QM/MM-PBSA method to estimate free energies for reactions in proteins

Kaukonen, Markus; Söderhjelm, Pär; Heimdal, Jimmy; Ryde, Ulf

Published in:
The Journal of Physical Chemistry Part B

DOI:
10.1021/jp802648k

2008

Document Version:
Peer reviewed version (aka post-print)

A QM/MM–PBSA method

to estimate free energies for reactions in proteins

Markus Kaukonen, Pär Söderhjelm, Jimmy Heimdal,

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,

Tel: +46 – 46 2224502, Fax: +46 – 46 2224543

2008-06-11

Abstract

We have developed a method to estimate free energies of reactions in proteins, called QM/MM–PBSA. It estimates the internal energy of the reactive site by quantum mechanical (QM) calculations, whereas bonded, electrostatic, and van der Waals interactions with the surrounding protein are calculated at the molecular mechanics (MM) level. The electrostatic part of the solvation energy of the reactant and the product is estimated by solving the Poisson–Boltzmann (PB) equation and the non-polar part of the solvation energy is estimated from the change in solvent accessible surface area (SA). Finally, the change in entropy is estimated from the vibrational frequencies. We test this method for five proton-transfer reactions in the active sites of [Ni,Fe] hydrogenase and copper nitrite reductase. We show that QM/MM–PBSA reproduces the results of a strict QM/MM free-energy perturbation method with a mean absolute deviation (MAD) of 8–10 kJ/mol if snapshots from molecular dynamics simulations are used and 4–14 kJ/mol if a single QM/MM structure is used. This is appreciably better than the original QM/MM results or if the QM energies are supplemented with a point-charge model, a self-consistent reaction field, or a PB model of the protein and the solvent, which give MADs of 22–36 kJ/mol for the same test set.

Key Words: QM/MM, MM/PBSA, proton transfer, [Ni,Fe]-hydrogenase, copper nitrite reductase, density functional theory.

Introduction

During recent years, computational chemistry has become a powerful complement to experiments to study the function of proteins. One of the most important strengths of theoretical chemistry is its ability to estimate energies (e.g. activation, reaction, and conformational energies), which are central to the understanding of protein function, because they govern all chemical processes. However, also with theoretical methods, it is a challenge to obtain accurate energies, especially the experimentally relevant *free* energies for proteins in solution, i.e. including a proper treatment of solvation effects and long-range electrostatics.

The most rigorous theoretical method to obtain free energies is free-energy perturbation (FEP).[1] It estimates the difference in free energy between two states by slowly changing one state to another through a number of unphysical intermediate states, performing extensive sampling at each intermediate. Therefore, the method is extremely time-consuming and it spends most of its time at uninteresting intermediate states.

More approximate methods have been developed to estimate the free energy from simulations of only the end-point states (reactant and product), e.g. the linear interaction energy (LIE) method[2] and the molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) method.[3] The latter approach is attractive, because it does not contain any adjustable parameters and it involves a set of physically well-defined terms: The free energy of each of the states is estimated as a sum of four terms:

$$G = \langle E_{MM} \rangle + \langle G_{Solv} \rangle + \langle G_{np} \rangle - T\langle S_{MM} \rangle \quad (1)$$

where G_{Solv} is the polar solvation energy of the molecule, estimated by the solution of the Poisson–Boltzmann (PB) equation,[4] G_{np} is the non-polar solvation energy, estimated form the solvent-accessible surface area of the molecule,[5] T is the temperature, S_{MM} is the entropy of the molecule, estimated from a normal-mode analysis of harmonic frequencies calculated at the molecular mechanics (MM) level, and E_{MM} is the MM energy of the molecule, i.e. the sum of the internal energy of the molecule (i.e. bonded terms, E_{int}), the electrostatics (E_{es}), and van der Waals interactions (E_{vdW}):

$$E_{MM} = E_{int} + E_{es} + E_{vdW} \quad (2)$$

All the terms in Eqn. 1 are averages of energies obtained from a number of snapshots taken from molecular dynamics (MD) simulations. The MM/PBSA method has successfully been applied to several different systems, in particular to the estimation of binding free energies of small ligands (drug candidates) to proteins.[3,6,7,8,9,10,11]

Owing to the extensive sampling needed for proper convergence of free energy methods, they are normally based on a molecular mechanics (MM) potential. The results will therefore depend on the quality of this potential and the application of these methods is restricted to cases, for which such a potential is available. For many of the most interesting systems, e.g. for transition states and metal sites, standard MM potentials perform poorly. Even if there are solutions to these problems,[12,13] it would be more attractive to employ the more versatile quantum mechanical (QM) methods for these systems instead.

QM methods have frequently been used for the study of protein function, but mainly with minimal model systems, including only a few residues from the active site of the protein.[14] At the next level of approximation, the whole protein is included in the calculations, e.g. by describing the active site by QM methods and the surrounding protein by a simple point-charge model, solving the Poisson–Boltzmann (PB) equation[15,16,17,18,19,20] (i.e. considering only the electrostatic effects), or by a full MM treatment (the QM/MM approach).[21,22] QM calculations of a whole protein have also started to appear, either at the semiempirical level[23,24] or at higher levels by fractionation approaches.[25] Recently, QM methods have also started to be combined with FEP methods to obtain free energies.[26,27,28,29,30,31,32,33,34] Most of these are based on single-point QM calculations or by sampling with semiempirical methods. However, recently also methods based on more accurate density functional theory (DFT) have started to emerge.[29,30,31,32,33] Naturally, such approaches are extremely time-consuming.

In this paper, we will investigate if it is possible to reduce the time-consumption of such

methods, while retaining the accuracy. In analogy with the pure MM methods, we will use the MM/PBSA approach and see if it can be improved by treating parts of the protein by QM methods. Thus, we will develop a QM/MM–PBSA approach, with the primary aim to study reactions in proteins. Such an approach has been tested before (and also a QM–PBSA approach[35]), but only for the simpler case of binding a ligand to a protein.[36,37] We compare different implementations of this approach and test their performance on two challenging reactions: the proton-transfer reactions between a metal ligand and a second-sphere residue in copper nitrite reductase[18] and [Ni,Fe] hydrogenase.[38] The performance of the method will be judged by comparing to the more rigorous QM/MM thermodynamic cycle perturbation approach (QTCP).[32,33,38]

Methods

[Ni,Fe] hydrogenase

The calculations on [Ni,Fe] hydrogenase (H2ase) were based on a crystal structure of the Ser499Ala mutant from *D. fructosovorans,*[39] which had the lowest resolution (1.81 Å) at the start of our investigation (Figure 1). The hydroxide group of Ser-499 forms a hydrogen bond to one of the CN^- ligands of Fe, but in the mutant, a water molecule replaces this group, leading to unchanged enzymatic properties, vibrational frequencies, and structure, but much better diffracting crystals.[39] In the calculations, Ser-499 was reinserted and the extra water molecule was deleted.

In the QM/MM calculations,[19] all residues located more than 27 Å from the Ni atom were deleted and solvation water molecules were added to the protein, forming a sphere with radius 33 Å (602 protein residues and 1042 water molecules, giving a total of 12178 atoms). Four metal sites are inside this cut-off radius, viz. the proximal [4Fe-4S] cluster, the [3Fe-4S] cluster, a six-coordinated Mg^{2+} ion at the C-terminus of the large subunit (its three water ligands were always considered a part of the protein), and the active [Ni,Fe] site. For these sites, we used Merz–Kollman electrostatic potential (ESP) charges, taken from QM calculations of truncated models of each site.

The protein was treated in exactly the same way as in our previous QTCP study:[38] Hydrogen atoms were added using the AMBER package.[40] The protonation status of histidine (His) residues was determined by inspection of the local surroundings and hydrogen-bond structure. This gave protonation of the $N^{\delta 1}$ atom for residues S92 (an initial S indicates the small subunit; residue numbers without S refer to the large subunit; cf. Figure 1), S243, 481, and 549, protonation of the $N^{\epsilon 2}$ atom for S13, 27, 66, 79, 113, 118, 121, 123, 210, 228, 349, 367, and 419, and protonation on both these atoms for 115, 204, 305, and 538. 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, which is involved in the proton-transfer path from Cys-543 to the protein surface. Before running the QM/MM calculations, the positions of the hydrogen atoms and solvation water molecules were optimised by a simulated-annealing molecular dynamics calculation followed by a MM minimisation, as has been described before.[41]

The smallest quantum system (called N) consisted of the Ni and Fe ions and their first-sphere ligands ($2CN^-$, CO, Cys-72, Cys-75, Cys-543 and Cys-546). Two second-sphere residues, Glu-25 and His-79, were also included because they share hydrogen atoms with Cys-543 and Cys-546, respectively. The Cys residues were modelled by CH_3S^-, whereas His was modelled by imidazole and Glu-25 by acetic acid. Thus, the QM system consisted of 46 atoms. The aim of this investigation was to study the relative stability of the two states in which the proton shared by Cys-546 and His-79 resides on either His-79 (called the HIP state) or on Cys-546 (called the HID state). The QM system is depicted in Figure 2. To gain more statistics,

five larger QM systems were also studied, comparing the results with those of our earlier QTCP calculations.[38] These QM systems are also shown in Figure 2 and they contain 59 (NR), 66 (NACG), 79 (NACGR), 91 (NCHACG), and 104 atoms (NCHACGR), respectively.

Nitrite reductase

Copper nitrite reductase (NIR) is a homotrimer, as is shown in Figure 3a. The active site contains a copper ion, bound to three histidine ligands and a solvent molecule (water or OH^-; Figure 3b). One of the His ligands, His-100, forms a hydrogen bond to the carboxylate group of Glu-279 and the solvent molecule forms a hydrogen bond to the carboxylate group of Asp-98. It is well-known that metals lower the pK_a of their ligands. Therefore, it is possible that the protons involved in these two hydrogen bonds may actually reside on the carboxylate groups. This gives rise to four possible protonation states of the active site, which we will call Wat (protons on water and His-100), Hyd (protons on Asp-98 and His-100), Imm (protons on water and Glu-279), and Both (protons on Asp-98 and Glu-279).[18] They are illustrated in Figure 4. The aim of this work is to estimate the relative free energy of these four states.

The calculations on copper nitrite reductase (NIR) were based on the crystal structure of oxidised NIR at pH 6.0 (Protein Data Bank file 1NIC at 1.9 Å resolution).[42] The protein is composed of three identical subunits (although only one subunit is present in the crystallographic unit cell), each with 333 residues and two copper ions (one type 1 blue copper ion, used for electron transfer, and the catalytic type 2 copper ion). The QM structures for NIR were obtained in our previous study[18] using the quantum-refinement approach. It is essentially a QM/MM minimisation, in which the structure is restrained to remain close to crystallographic raw data (the structure factors).

In the calculations, the whole trimeric protein (999 amino acids and 6 copper ions) was considered. This was accomplished by copying the structure of the QM system from the quantum-refinement calculation[18,38] to the other two subunits of the protein. All Asp, Glu, Lys, and Arg residues were assumed to be ionised, whereas the protonation status of the His residues were determined from a study of the solvent exposure and hydrogen-bond interactions in the crystal: His-100, 135, and 306 were protonated on the $N^{\delta 1}$ atom, His-60, 95, 145 and 255 on the $N^{\epsilon 2}$ atom, and the other His residues were protonated on both nitrogen atoms. The only Cys residue is a copper ligand and was therefore assumed to be deprotonated. For disordered residues, the conformation with the lowest occupancy was deleted (conformation A for Arg-54, and conformation B for Glu-197, Arg-250, and Arg-271). A sulphate ion and disordered crystal water molecules were also removed, except DIS-1250, 1275, and 1282. For the QM energies in QM/MM–PBSA, only subunit A was considered, to save computer time, whereas the MM energies were obtained by perturbing all three copper sites simultaneously (but the presented free energies correspond to a single Cu site).

QM calculations

To be consistent with our earlier calculations,[18,19,38] the QM calculations on the two proteins were slightly different. For H2ase, the calculations were performed with the Becke 1988–Perdew 1986 (BP86) density functional[43,44] together with the 6-31G* basis set[45] for H, C, N, O and S, and the DZP basis set for Fe and Ni.[46,47] The calculations were sped up by expansion of the Coulomb interactions in auxiliary basis sets, the resolution-of-identity approximation.[48,49] For NIR, we instead used the three-parameter hybrid B3LYP method, as implemented in the Turbomole package.[50,51,52] These calculations employed the 6-31G* basis set for all atoms,[45] except for copper, for which we used the DZP basis sets of Schäfer et al.,[46,47] enhanced with *p*. *d,* and *f*-type functions with exponents of 0.174, 0.132, and 0.39.

For both systems, the structures were optimised until the change in energy between two iterations was below 2.6 J/mol (10^{-6} a.u.) and the norm of the maximum norm of the internal

gradients was below 10^{-3} a.u.

MM and MD calculations

All MM calculations were run with the sander module in the AMBER 9 software,[40] using the Amber 1999 force field (FF99).[53,54] The QM system was represented by charges fitted to the electrostatic potential, calculated in 115 000–265000 points selected at random around the QM system to a distance of 8 Å. The fit used singular-value decomposition to ensure that all the fitted charges are significant. The charges were constrained to reproduce the QM dipole moment exactly and they were restrained to reproduce the QM quadrupole moment, as well as the Boltzmann-weighted electrostatic potential, the CHELP-BOW method.[55] The charge on each atom that was converted to a H atom in the QM calculations (C junction atom) was adapted so that the total charge of the amino acid (including both QM and MM atoms) was the same as the sum of QM charges of the corresponding QM fragment.[56] Thereby, we ensure that the total charge of the simulated system is an integer, but we still allow charge transfer within the QM system (the amino acids with QM atoms have non-integer total charges). Moreover, the charges on the C junction atoms are changed from what is typical for a hydrogen atom to what is more typical for carbon atoms.

All bond lengths involving hydrogen atoms were constrained by means of the SHAKE algorithm.[57] The water solvent was described explicitly using the TIP3P model.[58] The electrostatics were treated with particle-mesh Ewald method[59,60] with a grid size of 80^3 Å, a fourth-order B-spline interpolation, a tolerance of 10^{-5}, and a real-space cut-off of 8 Å. The temperature was kept constant at 300 K using the Berendsen weak-coupling algorithm[61] with a time constant of 1 ps. The MD time step was 2 fs and the non-bonded pair list was updated every 50 fs.

QM/MM calculations

The QM/MM calculations were carried out with the COMQUM program.[56,62] In this approach, the protein and solvent are split into three subsystems: The QM region (system 1) contains the most interesting atoms and is relaxed by QM methods. System 2 consists of the rest of the protein, whereas system 3 is comprised of the explicitly modelled water molecules. In the present investigation, both systems 2 and 3 have been fixed at the original (crystallographic) coordinates in the QM/MM geometry optimisations (this is necessary for the QTCP calculations, but the two systems are subsequently relaxed in the MD simulations). In the QM calculations, system 1 is represented by a wave function, whereas all the other atoms are represented by an array of partial point charges, one for each atom, taken from MM libraries. Thereby, the polarization of the quantum chemical system by the surroundings is included in a self-consistent manner.

When there is a bond between systems 1 and 2 (a junction), the hydrogen link-atom approach is employed: The quantum region is truncated by hydrogen atoms, the positions of which are linearly related to the corresponding carbon atoms in the full system.[56,63] In order to eliminate the non-physical effect of placing point charges on atoms bound to junction atoms (i.e., the closest neighbours of QM system), these charges are zeroed, and the resulting residual charges are smoothly distributed.[56,62]

The total energy is calculated as:

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM1+ptch23}} - E_{\mathrm{MM1\_noel1}} + E_{\mathrm{MM123\_noel1}} \quad (3),$$

where $E_{\mathrm{QM1+ptch23}}$ is the QM energy of the quantum system truncated by hydrogen atoms and embedded in the set of point charges, representing systems 2 and 3 (but excluding the self-energy of the point charges). $E_{\mathrm{MM1\_noel1}}$ is the MM energy of the quantum system, still truncated by hydrogen atoms, but without any electrostatic interactions. Finally, $E_{\mathrm{MM123\_noel1}}$ is the classical energy of all atoms in the system with normal atoms at the junctions and with the charges of the

quantum system set to zero (to avoid double-counting of the electrostatic interactions). By using this approach, which is similar to the one used in the Oniom method,[64] errors caused by the truncation of the quantum system should cancel out.

The QM/MM-PBSA approach

In the MM/PBSA method,[3] free energy of a system is calculated from Eqns. 1 and 2. In this paper, we will extend the MM/PBSA method to the calculation of reaction energies in metalloproteins with QM/MM methods. We will call this approach QM/MM-PBSA. In principle, this extension is simple: we only replace the E_{MM} term in Eqn. 1 with the corresponding total energy term $E_{\mathrm{QM/MM}}$ of a QM/MM approach, e.g. the ComQum energy in Eqn. 3:

$$G = \langle E_{\mathrm{QM/MM}} \rangle + \langle G_{\mathrm{Solv}} \rangle + \langle G_{\mathrm{np}} \rangle - T\langle S_{\mathrm{QM/MM}} \rangle \quad (4)$$

The entropy term in this equation ($S_{\mathrm{QM/MM}}$) consists of the entropy of the QM system (S_{QM1}), the entropy of the protein (S_{MM2}), and their coupling (S_{12}); the entropy of the solvent is included in the G_{np} term:

$$S_{\mathrm{QM/MM}} = S_{\mathrm{QM1}} + S_{\mathrm{MM2}} + S_{12} \quad (5)$$

As mentioned above, the entropy of the protein in the standard MM/PBSA approach is calculated from a normal-mode analysis of harmonic frequencies calculated at the MM level (typically on a truncated model of the protein). However, in our present investigation of a single proton transfer inside the protein and with fixed QM systems, it is unlikely that the entropy of the protein is significantly different for the two states. Therefore, we will ignore this term (S_{MM2}). Such an approach has been used in several previous applications of MM/PBSA.[65,66,67] Likewise, we will ignore the coupling term, S_{12}. On the other hand, it is possible that the internal entropy of the QM system is significantly different in the two states. Therefore, we estimated S_{QM1} from a normal-node analysis of harmonic frequencies of the isolated QM systems, optimised and calculated in vacuum.[38]

In analogy with the MM/PBSA approach, we must ensure that each interaction is only included once. All interactions with the solvent are included in the G_{Solv} and G_{np} terms; therefore, solvent molecules must be stripped away before the other terms are calculated. This leads to the following equation for the free energy:

$$G_{\mathrm{v1}} = \langle E_{\mathrm{QM1+ptch2}} \rangle - \langle E_{\mathrm{MM1\_noel1}} \rangle + \langle E_{\mathrm{MM12\_noel1}} \rangle + \langle G_{\mathrm{Solv}} \rangle + \langle G_{\mathrm{np}} \rangle - T\langle S_{\mathrm{QM1}} \rangle \quad (6),$$

where $E_{\mathrm{QM1+ptch2}}$ is the QM energy of the QM system in the field of a set of point charges, modelling the surrounding protein, but not any solvent molecules (possibly except crystal water molecules, see below), and omitting the self-energy of the point charges. Likewise, $E_{\mathrm{MM12\_noel1}}$ is the MM energy of the full protein, but without solvent molecules (again possibly except crystal waters), excluding electrostatic interactions with the quantum system. $E_{\mathrm{MM1\_noel1}}$ is the non-electrostatic MM energy of the QM system (truncated by H junction atoms) from Eqn. 3. We call this variant of our QM/MM-PBSA approach version 1.

There are two small disadvantages with such an approach: The QM system is not polarised by the solvent in the QM calculations and the electrostatic interactions with the surrounding protein include interactions with the H junction atoms and also some interactions between atoms that are connected by bonds, angles, or dihedrals (in variance with the standard MM approach). A simple way to solve these problems is to include all point charges in the QM calculation (also solvent water), but only when the wavefunction is optimised. Then, the point charges are removed and the energy is evaluated by a single SCF iteration (note, that this calculation must be done with the final integration grid size, 3 in Turbomole, not with the smaller iteration grid, m3). Thereby, the resulting energy, $E_{\mathrm{QM1,pol23}}$, includes only the internal QM energy of the QM system, but no interactions between the QM systems and surroundings. Consequently, the $E_{\mathrm{MM12\_noel1}}$ and $E_{\mathrm{MM1\_noel1}}$ terms in Eqn. 6 must also be modified:

$$G_{\mathrm{v2}} = \langle E_{\mathrm{QM1,pol23}} \rangle - \langle E_{\mathrm{MM1}} \rangle + \langle E_{\mathrm{MM12}} \rangle + \langle G_{\mathrm{Solv}} \rangle + \langle G_{\mathrm{np}} \rangle - T\langle S_{\mathrm{QM1}} \rangle \quad (7),$$

where E_{MM12} is the true MM energy of the full protein (i.e. including also all electrostatic

interactions and therefore identical to the energy calculated in the standard MM/PBSA approach in Eqn. 1) and E_{MM1} is the true MM energy of the QM system (including electrostatics) and with C junction atoms (i.e. the junction atoms are of the same type and in the same positions as in the full protein). It is important that the point charges of the QM system, used to calculate the E_{MM1} and E_{MM12} terms, are obtained in the field of the point charges of the surrounding protein and solvent (like $E_{QM1,pol23}$; but the point charges for the QM system are only fitted once, not for every MD snapshot). We call this variant of our QM/MM-PBSA approach version 2 in the following.

In the standard MM/PBSA approach,[3,27] all energy terms in Eqn. 1 are obtained as averages over a set of snapshots from a single MD simulation. Thus, even if both the reactants and products are studied, only a single set of MD snapshots (obtained either with the reactant or with the product) of the surrounding protein is studied. The philosophy behind this approximation is that the energies become much more stable, because the internal energy (E_{int} in Eqn. 2) cancels (for ligand-binding energies) and the time-consumption is reduced. However, this approach has been criticized, because it ignores the difference in geometry of the two states.[68] On the other hand, it has recently been suggested that even more computer time can be saved and more stable and accurate results can be obtained if minimised structures are used instead of snapshots from MD simulations.[69,70]

Therefore, we have tested three different approaches for the calculation of the QM/MM-PBSA free energies: In the first (MD2), we use different MD simulations for the reactant and product states. In the second approach, we use the same MD snapshots for the two QM states (MD1r or MD1p depending on whether we used the snapshots of the reactant or product state). Thereby, the MM self-energy of the protein outside the QM system cancels (i.e. the MM self-energy of system 2, but the MM interaction energy between system 1 and system 2 remains). In the third approach (MM), we instead use minimised structures (one structure for the reactant and one for the product), taken from the QM/MM optimised structures. The results presented in the tables are the averages from the reactant and product structures. The difference between these two results (the hysteresis) is typically ~10 kJ/mol, with a maximum of 15 kJ/mol.

Finally, crystal water molecules provide a problem for all PB approaches: Should they be considered as part of the protein (MMp) or as a part of the solvent (MMs)? We have tested both approaches in connection with the MM approach (in MD simulations, many crystal-water molecules exchange with solvent; therefore all water molecules are considered as solvent in those calculations).

In practice, the QM/MM-PBSA calculations with MD simulations were performed in the following way (see http://www.teokem.lu.se/~ulf/Methods/qmmm_pbsa.html for a more detailed description). We started from the structures previously optimised by QM/MM (H2ase) or quantum refinement (NIR).[18,19,38] To the latter structures, hydrogen atoms outside the QM system were added and optimised by conjugate-gradient minimisation until the root mean square force was less than 0.001 kcal/mol/Å. These structures were used directly as the MM structures in QM/MM-PBSA. The ESP charges for the QM systems were also obtained from these structures, but for NIR (to be consistent with the QTCP data[38]), a solvation shell of ~4000 water molecules was first added and optimised (together with the added hydrogen atoms) with simulated annealing and conjugate-gradient minimisation.

For the MD snapshots for QM/MM-PBSA, the proteins were further solvated in an octahedral water box, extending at least 9 Å from the protein (giving a total of 32062 atoms for H2ase and 62853 atoms for NIR). This system was then simulated with MD in the *NPT* ensemble (one atmosphere pressure and 300 K temperature) for 20 ps, restraining the heavy atoms by a harmonic force constant of 50 kcal/mol/Å^2 to the QM/MM structure. Thereafter, only the QM system was restrained and the system was simulated for another 50 ps in the *NPT* ensemble to allow the volume to equilibrate. After this, the QM system was moved back to the exact position in the QM/MM calculations (after translation and rotation; the QM system moves slightly during the constant-pressure simulation) and was fixed. Finally, the system was

equilibrated in the *NVT* ensemble for 200 ps and snapshots were saved every 10 ps for an additional time of 200 ps. Thus, following the QTCP approach[32,33] and in order to make the energies stable, we keep the geometry of the QM system fixed in the MD simulations. Thereby, we also avoid the need of MM parameters for the QM system.

Next, the energy terms in Eqns. 6 or 7 were calculated with Amber 9[40] and Turbomole 5.9 program packages.[52] First, the E_{MM12} energy was calculated by the MM/PBSA module of Amber. This required the set-up of topology files for the two states without any solvent water molecules (except crystal water molecules in the MMp simulations). This module also directs the calculations of the G_{np} term, using the molsurf program (included in the Amber distribution), and the PB G_{Solv} term using the Poisson–Boltzmann solver in Amber 9.[70,71] The E_{MM2_noel1} energy could be obtained with the MM/PBSA script after proper modifications. S_{QM1} was estimated from a normal-mode analysis of frequency obtained with Turbomole on a complex optimised in vacuum, as has been described before.[38] Likewise, $E_{QM1+ptch2}$ and $E_{QM1,pol23}$ were calculated with Turbomole, after the set-up of proper point-charge files. Finally, E_{MM_noel1} was taken directly from the corresponding QM/MM calculation, whereas E_{MM1} was obtained by sander calculations, using modified topology and coordinate files obtained by a direct truncation of the file for E_{MM12}. These two energies were calculated only once, because the QM systems were fixed in the MD simulations.

Results and Discussion

[Ni,Fe] hydrogenase

Hydrogenases are enzymes that catalyse the seemingly simple reaction[72]

$$H_2 \rightarrow 2\,H^+ + 2\,e^- \qquad (8).$$

There are several types of these enzymes. The [Ni,Fe] hydrogenases have an active site with one Ni ion coordinated to four Cys residues and a Fe ion that coordinates two of the same Cys ligands and also two CN^- ions and one CO molecule (Figure 2).[73] Many crystal structures of H2ases have been presented,[73] but several details of the reaction are still controversial because the reactants and products of Eqn. 8 are invisible in normal crystal structures. For example, the reaction involves protons, but it is not fully clear where these bind: In principle any of the four Cys ligands of Ni may be protonated and there are also other possible proton acceptors in the active site.

Many theoretical investigations of the H2ases have also been presented.[18,74,75,76,77,78] In principle, it should be possible to deduce the energetically most favourable protonation sites by theoretical methods. In practice, it is very hard to obtain converged energies in such calculations.[18] We have recently reported an attempt in that direction, using the QM/MM FEP method QTCP.[38] Such a method is very time-consuming. Therefore, we here investigate if we can reproduce those results with a less computer-intensive method, QM/MM-PBSA.

We study the transfer of a single proton between the Ni ligand Cys-546 and His-79 (Figure 2) and try to estimate the relative free energy of the state with the proton on Cys-546 (called the HID state) and the state with the proton on His-79 (called the HIP state). According to the QTCP calculations,[38] this free-energy difference is 31–49 kJ/mol (depending on the QM system), favouring the HIP state .

We have tested our new QM/MM-PBSA approach quite extensively on H2ase. Before discussing the results in detail, we note that the TS_{QM1} term in Eqns. 6 and 7 is constant in all calculations (because it is based on QM optimised structures in vacuum) and amounts to 4 kJ/mol in favour of the HID structure.[38] Furthermore, the zero-point energy favours the HID state by 5 kJ/mol. The basis-set and functional correction works in the opposite direction, favouring the HIP state by 5 kJ/mol. Thus the overall effect of these terms is 5 kJ/mol, favouring the HID state.[38] Thus, this term is relatively small and exactly the same term contributes to the QTCP free energy estimate (because the QM system is fixed there also).

Therefore, it does not contribute at all to the comparison with the QTCP results and it will not be further discussed.

Moreover, the G_{np} term (in Eqns. 6 and 7) turned out to be the same (within 1 kJ/mol) for the two states in all calculations, because it is obtained from the surface area of the whole protein and this area does not change much when a proton is moved by less than1 Å in the centre of the protein. Therefore, this term will not be listed or discussed below.

As discussed in the Methods section, we have tested two different versions of the QM/MM-PBSA method (Eqns. 6 and 7), differing in the treatment of the electrostatic interactions between the QM system and the protein. Moreover, we have performed energy calculations based on either the QM/MM minimised structure (MM) or on snapshots from MD simulations of the solvated protein. In the former case, we also tested whether crystal water molecules should be treated as a part of the protein (called MMp below) or as the solvent (MMs; in the MD simulations, many of the crystal water molecules exchanged with solvent molecules, so such an investigation was not possible). Finally, for the MD simulations, we calculated the energies either for different sets of coordinates for the HID and HIP states (MD2) or for the same set of coordinates for both states (which could be obtained either from the HID or the HIP simulations, called MD1d and MD1p below, respectively). Thus, we have 10 different estimates of the free energies, which are collected in Table 1.

If we concentrate on the total free energy difference to start with (ΔG_{v1} and ΔG_{v2} in Table 1), we see that the calculations with different MD structures (MD2) give very poor results with errors of ~200 kJ/mol). This is caused by a very large variation in the energies between the various snapshots (the standard deviation of the ΔG values is ~270 kJ/mol). On the other hand, all the other calculations (which are based on the same structure outside the QM region) give much more reasonable results, ranging from –19 to –35 kJ/mol, i.e. rather close to the QTCP reference, –37 kJ/mol with this QM system[38] (throughout this section, a negative energy means that the HIP state is most stable).

The difference between the two calculations based on the same MD snapshots (MD1d and MD1p) provides an estimate of the accuracy of the QM/MM-PBSA calculations. From Table 1, it can be seen that this difference amounts to ~7 kJ/mol, which is reasonable for this approximate method (the MM part of the QTCP calculations showed a hysteresis of 2 kJ/mol[38]). The root-mean-squared difference between the coordinates of all atoms in the two trajectories is 0.14 Å.

A further estimate of the accuracy of the QM/MM-PBSA results can be obtained by looking at the variation of the various terms for the 20 snapshots from the MD simulation, used in the energy calculations. The standard deviations of the various terms are also included in Table 1 (in brackets). It can be seen that the standard deviation of the total free energy in the four calculations based on MD snapshots is 6–8 kJ/mol. Of course, the standard deviation of the mean value is smaller by a factor of $\sqrt{20}$, i.e. 1–2 kJ/mol. Apparently, the precision of this method is not limited by the statistical uncertainty. It is notable that the standard deviation is almost a magnitude smaller than in ligand-binding studies (e.g. 47–62 kJ/mol[79]). The reason for this is that we study a reaction that only involves a small movement of one atom and takes place at the centre of the protein.

As was discussed in the Methods section, we have tested two variants of QM/MM-PBSA, differing in the way the QM system is polarised by the surroundings and how the electrostatics between the QM system and the surroundings are treated. In the first version (Eqn. 6), the QM system is polarised by the protein (and possibly by the explicit crystal water molecules, but not by the implicit solvent) and the electrostatics between QM system and MM part of the protein are treated in the QM calculation. In the second variant (Eqn. 7), the QM system is also polarised by (parts of) the solvent, which is more accurate. On the other hand, electrostatics between the QM system and the protein are treated by MM, which has both advantages and disadvantages. The bad thing is that the QM wavefunction has to be converted to a set of point charges, which of course is only an approximation. The good thing is that we fully avoid the

use of junction atoms with wrong positions and atom types, and that the electrostatics are treated in a more consistent way: Electrostatics between atoms separated by one or two bonds are omitted and electrostatic interactions between atoms three bonds apart are scaled down in the normal way of the Amber 1999 force field.

From Table 1, it can be seen that the change from version 1 to version 2 leads to changes in the QM and MM energies, but not in the solvation energy. The main difference is that we move the electrostatic interaction energy between the QM system and the protein from the QM energy in version 1 to E_{MM12} in version 2, but E_{MM1} also changes because electrostatics are included also in this energy in version 2 (and the type and position of the junction atoms change). It can be seen from Table 1 that the two versions give similar results (the difference is –7 to +3 kJ/mol) and we have too little data to judge which version gives the best results.

An advantage with version 2 is that the energies are easier to partition and understand. $\Delta E_{QM1,pol23}$ is the total internal energy of the QM system, polarised by the surroundings. As expected, it is almost constant (2–4 kJ/mol) and slightly favours the HID state. The standard deviation is less than 1 kJ/mol. The corresponding vacuum result is 4 kJ/mol, using the QM/MM geometries.[38] On the other hand, the HIP state is 19 kJ/mol lower in energy if the structures are optimised in vacuum.

Likewise, the difference in solvation energy between the HID and HIP states (ΔG_{Solv}) is rather constant in all calculations (–20 to –27 kJ/mol). The standard deviation is 2–5 kJ/mol if the same snapshot is used, but 90 kJ/mol if different MD simulations are used. Thus, this term is an important contribution to the stability of the HIP state. This is perhaps somewhat unexpected, considering that the active site is ~30 Å from the surface of the protein, but it comes from an increased dipole moment of the HID state (19 D compared to 14 D).

The E_{MM_v2} term (i.e. $E_{MM12} - E_{MM1}$) also stabilises the HIP state with a contribution similar to that of the solvation term (–2 to –18 kJ/mol). This is also the term that gives the main variation. It can be further divided into various MM energy components, which are shown in Table 2. It can be seen that this term comes mainly from electrostatic interactions. In particular, the electrostatic 1,4 interactions are large, indicating that the main effect comes from the MM part of the Cys-546 residue and from the neighbouring Ala-545 and Gly-547 residues. The same results were obtained in the QTCP calculations[38] and it indicates that the results may be improved by including these residues also in the QM system. Since E_{MM_v2} is a pure MM term, it can also be divided into contributions from each residue in the protein, as was also done for the QTCP energies.[38] The difference between the MMs and MMp calculations (i.e. whether the crystal waters were treated as part of the protein or solvent) comes entirely from the electrostatic energy.

To gain more statistics to compare the two versions of QM/MM-PBSA (Eqns. 6 or 7) and decide how the crystallographic water molecules should be treated in the MM calculations, MMs, MMp, and MD1 calculations were performed with the five larger QM systems, depicted in Figure 2. The results for these calculations are gathered in Table 3. Based on the mean absolute deviations (MAD) from the QTCP data, it seems that it is better to treat the crystal water molecules as a part of the protein (MAD drops from ~13 kJ/mol to ~4 kJ/mol). However, the two versions of QM/MM-PBSA still perform equally well, with MADs differing by less than 2 kJ/mol.

In conclusion, the QM/MM-PBSA method seems to work quite well, compared to QTCP. A reasonable approximation may be to take the average of the MD1d and MD1p results as the final result, leading to a MAD of 9–10 kJ/mol compared to the QTCP value. However, even better results can be obtained by using simply the QM/MM minimised structures, treating the crystal water molecules as a part of the protein (MAD = 4–6 kJ/mol).

Nitrite reductase

Copper nitrite reductase (NIR) is a bacterial enzyme that catalyses the one-electron

reduction of nitrite to gaseous NO:[80]

$$NO_2^- + e^- + 2\ H^+ \rightarrow NO + H_2O \quad (9).$$

It contains two copper ions, one electron-transfer blue-copper site and the catalytic site. In the latter, the copper ion is bound to three His ligands and a solvent molecule (Figure 3b). One of the His ligands, His-100, forms a hydrogen bond to the carboxylate group of Glu-279 and the solvent molecule forms a hydrogen bond to the carboxylate group of Asp-98. The status of the solvent molecule (water or OH^-) is important for the catalysis.[80] Therefore, we have in a previous article studied the protonation of this residue and the His-100 – Glu-279 pair with quantum refinement of crystal structures and various QM(/MM) calculations.[18] This gave detailed information about the structures of the various protonation states, but it was not possible to obtain reliable estimates of the relative energies of the four possible protonation states: Wat (protons on water and His-100), Hyd (protons on Asp-98 and His-100), Imm (protons on water and Glu-279), and Both (protons on Asp-98 and Glu-279), cf. Figure 4. On the contrary, different treatments of the electrostatics and solvation gave estimates that differed by up to ~100 kJ/mol.[18] Thus, this is a perfect test case for various methods to estimate accurate energies in proteins. Recently, we have studied this system with the QTCP method, which should provide a good estimate of the true free-energy differences between the four protonation states.[38] Therefore, we will use also this system as a test case for the QM/MM-PBSA approach, in order to gain more statistics and probe its performance on another system.

We have performed the same QM/MM-PBSA calculations for NIR as for H2ase, i.e. calculations based on the two versions of the approach and calculations based directly on quantum-refined minimised structures or on MD sampled snapshots, in the former case treating the crystal water molecules either as part of the protein or the solvent, and in the latter case, based either on the same or different sets of coordinates for both the reactant and product states. The results of these calculations are collected in Table 4.

As for H2ase, the MD2 results (based on different MD simulations) are poor, with MADs of 34–36 kJ/mol and with very large standard deviations (152–230 kJ/mol). They will not be further discussed.

The MD1 simulations gave much better results, with MADs of 8–14 kJ/mol. There is no significant difference between the performance of the calculations based on the reactant or product states (MD1r and MD1p). The average of the two MD1simulations gives consistently good results with a MAD of 8 kJ/mol and a maximum error of 16 kJ/mol for version 2.

However, the MD1r and MD1p calculations sometimes give quite different results, with a difference of up to 28 kJ/mol. In fact, the Hyd→ Both and Imm→ Wat reactions consistently give a higher difference (19–28 kJ/mol) than the other two reactions (0–6 kJ/mol). Likewise, the former two reactions give larger errors than the other two (the MAD for all methods except MD2 is 14–23, compared to 6–9 kJ/mol). The reason for this is that the Hyd→ Both and Imm→ Wat reactions both involve a proton transfer between Glu-279 and His-100, which is much more affected by the protein than the other proton transfer,[38] because Glu-279 forms a hydrogen bond to Lys-269. The standard deviations of the MD simulations are similar to those of H2ase, 3–8 kJ/mol, confirming that the precision of the QM/MM-PBSA approach for protein reactions is much better than that of MM/PBSA for ligand binding.

Likewise, the results based on minimised structures (MM) are of a similar quality to those obtained with the MD snapshots (the MADs are 7–22 and 8–14 kJ/mol, respectively. This is of the same order as the hystereses of the QTCP results, 5–15 kJ/mol, so it is hard to pinpoint any method as particularly good, but it indicates that also for NIR, a sampling of the dynamics of the surrounding protein and solvent is not really required to obtain reasonable energies. Interestingly, for NIR, the MMs results are better than the MMp results (the MADs are 6 kJ/mol lower for both versions of QM/MM-PBSA; for H2ase , the opposite was observed. Therefore, we must conclude that it cannot be predicted beforehand which of those two approaches is preferable. However, it seems most reasonable to treat all, or at most (those that are solvent-exposed) crystal water molecules as a part of the solvent.

Finally, we have also compared the two versions of QM/MM-PBSA (Eqn, 7). From Table 4, it can be seen that version 2 gives slightly lower MADs than version 1 for NIR, with a difference of 1–9 kJ/mol. Moreover, version 2 consistently gives lower standard deviations than version 1. Therefore this version is probably preferable.

Conclusions

We have developed a new method to obtain free energies for reactions in proteins, by combining the QM/MM[21,22] and MM/PBSA[3] approaches. We have applied and tested the approach on five proton-transfer reactions in [Ni,Fe] hydrogenase and copper nitrite reductase, i.e. to reactions that have been shown to be strongly influenced by the surroundings and therefore hard to treat with theoretical methods.[18,19,38] In fact, different quite reasonable methods can give results that differ by almost 100 kJ/mol and no simple method gave results with a MAD less than ~20 kJ/mol compared to a strict QM/MM free-energy perturbation method (QTCP) for these five reactions.[38]

In view of this, the QM/MM-PBSA results are quite impressive: All variants (except MD2) give MADs of 4–22 kJ/mol. This approaches the uncertainty of the QTCP data (especially for NIR) and it is actually not much worse than QM/MM-FE approach (MADs of 5 and 9 kJ/mol[38]), which is based on strict free-energy perturbations, although only at the MM level. Thus, we can conclude that QM/MM-PBSA is a promising approach to obtain reliable free energies in proteins at the QM/MM level, without actually performing FEPs, and clearly better than other approximate methods (e.g. the raw QM/MM data, QM data with or without point charges or with the surroundings treated by a reaction field, or a combination of the QM energies and a PB estimate of the electrostatic and solvation energies, which had MADs of 29–32, 14–36, and 32–39 kJ/mol, respectively[38]).

We have tested several variants of the QM/MM-PBSA approach. First, we tested whether we can use different MD simulations for the two states (MD2) or if it is better to use the same MD snapshots for the energy calculations (MD1), as normally is done in MM/PBSA. The results in Tables 1 and 4 clearly show that the MD2 approach give unreliable results and very large standard deviations (~200 kJ/mol). Thus, the MD2 approach should be avoided. Fortunately, the MD1 approach gave quite similar results, independently whether it was based on reactant or product state, with average MADs of 9 and 10 kJ/mol for H2ase and 14 and 11 kJ/mol for NIR, respectively. The best approach seems to take the average of the two MD1 results.

Second, we have tested two variants of the QM/MM-PBSA approach, which differ in the treatment of electrostatic interactions. In the first variant (Eqn. 6), electrostatic interactions between the QM system and the protein are estimated in the QM calculation, although the calculations involve some (1–2, 1–3, and 1–4) interactions that are normally not included. In the other version (Eqn. 7), the wavefunction is polarised by a point-charge model of the surrounding protein and solvent, but all electrostatic interactions are calculated by MM, including only the relevant interactions. From Tables 1, 3 and 4, it can be seen that the results of the two versions are similar for H2ase, but version 2 is slightly better for NIR (by 1–9 kJ/mol). Moreover, the results of version 2 are easier to interpret (to divide the MM energies in components and contributions from the various residues).

Finally, we tested whether it is possible to base the QM/MM-PBSA calculations on the QM/MM (or quantum-refined) structures, rather than on snapshots from MD simulations. The results in Tables 1, 3 and 4 show that such an approach gave results of nearly the same quality as the MD1 averages. With version 2, average MADs were 4–14 kJ/mol. This can be compared to averaged MD results (of version 2), which showed MADs of 8–10 kJ/mol for both proteins. This is a major improvement, compared to the raw QM/MM energies, which gave a MAD of 32 kJ/mol for H2ase.[38] Unfortunately, our calculations do not allow us to decide whether it is advantageous to treat the crystal water molecules as a part of the protein or the solvent.

An important question is how much time is gained by the QM/MM-PBSA approach, compared to the QTCP approach. Timings for the smallest (N) and largest (NCHACGR) QM systems for H2ase (cf. Figure 2) are shown in Table 5. It can be seen that QM/MM-PBSA with MD sampling uses only ~13% of the QTCP time for the free-energy calculation for both the small and large QM systems. However, both methods use a similar type of MD sampling (in fact, in this application, we used the same MD trajectories for both calculations) and the time for the sampling dominates that of the free-energy calculation. Therefore, the total saving in time is quite small, especially for the small QM system. However, if only minimised structures are used for QM/MM-PBSA (MM in Table 5), the saving is extensive, giving a time consumption of only ~1% of that of QTCP. This is the reason why this approach is so interesting. Finally, we note that the generation of the QM/MM structures also takes a significant time, which may dominate the total time consumption, especially for a large QM system. However, the largest savings of the QM/MM-PBSA approach can be expected in cases for which many intermediate states need to be studied to obtain a proper convergence of the free energies in QTCP – QM/MM-PBSA is an end-point approach that does not require any calculations on (uninteresting) intermediates.

In conclusion, we have shown that QM/MM-PBSA is a promising approximate method to estimate free energies in proteins, without doing extensive sampling of intermediate states as in FEP. In fact, it seems to give results with a MAD of ~10 kJ/mol, even if the original QM/MM structures are directly employed (i.e. without any sampling at all). It is also notable that with MD sampling, the standard deviations are quite small, 3–9 kJ/mol, giving a final statistical uncertainty of only 1–2 kJ/mol, which is much better than typical MM/PBSA results for ligand binding.[79] Therefore, we recommend the QM/MM-PBSA approach for the calculation of reaction free energies in proteins, especially if many different states are studied or if a complicated reaction coordinates are involved, making a traditional FEP approach difficult. A typical example of the latter type is protonation or deprotonation reactions, which often leads to the inversion of a whole chain of hydrogen bond interactions. Such complicated reactions can probably be studied in a single step by the QM/MM-PBSA approach.

Acknowledgements

This investigation has been supported by grants from the Swedish research council, the Wenner-Gren foundation, the Crafoord foundation, and by computer resources of Lunarc at Lund University.

References

1. D. L. Beveridge, F. M. Dicapua, Annual Review of Biophysics and Biophysical Chemistry 1989, 18, 431-492.
2. T. Hansson, J. Marelius, J. Åqvist, Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided mol. Design* **1998**, *12*, 27-35.
3. P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A . Case, T. E. Cheatham, Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Acc. Chem. Res. 2000, 33(12), 889-897.
4. Gilson, M. K; Honig, B. Calculations of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins, Struct. Funct. Gen.* **1998**, *4*, 7-18.
5. Hermann, R. B. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. J. Phys. Chem. 1972, 76, 2754-2759.
6. B. Kuhn, P. A. Kollman, J. Med. Chem. 2000, 43, 3786-3791.
7. J. Srinivasan, T. E. Cheatham 3rd, P. Cieplak, P. A. Kollman, D. A. Case, Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices, J. Am. Chem. Soc. 1998, 120, 9401-9409.
8. O. A. T. Donini, P. A. Kollman, Calculation and Prediction of Binding Free Energies for the Matrix metalloproteinases, J. Med. Chem. 2000, 43, 4180-4188.
9. J. Wang, P. Morin, W. Wang, P. A. Kollman, Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA, J. Am. Chem. Soc. 2001, 123, 5221-5230.
10. S. Huo, J. Wang, P. Cieplak, P. A. Kollman, I. D. Kuntz, Molecular Dynamics and Free Energy Analyses of cathepsin D-Inhibitor Interactions: Insight into Structure-Based Ligand Design, J. Med. Chem. 2002, 45, 1412-1419.
11. Brown, S. P; Muchmore, S. W. High-throughput calculation of protein-ligand binding affinities: Modification and adaptation of the MM-PBSA protocol to enterprise grid computing. J. Chem. Inf. Model. 2006, 46, 999-1005.
12. Jensen F, Norrby PO, Theor. Chem. Acc., 2003, 109, 1-7.
13. Deeth, RJ, Chem. Comm. 2006, 2551-2553.
14. Siegbahn, P. E. M.; Blomberg, M. R. A. *Chem. Rev.* **2000**, *100*, 421-437.
15. D. Bashford, D. A. Case, C. Dalvit, L. Tennant, P. E. Wright, (1993) Biochem., 32, 8045-8056.
16. Noodleman, L.; Lowell, T.; Han, W.-G.; Li, J; Himo, F. Chem. Rev. 2004, 104, 459-508.
17. G. M. Ullmann, L. Noodleman, D. A. Case (2002) J. Biol. Inorg. Chem., 7, 632-639.
18. N. Källrot, K. Nilsson, T. Rasmussen & U. Ryde (2005) "The structure of the catalytic copper site in nitrite reductase, studied by quantum refinement", Intern. J. Quant. Chem., 102, 520-541.
19. P. Söderhjelm & U. Ryde (2006) J. Mol. Struct. Theochem, 770, 199-219.
20. L. Olsen, T. Rasmussen, L. Hemmingsen & U. Ryde (2004) "Binding of benzylpenicillin to metallo-beta-lactamase: a QM/MM study", J. Phys. Chem. B. 108, 17639-17648
21. A. J. Mulholland, in *Theoretical biochemistry – processes and properties of biological systems* (Theoretical and computational chemistry, Vol. 9). L. A. Eriksson, ed. Amsterdam: Elsevier Science; 2001: 597-653.
22. U. Ryde (2003), Curr. Opin. Chem. Biol., 7, 136-142.
23. Raha, K.; Merz, K. M. J. Med. Chem. 2005, 48, 4558-4575.
24. Yu, N.; Yennawar, H. P.; Merz, K. M.; Acta Crystal. D 2005, 61, 322-332.
25. Nemoto, T.; Fedorov, D. G.; Uebayasi, M.; Kanazawa, K.; Kitaura, K.; Komeiji, Y. Comp. Biol. Chem. 2005, 29, 434-439.

26. Chandrasekhar, J; Jorgensen, W. L. J. Am. Chem. Soc. 1985, 107, 2974-2975.
27. Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. Acc. Chem. Res. 2001, 34, 72-79.
28. Muller, R. P.; Warshel, A. J. Phys. Chem. 1995, 99, 17516-17524.
29. Olsson, M. H. M.; Hong, G.; Warshel, A. J. Am. Chem. Soc. 2003, 125, 5025-5039.
30. Zhang, Y.; Liu, H.; Yang, W. J. Chem. Phys. 2000, 112, 3483-3492.
31. Ishida, T.; Kato, S. J. Am. Chem. Soc. 2003, 125, 12035-12048.
32. T. H. Rod & U. Ryde (2005) "Quantum mechanical free energy barrier for an enzymatic reaction", Phys. Rev. Lett., 94, 138302
33. T. H. Rod & U. Ryde (2005) "Free energy barriers at the density functional theory level: methyl transfer catalyzed by catechol O-methyltransferase", J. Chem. Theory Comput., 1, 1240-1251.
34. Yang W.; Bitetti-Putzer R.; Karplus M. J. Chem. Phys. 2004, 120, 9450-9453.
35. Díaz, N.; Suárez, D.; Merz, K. M.; Sordo, T. L. J. Med. Chem. 2005, 48, 780-791.
36. Gräter, F.; Schwarzl, S. M.; Dejaegere, A.; Fischer, S.; Smith, J. C. J. Phys. Chem. B 2005, 109, 10474-10483.
37. Wang, M. L.; Wong, C. F. J. Chem. Phys. 2007, 126, 026101.
38. M. Kaukonen, J. Heimdal, P. Söderhjelm, U. Ryde, J. Chem. Theory Comput., submitted.
39. Volbeda A.; Montet, Y.; Vernède, X.; Hatchikian, E. C.; Fontecilla-Camps, J. C. *International Journal of Hydrogen Energy* **2002**, *27*, 1449-1461.
40. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman (2006), AMBER 9, University of California, San Francisco.
41. K. Pierloot, J. O. A. De Kerpel, U. Ryde, M. H. M. Olsson, B. O. Roos (1998) J. Am. Chem. Soc. 120, 13156-13166.
42. E. T. Adman, J. W. Godden, S. Turley (1995) J. Biol. Chem. 270, 27458-27474.
43. Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098-3100.
44. Perdew J. P. *Phys. Rev. B* **1986**, *33*, 8822-8824.
45. Hehre, W. J; Radom L.; Schleyer P. v. R.; Pople J. A. In Ab initio molecular orbital theory, Wiley-Interscience, New York, **1986**.
46. A. Schäfer, H. Horn, and R. Alrichs, J. Chem. Phys., 97 (1992) 2571.
47. A. Schäfer, C. Huber & R. Ahlrichs, J. Chem. Phys. 100 (1994) 5829.
48. Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283-290.
49. Eichkorn K , Weigend F, Treutler, O , Ahlrichs R (1997) Theor Chem Acc. 97:119-126
50. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.
51. R. H. Hertwig and W. Koch, Chem. Phys. Lett., 268 (1997) 345.
52. Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346-354.
53. Cornell, W. D.; Cieplak, PI.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, PI.. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
54. Wang, J.; Cieplak, P.; Kollman, P. A. J. Comput. Chem. 2000, 21, 1049-1074.
55. E. Sigfridsson & U. Ryde (1998) "A comparison of methods for deriving atomic charges from the electrostatic potential and moments". J. Comp. Chem. 19, 377-395.
56. Ryde, U. *J. Comput.-Aided Mol. Design* **1996**, *10*, 153-164.
57. Ryckaert, J. P.; Ciccotti, G.; & Berendsen, H. J. C. (1977) J. Comput. *Phys*. **23,** 327-341.
58. W.L. Jorgensen, J. Chandrasekhar, J. Madura & M.L. Klein. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 79, 926-935 (1983).
59. Darden, T., York, D., Pedersen, L. J. Chem. Phys. 98:10089-10092, 1993.

60. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., Pedersen, L. G. A smooth particle mesh Ewald potential. J. Chem. Phys. 103:8577-8592, 1995.
61. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 1984, 81, 3684-3690.
62. Ryde, U.; Olsson, M. H. M. *Int. J. Quantum Chem.* **2001**, *81*, 335-347.
63. Reuter, NI.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem.* **2000**, *104*, 1720-1735.
64. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357-19363.
65. Huo, S.; Massova, I.; Kollman, P. A. J. Comput. Chem. 2002, 23, 15-27.
66. Brigo, A.; Lee, K. W.; Fogolari, F.; Mustata, G. I.; Briggs, J. M. Proteins, Struct. Funct. Bioinf. 2005, 59, 723-741.
67. Obiol-Pardo, C.; Rubio-Martinez, J. J. Chem. Inf. Model. 2007, 47, 134-142.
68. Pearlman, D. A. *J. Med. Chem.* **2005**, *48*, 7796-7807.
69. B. Kuhn, P. Gerber, T. Schulz-Gasch, M. Stahl, J. Med. Chem. 2005, 48, 4040-4048.
70. R. Luo, L. David and M.K. Gilson, (2002) J. Comput. Chem. 23, 1244-1253.
71. E. Gallicchio, M.M. Kubo and R.M. Levy, (2000) J. Phys. Chem. 104, 6271-6285.
72. P.M. Vignais, B. Billoud and J. Meyer, FEMS Microbiol. Rev. 25 (2001) 455-501.
73. A. Volbeda & JC Fontecilla-Camps, Coord. *Chem Rev (2005), 249, 1609-1619.*
74. Siegbahn PEM (2004) Adv. Inorg Chem 56:101-125
75. Stein, M.; Lubitz, W. *Curr. Opin. Chem. Biol.* **2002**, *6*, 243-249.
76. Fan, H.-J.; Hall, M. B. *J. Biol. Inorg. Chem.* **2001**, *6*, 467-473.
77. M. Bruschi, G. Zampella, P. Fantucci, L. De Gioia (2005) Coord. Chem. Rev. 249, 1620-1640.
78. Amara, P.; Volbeda, A.; Fontecilla-Camps, J. C.; Field, M. J. *J. Am. Chem. Soc.* **1999**, *121*, 4468-4477.
79. A. Weis, K. Katebzadeh, P. Söderhjelm, I. Nilsson, U. Ryde (2006) "Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field", J. Med. Chem., 49, 6596-6606
80. E. T. Adman, M. E. P. Murphy (2001) In: Messerschmidt A, Huber R, Poulos T, Wieghardt K (eds) Handbook of Metalloproteins. J. Wiley & Sons, Chichester, pp 1381-1390.

Table 1. The QM/MM-PBSA energies for the proton transfer between Cys-546 and His-79 in H2ase for the normal (N) size of the QM system. The energy terms (kJ/mol) are defined in Eqns. 6 and 7, except that $E_{MM\_v1} = E_{MM12\_noel1} - E_{MM1\_noel1}$ and $E_{MM\_v2} = E_{MM12} - E_{MM1}$. All energy terms are presented as the difference between the HIP and HID states. A negative number indicates that the proton prefers to be on His-79 (HIP state). The corresponding QTCP result for the total ΔG is –37 kJ/mol.[38] Two versions of QM/MM-PBSA have been used (Eqns. 6 and 7), and two types of coordinates: either from the QM/MM minimised structure (MM) of from snapshots of a MD simulations. In the former case, crystal water have (MMp) or have not (MMs) been considered as parts of the protein. For the MD coordinates, the QM/MM-PBSA calculations can be based on two different sets of coordinates for the reactant and product states (MD2), or on the same set of coordinates, which could either be those of the HID (MD1d) or the HIP (MD1p) state. MD1av is the average of those two calculations. Values in brackets are the standard deviation over the 20 snapshots.

Version	Coordinates	$\Delta E_{QM1+ptch2}$	ΔE_{MM_v1}	ΔG_{Solv}	ΔG_{v1}
1	MMs	10.6	-4.2	-25.9	**-19.4**
	MMp	-5.1	-4.7	-19.7	**-29.5**
	MD1d	11.3 (3.1)	-13.5 (4.0)	-27.1 (5.2)	**-29.3 (5.9)**
	MD1p	8.3 (3.2)	-18.2 (2.9)	-25.6 (5.5)	**-35.5 (8.1)**
	MD1av	9.8	-15.8	-26.4	**-32.4**
	MD2	0.4 (28)	170.1 (261)	-26.5 (90)	**143.9 (269)**
		$\Delta E_{QM1,pol23}$	ΔE_{MM_v2}	ΔG_{Solv}	ΔG_{v2}
2	MMs	3.6	-1.6	-25.9	**-24.0**
	MMp	3.6	-18.1	-19.7	**-34.2**
	MD1d	2.3 (0.4)	-1.3 (5.1)	-27.1 (5.2)	**-26.1 (6.2)**
	MD1p	2.1 (0.3)	-9.8 (4.6)	-25.6 (5.5)	**-33.3 (8.2)**
	MD1av	2.2	-5.5	-26.4	**-29.7**
	MD2	7.4 (5.2)	156.1 (263)	-26.5 (90)	**137.0 (272)**
QTCP					**-37.3**

Table 2. The various contributions to the QM/MM-PBSA $E_{MM\_v2} = \langle E_{MM12} \rangle - \langle E_{MM1} \rangle$ energy term for the results of Table 1, version 2. A negative number indicates that the HIP state is stabilized by the interaction. The bonded energy in the first column is the sum of the classical bond, angle and dihedral energies. The second and third contain the classical Coulomb interaction. The van der Waals energies are shown in the 4th and 5th columns. For comparison, the corresponding QTCP results are also included.[38]

	bonded	1–4 electrostatics	Electrostatics	1–4 van der Waals	van der Waals
MMs	-1.7	-12.0	14.4	-0.4	-2.1
MMp	-1.7	-12.0	-1.5	-0.4	-2.6
MD1d	-1.3 (1.2)	-11.6 (1.2)	14.7 (3.8)	0.3 (0.4)	-3.4 (3.7)
MD1p	-2.0 (1.1)	-11.9 (1.3)	11.3 (3.2)	-0.1 (0.3)	-7.1 (2.9)
QTCP	-1.5	-12.1	+11.8	+0.1	-5.2

Table 3. The QM/MM-PBSA energies as a function of the size of the QM system size. The calculations were performed on the six QM systems (N, NR, NACG, NACGR, NCHACG and NCHACGR) in Figure 2, using the MMp, MMs (crystal water molecules were treated as a part of the protein or the solvent), and MD methods (standard deviations in brackets) and the two versions of QM/MM-PBSA (Eqns. 6 and 7). The energy terms are the same as in Table 1, but the difference from the QTCP energies ($\Delta\Delta E$; –37.3, –30.6, –40.5, –41.1, –48.7, and –43.9 kJ/mol, respectively)[38] and the mean absolute deviations from those values (MAD) are also included.

QM system	Version	Coordinates	$\Delta E_{QM1+ptch2}$	ΔE_{MM_v1}	ΔG_{Solv}	ΔG_{v1}	$\Delta\Delta E$	MAD
N	1	MMs	10.6	-4.2	-25.9	**-19.4**	17.9	
NR			12.1	-4.9	-26.5	**-19.3**	11.3	
NACG			9.4	-3.1	-31.2	**-24.5**	16.0	
NACGR			7.4	-2.4	-31.6	**-26.5**	14.6	
NCHACG			-10.9	-1.2	-33.8	**-45.9**	2.8	
NCHACGR			-12.3	-1.1	-11.8	**-25.4**	18.5	13.5
N	1	MMp	-5.1	-4.7	-19.7	**-29.5**	7.8	
NR			-6.4	-5.3	-20.2	**-31.9**	-1.3	
NACG			-8.1	-5.3	-22.5	**-34.1**	6.4	
NACGR			-9.8	-3.0	-26.3	**-39.1**	2	
NCHACG			-33.1	-0.6	-27.7	**-61.4**	-12.7	
NCHACGR			-33.6	-0.5	-11.7	**-45.9**	-2.0	5.4
N	1	MD1av	9.8 (3.2)	-15.8 (3.5)	-26.4 (5.3)	**-32.4 (7.0)**	4.9	
NR			8.7(2.8)	2.6 (3.3)	-25.8 (4.0)	**-14.5 (4.6)**	16.1	
NACG			4.7 (4.3)	-4.9 (3.0)	-26.3 (4.2)	**-26.4 (8.0)**	14.1	
NACGR			4.8 (3.6)	5.2 (2.3)	-29.7 (5.1)	**-19.8 (6.2)**	21.3	
NCHACG			-14.7(3.0)	-3.4 (3.7)	-29.6 (3.2)	**-47.6 (6.4)**	1.1	
NCHACGR			-16.1 (3.2)	-1.8 (3.0)	-27.7 (2.3)	**-45.6 (4.4)**	-1.7	9.3
			$\Delta E_{QM1,pol23}$	ΔE_{MM_v2}		ΔG_{v2}		
N	2	MMs	3.6	-1.6	-25.9	**-24.0**	13.3	
NR			-28.8	30.8	-26.5	**-24.5**	6.1	
NACG			-13.2	17.7	-31.2	**-26.8**	13.7	
NACGR			-45.2	44.4	-31.6	**-32.4**	8.7	
NCHACG			-31.4	30.3	-33.8	**-34.9**	13.8	
NCHACGR			-62.2	54.7	-11.8	**-19.3**	24.6	13.4
N	2	MMp	3.6	-18.1	-19.7	**-34.2**	3.1	
NR			-28.8	14.3	-20.2	**-34.7**	-4.1	
NACG			-13.2	-0.7	-22.5	**-36.4**	4.1	
NACGR			-45.2	26.6	-26.3	**-44.9**	-3.8	
NCHACG			-31.4	9.1	-27.7	**-50.0**	-1.3	
NCHACGR			-62.2	34.3	-11.7	**-39.6**	4.3	3.5
N	2	MD1av	2.2 (0.4)	-5.5 (4.8)	-26.4 (5.3)	**-29.7 (7.2)**	7.6	
NR			-25.2 (0.9)	27.3 (4.2)	-25.8 (4.0)	**-23.7 (5.0)**	6.9	
NACG			-13.9 (0.5)	11.1 (6.3)	-26.3 (4.2)	**-29.0 (8.7)**	11.5	
NACGR			-42.1 (1.3)	39.7 (4.6)	-29.7 (5.1)	**-32.1 (7.1)**	9.0	
NCHACG			-30.0 (1.2)	24.6 (5.6)	-29.6 (3.2)	**-34.9 (7.1)**	13.8	
NCHACGR			-57.9 (2.7)	50.4 (4.8)	-27.7 (2.3)	**-35.3 (6.2)**	8.6	9.6

Table 4. Relative free energies of the four protonation states in NIR calculated with QM/MM-PBSA. Results for the corresponding QTCP calculations are included as a reference.[38] The two versions of QM/MM-PBSA have been used (Eqns. 6 and 7), and two types of coordinates: either from the QM/MM minimised structure (MM) or from snapshots of a MD simulations. In the former case, crystal water molecules have (MMp) or have not (MMs) been considered as parts of the protein. In the latter case, the calculations were based on two different sets of coordinates for the reactant and product states (MD2), or on the same set of coordinates, which could either be the reactant (MD1r) or the product (MD1p) set. MD1av is the average of those two values. Values in brackets are the calculated standard deviation over the 20 snapshots. The last column gives the mean absolute deviation (MAD) compared to the QTCP results.

Version	Coordinates	Hyd→ Both	Hyd→ Wat	Both→ Imm	Imm→ Wat	MAD
1	MMs	17.3	54.9	42.7	-6.0	16.3
	MMp	1.9	47.1	30.2	12.4	22.3
	MD1r	33.5 (6.8)	51.2 (5.0)	33.7 (4.5)	11.0 (5.7)	14.2
	MD1p	11.0 (7.8)	47.2 (2.9)	28.2 (3.6)	-16.9 (7.5)	13.2
	MD1av	22.3	49.2	31.0	-3.0	13.7
	MD2	71.6 (152)	109.3 (220)	48.0 (188)	-10.3 (230)	33.9
2	MMs	32.9	54.0	37.6	-20.2	7.4
	MMp	21.8	35.3	24.0	-12.9	13.0
	MD1r	48.3 (4.5)	54.4 (2.8)	30.1 (3.1)	-5.8 (3.6)	13.7
	MD1p	26.1 (3.6)	51.4 (2.7)	29.8 (2.9)	-25.1 (4.3)	8.1
	MD1av	37.2	52.9	30.0	-15.5	8.1
	MD2	76.0 (155)	110.4 (206)	32.9 (186)	1.5 (228)	35.7
	QTCP	37.0	42.2	35.3	-31.5	0.0

Table 5. Approximate timings (CPU hours) for the QTCP method and the QM/MM-PBSA approach with either MD sampling (MD) or only MM minimisation (MM). Timing for the smallest (N) and largest (NCHACGR) QM systems are given.

QM system	N			NCHACGR		
Approach	QTCP	MD	MM	QTCP	MD	MM
QM/MM	70	70	70	220	220	220
MD sampling	100	100	0	100	100	0
Free energy calculation	15	2	1	60	8	2

Figure 1. The general structure of [Ni,Fe] hydrogenase, with the two subunits (S and L) depicted, as well as the quantum system and the three iron–sulphur clusters.

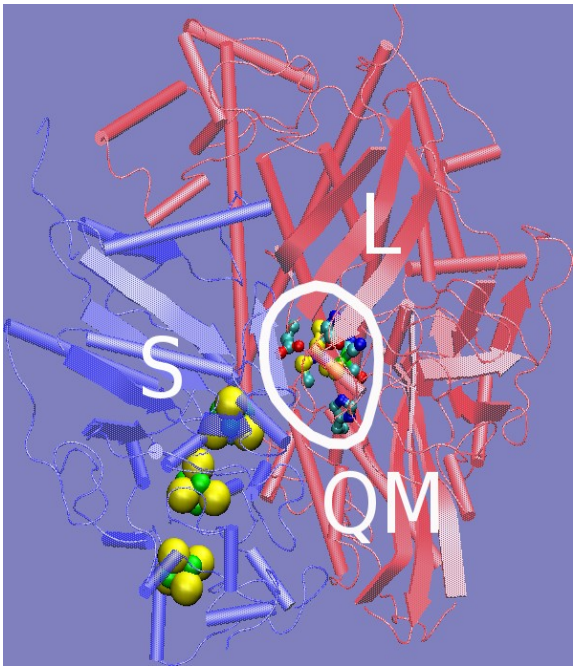
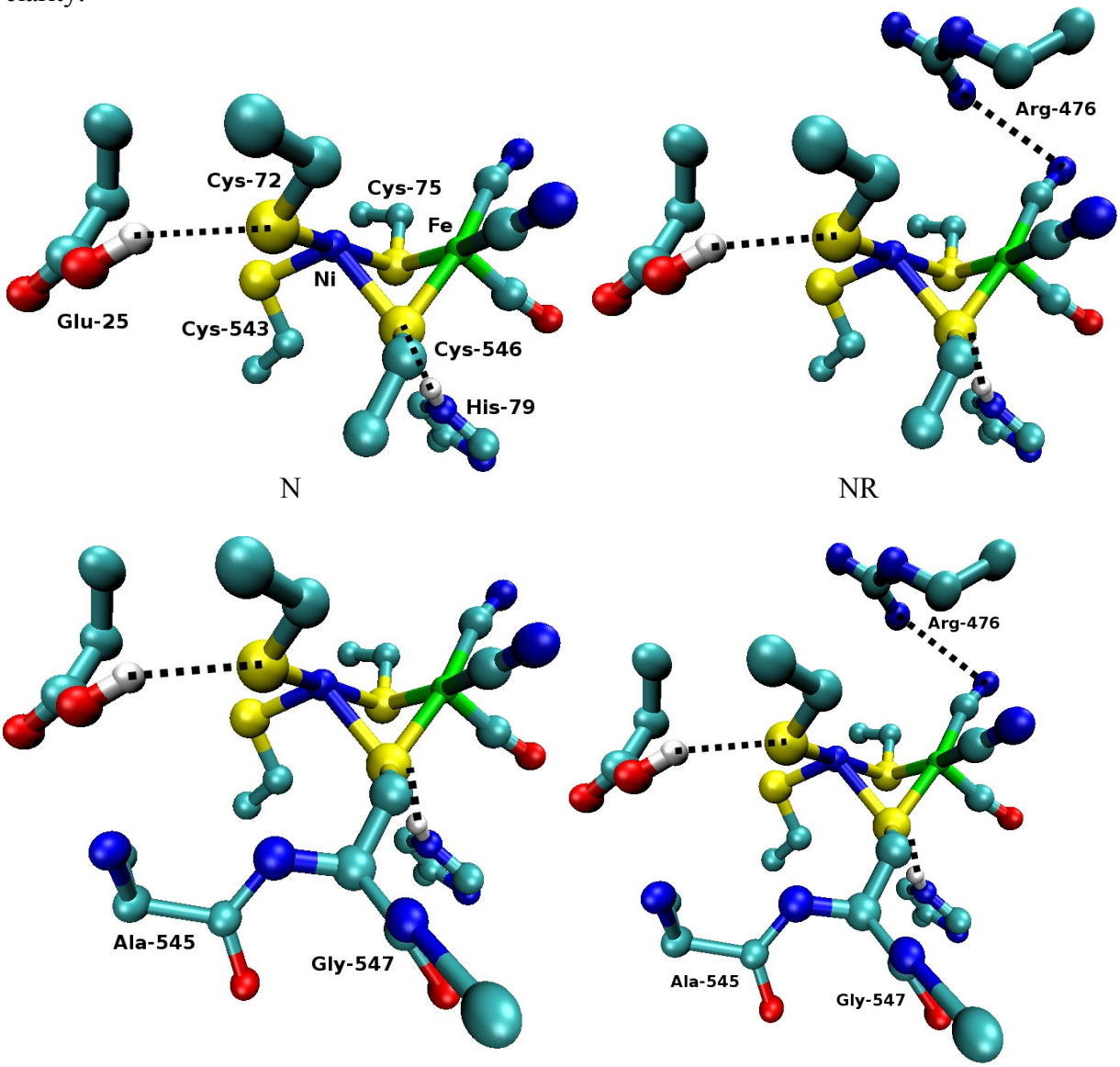
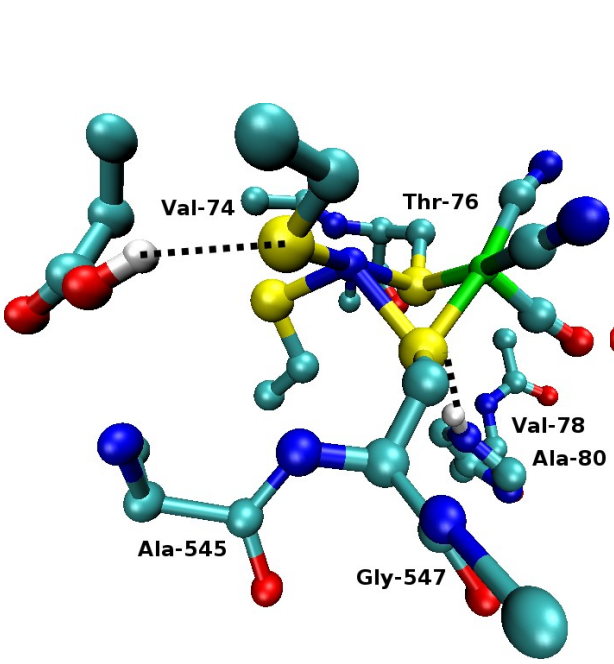
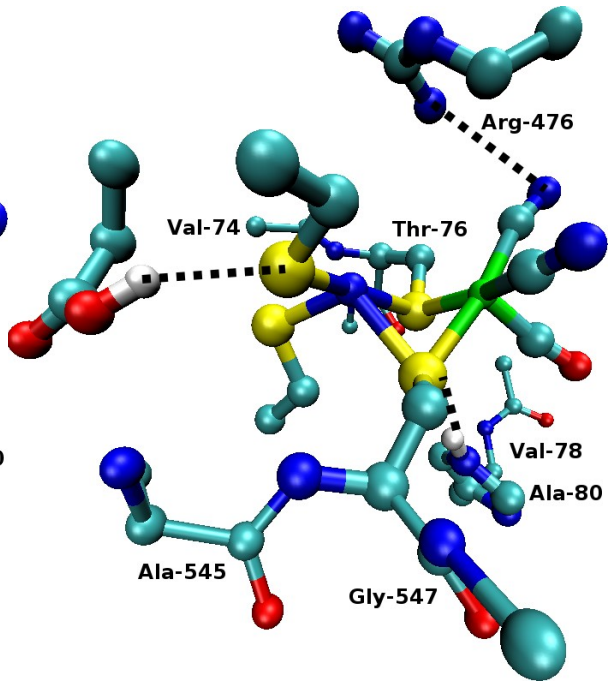


Figure 2. The six different QM systems used for [Ni,Fe] hydrogenase, N, NR, NCAG, NCAGR, NCHACG, and NCHACGR. For the N model, all active-site residues are marked. In the other five models, only additional groups are marked. Hydrogen atoms are omitted for clarity.



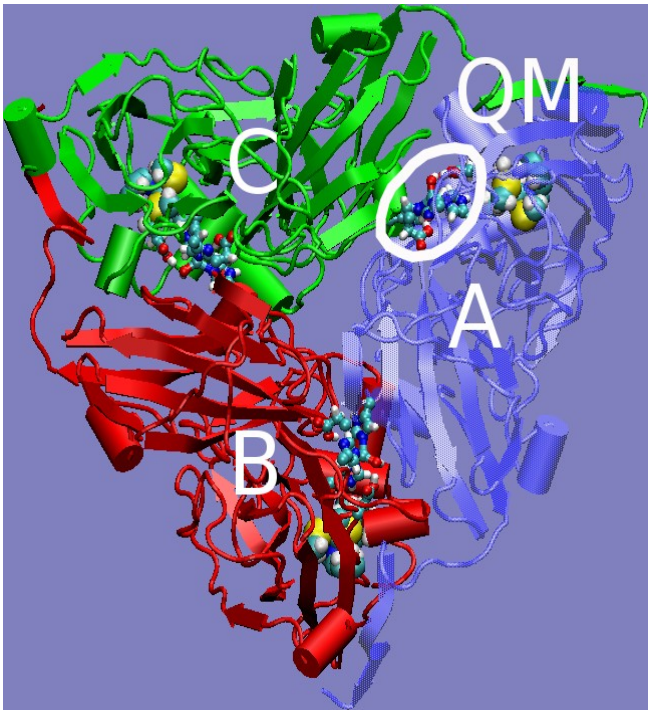


NCHACG



NCHACGR

Figure 3. (a) The general structure of Cu nitrite reductase, with the three subunits (A, B, and C) depicted, as well as the quantum system. (b) A detailed picture of the normal (N) quantum system in the Both state.



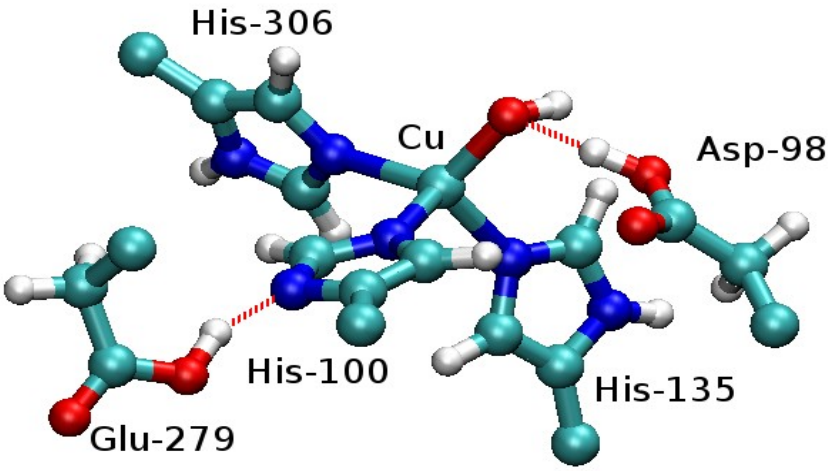


Figure 4. The QM system used for the active site of nitrite reductase in this study. The moving protons are included in the ellipses of the figure. The various protonation combinations are depicted in the lower part of the figure.