protein are ignored – for the present test case, no QM-cluster approach

would give reliable results.

Instead, we recommend the QM/MM method for the geometry optimisations, which seems to give reliable and stable structures, even with small QM systems. However, the QM/MM energies are quite unstable and we recommend the big-QM approach [17] (preferably combined with free-energy methods) for reliable energies. In particular, we emphasize the importance of moving junctions away from the active site. A strong advantage with the QM/MM structures is that coordinates for all atoms are obtained so that the effect of using different QM systems in single-point energy calculations can easily be tried without redoing the optimisations. The present results also show that keeping the surrounding protein fixed at the crystal structure is a reasonable approximation which minimises the risk of ending up in different local minima for the various states in an enzyme mechanism, an important problem for the QM/MM approach [7,8].

Of course, the present results are strictly only valid for the present [Ni,Fe] hydrogenase test case. The investigated reaction is special in that it involves a minor change in the structure of the active site and that it is very sensitive to the surroundings. [Ni,Fe] hydrogenase is also unusual in that it contains a large number of charges buried inside the protein. Still, our results point out possible problems of the various approaches and how they can be solved, indicating that both standard QM-cluster and QM/MM studies must be interpreted with caution and that a thorough investigation of the convergence of the results is recommended. In future publications, we will test similar methods on other systems.

## Acknowledgements

This investigation has been supported by grants from the Swedish research council (project 2010-5025) and by computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc at Lund University.

## References

- 1 Siegbahn, P. E. M.; T. Borowski, *Acc. Chem. Res.* **2006**, 39, 729-738.
- 2 P. E. M. Siegbahn, F. Himo, *J. Biol. Inorg. Chem.* 2009, 14, 643-651.
- 3 P. E. M. Siegbahn, F. Himo, *WIREs Comput. Mol. Sci.* 1 (2011) 323-336.
- 4 Ramos, M. J.; Fernades, P. A. *Acc. Chem. Res.* **2008**, 41, 689-698.
- 5 Lonsdale, R.; Ranaghan, K.E.; Mulholland, A. J. *Chem. Comm.* **2010**, 46, 2354-2372
- 6 Lin H.; Truhlar D. G. *Theor. Chem. Acc.*, 2007, 117, 185-199
- 7 Senn H. M.; Thiel W. *Angew. Chem. Int. Ed.* 2009, 48, 1198-1229.
- 8 Ryde, U. in *Computational inorganic and bioinorganic chemistry*, E. I. Solomon, R. B. King, R. A. Scott, eds. J. Wiley, Sons, Ltd. Chichester, UK, **2009**, pp. 33-42.
- 9 U. Ryde, K. Nilsson (2003) *J. Am. Chem. Soc.*, 125, 14232-14233
- 10 Siegbahn, P. E. M.; Blomberg, M. R. A.; Chen, S.-L. *J. Chem. Theory Comput.* 2010, 6, 2040-2044.
- 11 Grimme, S. *WIREs Comput. Mol. Sci.* 1 (2011) 211-228.
- 12 Li, J.-L.; Mata, R. A.; Ryde, U. *J. Chem. Theory Comput.*, 2013, DOI 10.1021/ct301094r.
- 13 Sevastik R, Himo F. Quantum chemical modeling of enzymatic reactions: the case of 4-oxalocrotonate tautomerase. *Bioorg Chem* 2007, 35:444-457.
- 14 Hopmann KH, Himo F. Quantum chemical modeling of the dehalogenation reaction of haloalcohol dehalogenase. *J Chem Theor Comp* 2008, 4:1129-1137.
- 15 Georgieva P, Himo F. Quantum chemical modeling of enzymatic reactions: the case of histone lysine methyltransferase. *J Comp Chem* 2010, 31:1707-1714.
- 16 L. Hu, J. Eliasson, J. Heimdal, U. Ryde (2009) "Do quantum mechanical energies

- 
- calculated for small models of protein active sites converge?", *J. Phys. Chem. A*, **113**, 11793–11800
- 17 L. Hu, P. Söderhjelm, U. Ryde "Accurate reaction energies in proteins obtained by combining QM/MM and large QM calculations" *J. Chem. Theory Comput.* **2013**, *9*, 640-649
- 18 Hu H.; Yang W. *Annu. Rev. Phys. Chem.*, **2008**, *59*, 573-601.
- 19 Rod, T. H.; Ryde, U. *J. Chem. Theory Comput.* **2005**, *1*, 1240-1251
- 20 Liao, R.-Z.; Yu, G.; Himo, F. *J. Chem. Theory Comput.* **2011**, *7*, 1494-1501.
- 21 Sumowski, C. V.; Ochsenfeld, C. *J. Phys. Chem. A* **2009**, *113* 11734-11741.
- 22 Hu, L.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.*, **2011**, *7*, 761-777
- 23 Tian, B.; Strid, Å; Eriksson, L. A. *J. Phys. Chem. B* **2011**, *115*, 1918-1926
- 24 Sumowski, C. V.; Schmitt, B. B. T.; Schweizer, S.; Ochsenfeld, C. *Angew. Chem. Int. Ed.* **2010**, *49*, 9951-9955.
- 25 Flaig, D.; Beer, M.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2012**, *8*, 2260-2271.
- 26 Kaukonen, M.; Söderhjelm, P.; Heimdal, J.; Ryde, U. *J. Chem. Theory Comput.* **2008**, *4*, 985-1001
- 27 M. Kaukonen, P. Söderhjelm, J. Heimdal, U. Ryde (2008) "A QM/MM-PBSA method to estimate free energies for reactions in proteins", *J. Phys. Chem. B*, **2008**, *112*, 12537-12548.
- 28 Genheden, S.; Ryde, U. "Improving efficiency of protein–ligand binding free-energy calculations by system truncation", *J. Chem. Theory Comput.*, **2012**, *8*, 1449-1458
- 29 Liao, R.-Z.; Thiel, W. *J. Chem. Theory Comput.* **2012**, *8*, 3793-3803.
- 30 Söderhjelm, P.; Ryde, U. *J. Mol. Struct. Theochem* **2006**, *770*, 199-219.
- 31 Volbeda A.; Montet, Y.; Vernède, X.; Hatchikian, E. C.; Fontecilla-Camps, J. C. *Intern. J. Hydrogen Energy* **2002**, *27*, 1449-1461.
- 32 Ryde, U. *J. Comput.-Aided Mol. Design* **1996**, *10*, 153-164.
- 33 Ryde, U.; Olsson, M. H. M. *Int. J. Quantum Chem.* **2001**, *81*, 335-347.
- 34 Reuter, N. I.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem.* **2000**, *104*, 1720-1735.
- 35 Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357-19363.
- 36 Case, D. A.; Darden, T. A.; Cheatham, III T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *Amber 10 University of California San Francisco*, **2008**.
- 37 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.
- 38 Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.
- 39 Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346-354.
- 40 Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098-3100.
- 41 Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822-8824.
- 42 Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305
- 43 Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283-290.
- 44 Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119-126.
- 45 Klamt, A.; Schüürmann, J. *J Chem Soc Perkin Trans 2* **1993**, *5*, 799-805.
- 46 Schäfer, A.; Klamt A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187-2193.

- 
- 47 Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem.*, **1998**, *102*, 5074-5085.
- 48 Sigfridsson, E.; Ryde, U. *J. Comp. Chem.* **1998**, *19*, 377-395.
- 49 Nilsson Lill SO, Siegbahn PEM (2009) *Biochemistry* 48:577-590
- 50 Siegbahn PEM (2011) *Chem Phys Chem* 12:3274-3280
- 51 S. Grimme, J. *Comput. Chem.* 25, 1463 2004.
- 52 Pelmenschikov V, Blomberg MRA, Siegbahn PEM (2002) *J Biol Inorg Chem* 7:284-298

**Table 1.** Results of the QM-cluster optimisations of QM systems D1–D20. The table lists the average movement of the not-fixed atoms ( $\Delta r$  in Å) and the energy difference between the HID and HIP states ( $\Delta E$  in kJ/mol, obtained in the COSMO continuum solvent) for the present calculations with geometry optimisation (Opt), the old calculations without geometry optimisation (No opt)[16], and the difference (Diff).

QM system	$\Delta r$ (Å)		$\Delta E$ (kJ/mol)		
	Hid	Hip	Opt	No opt	Diff
S-fix					
D1	0.46	0.50	66.0	58.5	3.3
D2	0.49	0.39	17.2	71.2	-62.0
D3	0.46	0.38	14.5	70.5	-65.3
D4	0.38	0.39	82.2	85.0	-9.8
D5	0.39	0.33	23.4	86.4	-70.8
D6	0.46	0.50	60.4	86.0	-26.6
D7	0.44	0.42	58.7	87.7	-29.1
D8	0.53	0.50	41.5	90.9	-48.2
D9	0.53	0.50	41.4	90.9	-48.3
D10	0.38	0.35	34.1	101.1	-61.8
D11	0.29	0.25	90.3	103.6	-7.3
D12	0.28	0.25	96.5	103.7	0.4
D13	0.43	0.39	101.2	116.8	-7.1
D14	0.42	0.39	103.8	118.6	-6.7
D15	0.37	0.36	119.0	120.5	8.7
D16	0.37	0.36	119.0	120.5	8.7
D17	0.39	0.37	117.6	120.4	6.8
D18	0.39	0.38	118.1	120.5	8.7
D19	0.37	0.37	119.9	121.2	7.0
D20	0.32	0.32	112.8	121.2	0.4
H-fix					
D1	0.93	1.08	64.9	58.5	6.4
D2	0.80	0.71	91.4	71.2	25.9
D3	0.66	0.64	98.0	70.5	32.9
D4	0.56	0.62	68.8	85.0	-10.4
D5	0.46	0.55	66.0	86.4	-13.8
D6	0.55	0.49	55.1	86.0	-23.6
C-fix					
D1	0.74	0.88	80.2	58.5	21.7
D2	0.77	0.77	93.2	71.2	27.7
D3	0.94	0.73	26.9	70.5	-38.2
D4	0.66	0.68	65.7	85.0	-13.5
D5	0.59	0.62	62.2	86.4	-17.6
D6	0.66	0.76	87.4	86.0	8.7

**Table 2.** Results of the QM/MM optimisations of QM systems D1–D20. The table lists the residue that is added for each QM system, the number of atoms (#atm), the minimum distance between the added group and the minimal QM system Q0 (dist in Å), the average movement of the optimised atoms ( $\Delta r$  in Å) and the energy difference between the HID and HIP states ( $\Delta E$  in kJ/mol). For the latter, we give five measures from the present calculations with geometry optimisation (Optimised) and two from the old calculations without geometry optimisation [16] (No opt): energies for the isolated QM system, calculated either in vacuum (Vac) or in the COSMO continuum solvent with a dielectric constant of 4 (COS), QM/MM energies and the  $E_{\text{QM1+ptch2}}^{\text{HL}}$  QM/MM energy component in Eqn. 1 (QMq), and the big-QM energy.

QM system	Residue	#atm	dist Å	$\Delta r$		$\Delta E$ (kJ/mol)						
				HID	HIP	Optimised					No opt	
						Vac	COS	QM/MM	QMq	big-QM	Vac	COS
D1	Cys-72 <sup>a</sup>	63	0.00	0.10	0.12	13.5	62.8	9.1	9.4	55.8	2.5	58.5
D2	Cys-75 <sup>a</sup>	82	0.00	0.11	0.15	35.4	79.2	14.9	20.5	47.1	20.4	71.2
D3	Cys-543 <sup>a</sup>	99	0.00	0.12	0.14	34.1	79.9	2.8	8.9	53.2	18.9	70.5
D4	Cys-546 <sup>a</sup>	118	0.00	0.13	0.14	48.2	92.0	10.5	14.8	56.1	39.8	85.0
D5	Glu-25 <sup>a</sup>	138	0.00	0.15	0.16	51.5	94.2	0.7	6.4	61.1	41.3	86.4
D6	His79 <sup>a</sup>	158	0.00	0.15	0.16	51.3	87.0	41.5	46.4	48.9	44.0	86.0
D7	Thr-S18 <sup>b</sup>	164	1.75	0.14	0.15	52.4	87.7	44.5	48.7	49.0	45.7	87.7
D8	Wat	167	1.89	0.15	0.16	54.5	89.6	44.4	48.8	48.9	49.8	90.9
D10	His-481 <sup>b</sup>	176	1.99	0.15	0.16	66.8	96.0	48.3	50.7	55.5	67.9	101.1
D11	Arg-476 <sup>a</sup>	188	2.07	0.15	0.15	70.9	97.6	46.1	49.1	56.5	73.0	103.6
D12	Val-497 <sup>b</sup>	202	2.11	0.15	0.16	75.8	96.1	54.0	54.5	57.2	78.3	103.7
D13	Arg-476 <sup>b</sup>	219	2.18	0.16	0.16	98.8	108.3	52.4	52.7	59.8	104.4	116.8
D14	Val-78 <sup>b</sup>	228	2.31	0.17	0.17	102.7	110.4	52.3	52.3	61.1	106.7	118.6
D15	Ala-499 <sup>a</sup>	242	2.32	0.16	0.17	103.6	110.3	53.4	55.0	58.1	108.6	120.5
D16	Val-74 <sup>a</sup>	242	2.39	0.16	0.17	103.5	110.4	53.4	54.9	58.2	108.6	120.5
D17	Ala-545 <sup>b</sup>	245	2.39	0.17	0.17	103.1	110.9	50.3	51.8	58.7	108.2	120.4
D18	Val-74 <sup>b</sup>	254	2.43	0.17	0.17	103.7	109.4	51.5	53.5	59.6	108.1	120.5
D19	Leu-479 <sup>b</sup>	271	2.48	0.17	0.17	108.1	112.9	61.5	64.0	58.8	110.1	121.2
D20	Ala-499 <sup>b</sup>	274	2.54	0.17	0.17	107.8	112.4	57.2	58.2	59.0	109.7	121.2

<sup>a</sup> Back-bone , not side chain.

<sup>b</sup> Side chain.

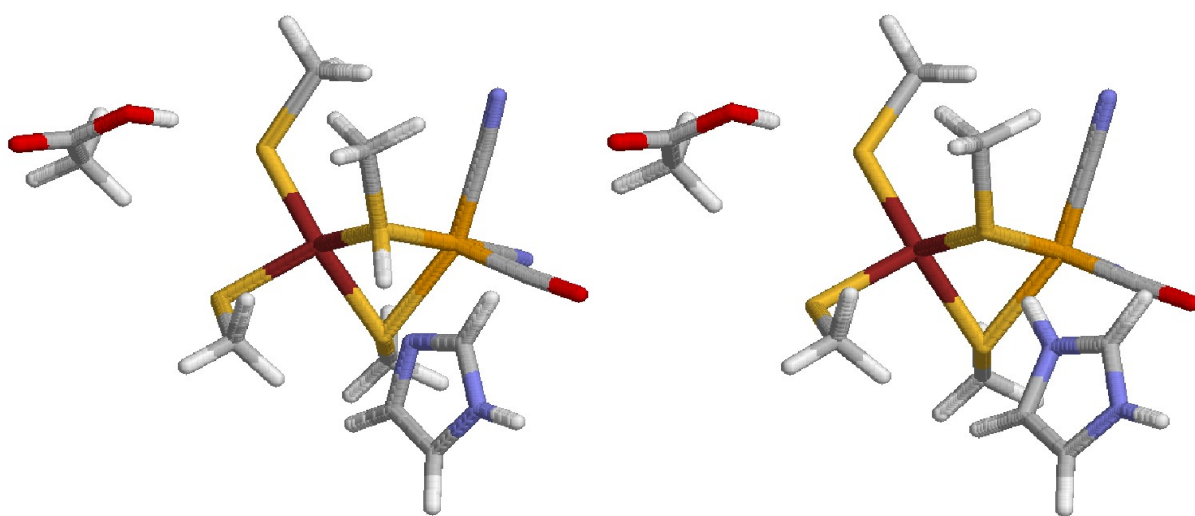
**Table 3.** Results of the QM/MM optimisations of QM systems E1–E30. The entries are the same as in Table 2.

QM system	Residue	#atm	dist	$\Delta r$				$\Delta E$				
				HID	HIP	Optimised			Not opt			
						Vac	COS	QM/MM	QMq	big-QM	Vac	COS
E1	Ile-544 <sup>a</sup>	56	1.75	0.08	0.09	3.2	55.9	11.1	10.7	56.4	-3.3	54.1
E2	Arg-476	69	2.18	0.10	0.10	30.5	69.1	12.0	11.5	58.6	26.8	69.5
E3	Cys-546 <sup>a</sup>	84	1.66	0.13	0.13	37.8	76.3	8.9	7.6	60.9	38.8	79.5
E4	Asp-114	91	3.32	0.14	0.14	19.1	69.2	4.2	3.9	60.6	19.0	70.9
E5	Cys-75 <sup>a</sup>	101	1.63	0.13	0.14	24.8	74.9	2.5	3.4	61.0	23.3	76.1
E6	Asp-541	108	4.82	0.14	0.16	5.7	67.2	-0.9	0.7	62.3	2.1	62.9
E7	Cys-72 <sup>a</sup>	118	1.64	0.15	0.15	-0.2	64.0	-17.2	-14.4	68.7	-3.6	58.6
E8	Gln-69 <sup>a</sup>	130	3.21	0.14	0.14	8.5	71.7	3.6	6.6	65.9	2.6	64.3
E9	Cys-543 <sup>a</sup>	137	1.61	0.14	0.14	9.7	73.3	-0.7	2.4	66.3	2.8	65.7
E10	Glu-S22	144	5.64	0.13	0.14	-3.8	61.4	1.5	3.6	63.9	-6.5	54.6
E11	His-481	153	1.99	0.14	0.13	15.3	68.7	12.0	13.3	64.5	8.9	65.3
E12	His-79 <sup>a</sup>	156	1.60	0.14	0.14	18.0	70.4	20.1	19.9	58.7	21.9	70.6
E13	Wat	159	2.70	0.15	0.16	25.0	75.8	39.3	51.5	53.5	10.8	65.7
E14	Pro-542 <sup>a</sup>	164	3.46	0.15	0.16	24.6	73.8	41.1	52.1	51.9	22.3	76.0
E15	Mg <sup>b</sup>	202	4.15	0.15	0.15	39.4	74.8	44.9	55.7	54.8	31.0	70.5
E16	Arg-428	215	7.85	0.15	0.15	47.2	74.0	45.1	54.5	55.4	38.7	76.5
E17	Arg-70	228	7.26	0.15	0.15	42.4	66.8	43.9	53.4	52.6	37.2	72.9
E18	His-115	238	6.36	0.15	0.15	47.9	74.9	45.3	54.5	57.5	44.5	75.8
E19	His-538	248	8.85	0.15	0.15	55.7	79.1	46.0	56.4	53.9	52.2	79.4
E20	Ala-80 <sup>a</sup>	258	2.91	0.15	0.15	50.9	73.8	40.0	38.2	57.3	50.2	72.9
E21	Arg-103	271	11.98	0.15	0.16	48.6	69.0	47.8	64.0	42.6	45.2	68.4
E22	Glu-S46	278	8.73	0.15	0.15	45.1	59.9	40.4	38.0	51.7	37.8	64.1
E23	Glu-S75	285	6.98	0.15	0.15	39.8	53.2	38.4	34.8	53.3	33.0	56.9
E24	Asp-88	292	13.92	0.16	0.15	33.0	60.5	38.0	35.9	54.2	35.0	61.0
E25	Arg-85	305	9.56	0.16	0.15	36.0	49.9	37.4	32.8	53.6	35.2	56.1
E26	Val-78 <sup>a</sup>	319	4.40	0.15	0.15	26.0	35.7	51.4	50.0	46.7	31.9	41.6
E27	Ala-71 <sup>a</sup>	335	4.24	0.16	0.16	24.0	42.0	50.4	49.7	53.9	28.9	49.2
E28	Arg-23	348	11.91	0.16	0.16	25.2	48.2	50.7	50.7	54.8	34.1	56.2
E29	Asp-126	355	13.98	0.16	0.16	25.4	41.6	50.0	49.7	53.2	27.9	48.8
E30	Asp-63	362	13.03	0.16	0.16	27.3	45.7	49.9	48.9	55.1	29.0	51.0

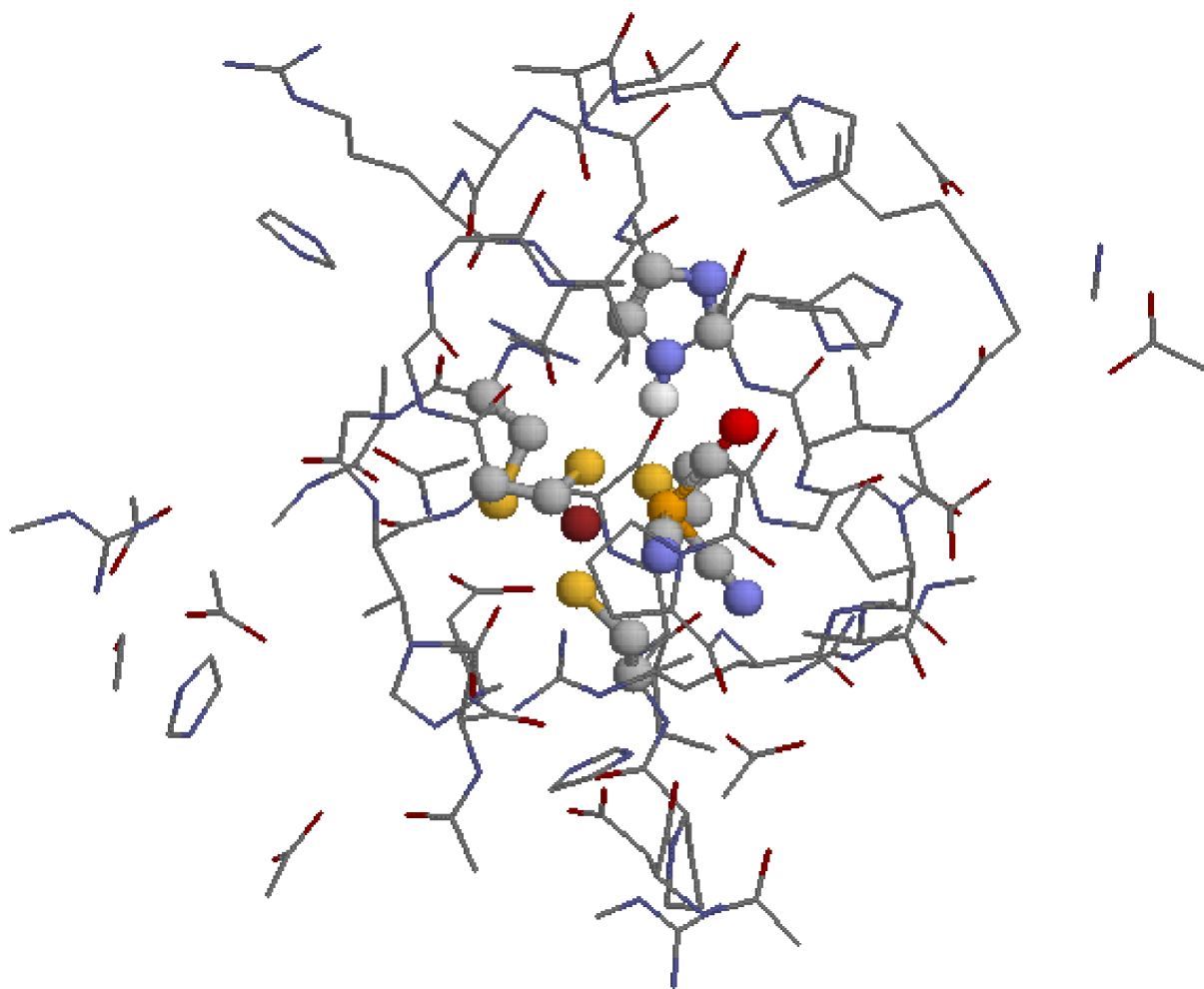
<sup>a</sup> Back bone, rather than side chain.

<sup>b</sup> The Mg site, consisting of a Mg<sup>2+</sup> ion, three water molecules, as well as the side chains of Glu-53 and His-639, and the back-bone O atom of Gln-540 (modelled by CH<sub>3</sub>CONHCH<sub>3</sub>).

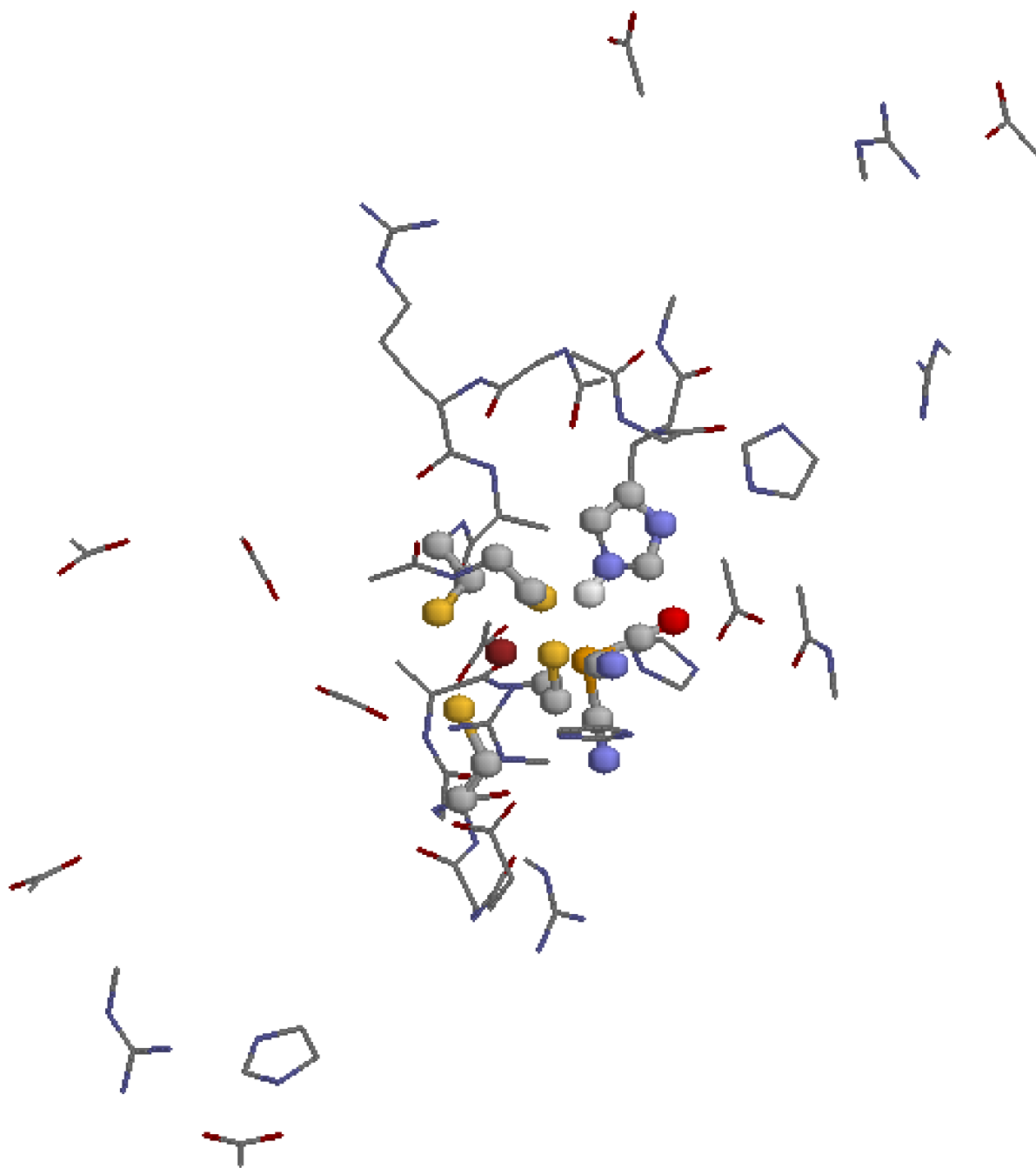
**Figure 1.** QM system Q0 in the HID (left) and HIP states (right).



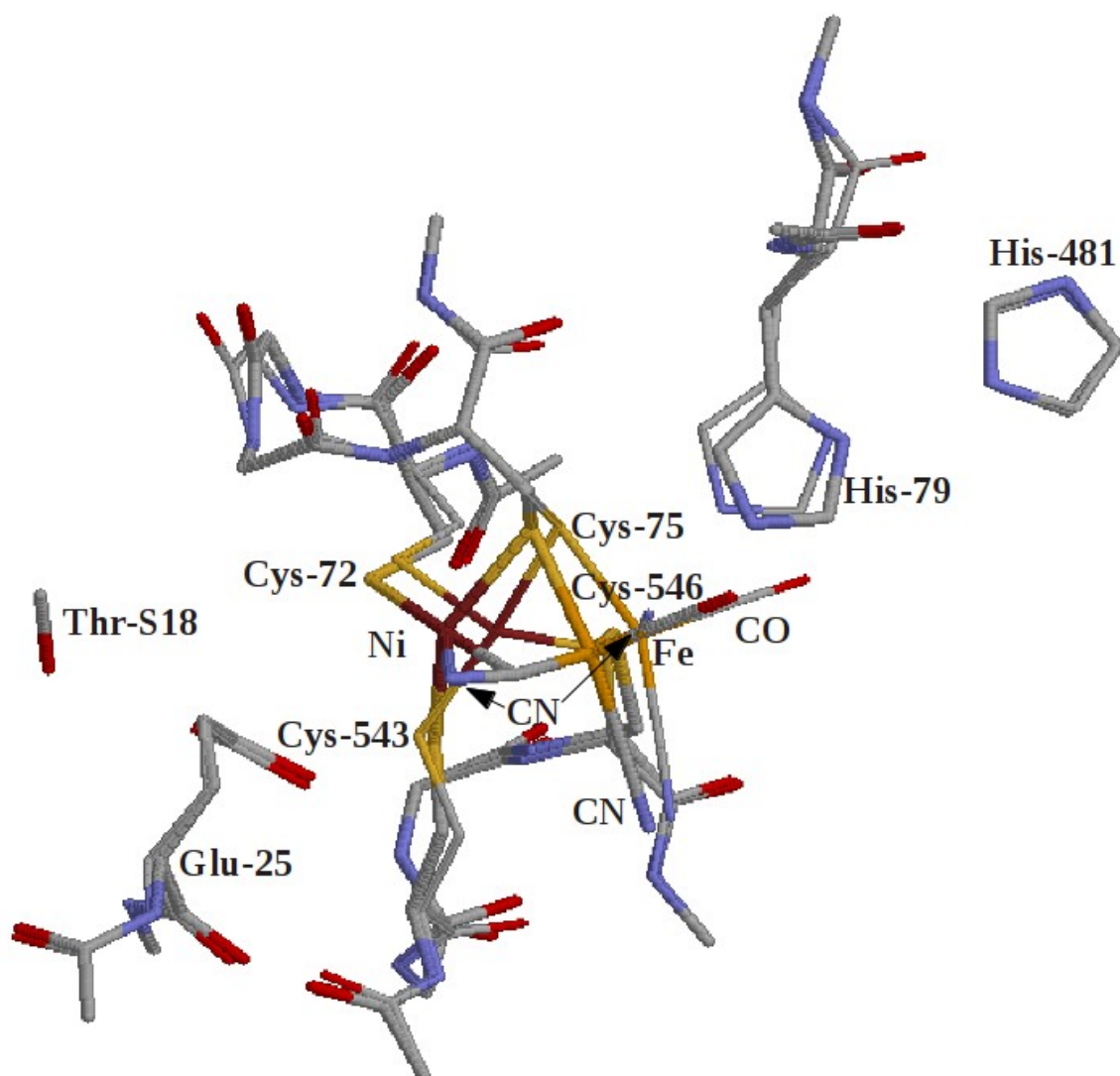
**Figure 2.** The 764-atom big-QM system. Hydrogen atoms are ignored and the active-site atoms are showed in ball-and-stick representation.



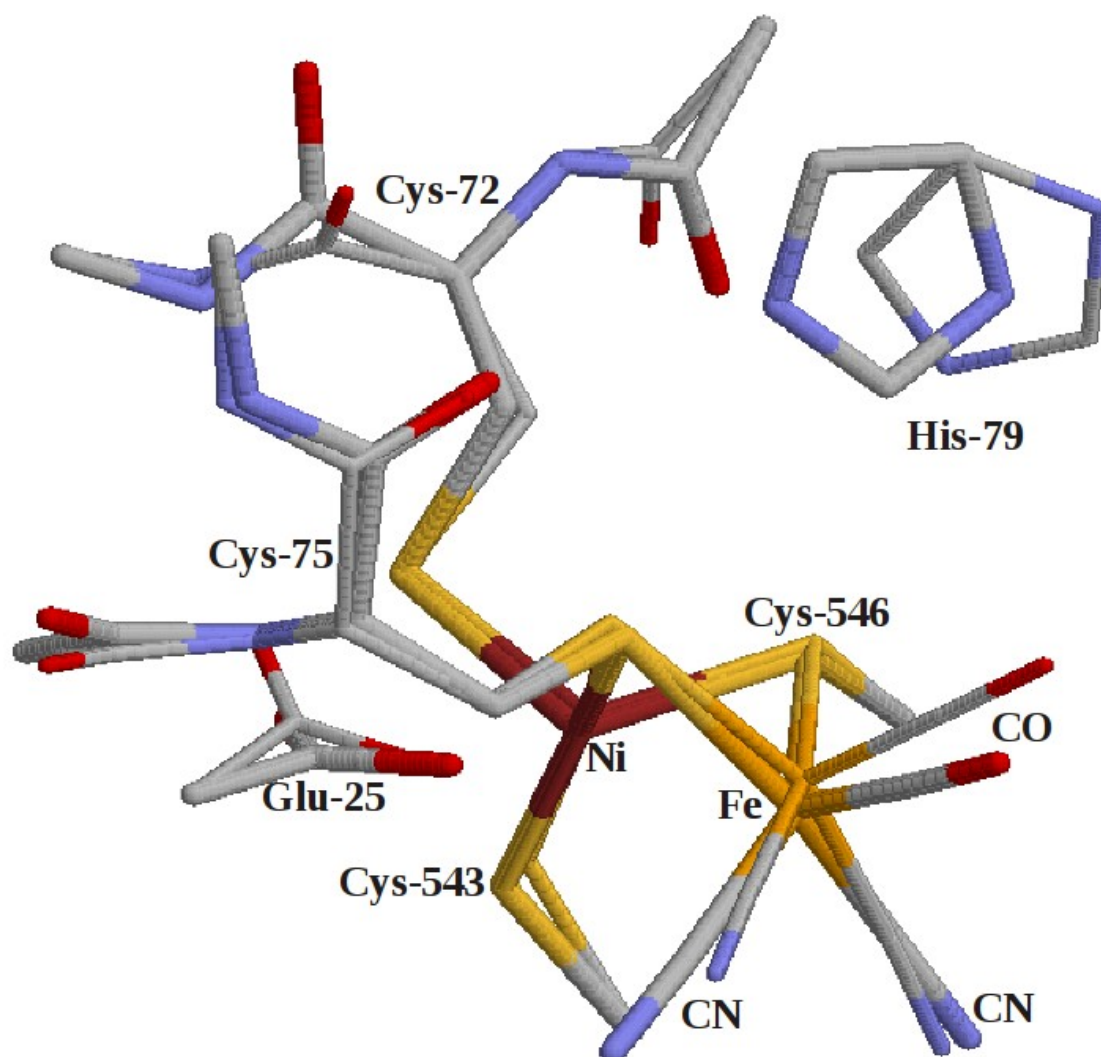
**Figure 3.** The largest E30 QM system in the HIP state. Hydrogen atoms are ignored and the active-site atoms are showed in ball-and-stick representation.



**Figure 4.** Bridging of one of the  $\text{CN}^-$  ligands in the D10 HID model. The geometries before and after the geometry optimisation are shown in thin and thick lines, respectively. Hydrogen atoms are omitted for clarity.



**Figure 5.** Rotation of the His-79 ring in the D2 HIP model optimised with the QM-cluster method and the C-fix scheme. The geometries before and after the geometry optimisation are shown in thin and thick lines, respectively. Hydrogen atoms are omitted for clarity.



**Figure 6.** Convergence of the various energy differences for the D1–D20 and E1–E30 systems.

