JCTC Journal of Chemical Theory and Computation

Proton Transfer at Metal Sites in Proteins Studied by Quantum Mechanical Free-Energy Perturbations

Markus Kaukonen, Pär Söderhjelm, Jimmy Heimdal, and Ulf Ryde*

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

Received December 22, 2007

Abstract: Catalytic metal sites in enzymes frequently have second-sphere carboxylate groups that neutralize the charge of the site and share protons with first-sphere ligands. This gives rise to an ambiguity concerning the position of this proton, which has turned out to be hard to settle with experimental, as well as theoretical, methods. We study three such proton-transfer reactions in two proteins and show that, in [Ni,Fe] hydrogenase, the bridging Cys-546 ligand is deprotonated by His-79, whereas in oxidized copper nitrite reductase, the His-100 ligand is neutral and the copper-bound water molecule is deprotonated by Asp-98. We show that these reactions strongly depend on the electrostatic interactions with the surrounding protein and solvent, because there is a large change in the dipole moment of the active site (2−6 D). Neither vacuum quantum mechanical (QM) calculations with large models, a continuum solvent, or a Poisson−Boltzmann treatment of the surroundings, nor combined QM and molecular mechanics (QM/MM) optimizations give reliable estimates of the proton-transfer energies (mean absolute deviations of over 20 kJ/mol). Instead, QM/MM free-energy perturbations are needed to obtain reliable estimates of the reaction energies. These calculations also indicate what interactions and residues are important for the energy, showing how the quantum system may be systematically enlarged. With such a procedure, results with an uncertainty of ∼10 kJ/mol can be obtained, provided that a proper QM method is used.

Introduction

The understanding of reaction mechanisms of enzymes has been a major goal in biochemistry for a long time. Most enzymes contain ionizable groups in their active site, and they often have perturbed properties (e.g., acid constants) compared to the same group in a water solution. This is especially pronounced in the neighborhood of metal sites, owing to the strong electrostatic field from the metal. Therefore, it is often crucial for the understanding of enzyme mechanisms to predict the acid constants of active-site residues, or at least the location of protons in the active site. This is a formidable task both for experimental and theoretical methods, and therefore many different methods have been developed with this aim.

A common way to study the reaction mechanism in protein is to perform a quantum mechanical (QM) calculation of only a few residues from the active site.[1] Such a procedure has the advantage of allowing for calculations with big basis sets and the inclusion of zero-point energies. On the other hand, the surrounding protein is only modeled by a few explicit residues, whereas the rest of the protein is either ignored or modeled as a featureless continuum with a dielectric constant of ∼4. Likewise, the dynamics of the protein are ignored, and entropic and thermal effects are estimated from a harmonic analysis of the normal-mode frequencies of the modeled system, a convenient but approximate approach.

At the next level of approximation, the whole enzyme is included in the calculations. This can be done in different ways. First, the whole protein may be described by molecular mechanics (MM) or the active site by QM methods and the surrounding protein by a point-charge model or by MM (QM/MM methods).[2–4] Second, a single conformation (a mini-

* Corresponding author tel.: +46 - 46 2224502; fax: +46 - 46 2224543; e-mail: Ulf.Ryde@teokem.lu.se.

10.1021/ct700347h CCC: $40.75
Published on Web 05/21/2008

mized or experimental structure) or a thermodynamic ensemble of structures can be studied. Finally, the investigation may be restricted to enthalpies or free energies may be estimated by either strict statistical mechanics methods (e.g., free-energy perturbations or thermodynamic integration) or more approximate methods (e.g., from the change in solvent-exposed surface area).

One popular way to estimate acid constants in proteins is to consider only electrostatic effects, estimated by the solution of the Poisson−Boltzmann (PB) equation.[5–9] The charges of the active site are typically obtained from a QM calculation, and the solvation energies are combined with the corresponding QM energies. Normally, only a single structure is considered, and this is compensated for by the use of an effective dielectric constant (ϵ) of 4 or higher inside the protein.

The most accurate estimates of free energies in proteins are obtained by free-energy perturbations (and similar methods) based on a QM description of at least part of the protein. Several approaches along these lines have been suggested, differing in the level of QM calculations, the interaction between the QM and MM systems, and the methods to perform the sampling and estimate the free energies.[10–19]

A problem with all methods that include explicitly the surrounding protein is the magnitude of the electrostatic interactions: Electrostatic atom−atom interactions are often ∼50 kJ/mol in a point-charge approximation, but the great majority of these interactions cancel. This makes the calculations sensitive to the treatment of solvation and the value of the dielectric constant in continuum approaches. This was succinctly illustrated in a recent study of the active copper site in nitrite reductase.[8] In this, the position of a proton in a hydrogen bond between the imidazole group of a histidine (His) copper ligand and the carboxylate group of a glutamate (Glu) residue was studied. Although the two structures differ only in the movement of a proton by less than 0.7 Å, estimates of the energy difference of the two states with different methods differed by up to 93 kJ/mol. In fact, a scaling of the point-charge model of the surrounding protein with a factor (effective dielectric constant) of 4 changed the relative energy by 88 kJ/mol. Likewise, theoretical estimates of pK_a values of metal ligands may easily be off by over 10 pK_a units.[20]

In this paper, we study the transfer of a proton within hydrogen bonds close to active-site metal ions with various theoretical methods, ranging from pure QM methods to a QM/MM free-energy perturbation approach (QTCP).[18,19] We study two test cases, copper nitrite reductase, mentioned above, and the transfer of a proton between a bridging cysteine ligand and a nearby His residue in [Ni,Fe] hydrogenase. By comparing the QTCP results with those obtained with cheaper methods, we may discuss the usefulness of various approximations. We also suggest a method to decide the size of the QM system and discuss whether large proteins may be truncated in the calculations.

Methods

QM/MM Calculations. The QM/MM calculations were carried out with the ComQum program.[21,22] In this approach, the protein and solvent are split into three subsystems: The QM system (system 1) contains the most interesting atoms and is relaxed by QM methods. System 2 consists of all residues within 15 Å of any atom in system 1 and is relaxed by a full MM minimization in each step of the QM/MM geometry optimization (in the QM/MM_free calculations; in the other calculations, only system 1 was optimized). Finally, system 3 contains the remaining part of the protein and surrounding solvent molecules and is kept fixed at the original (crystallographic) coordinates.

In the QM calculations, system 1 is represented by a wave function, whereas all of 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 QM 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 QM system is capped with hydrogen atoms (H junction atoms), the positions of which are linearly related to the corresponding carbon atoms (C junction atoms) in the full system.[21,23] Charges on atoms bound to junction atoms are zeroed, and the charges are evenly distributed on the other MM atoms in that residue.[21,22]

The total QM/MM energy in ComQum is calculated as:

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM1+ptch}} - E_{\mathrm{MM1}} + E_{\mathrm{MM123}} \qquad (1)$$

where $E_{\mathrm{QM1+ptch}}$ is the QM energy of the QM system truncated by hydrogen atoms and embedded in the set of point charges (but excluding the self-energy of the point charges). E_{MM1} is the MM energy of the QM system, still truncated by hydrogen atoms, but without any electrostatic interactions. Finally, E_{MM123} is the classical energy of all atoms in the system with C junction 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,[24] errors caused by the truncation of the QM system should cancel.

The geometry optimizations were continued until the energy change between two iterations was less than 2.6 J/mol (10^{-6} au) and the maximum norm of the Cartesian gradients was below 10^{-3} au. Restrained structures were optimized with a harmonic restraint with a force constant of 9375 kJ/(mol·Å^2) = 1 au, giving a deviation of less than 0.001 Å in the final structure.

QM Calculations. To be consistent with our earlier calculations,[8,25] the QM calculations on the two proteins were slightly different. For [Ni,Fe] hydrogenase (H2ase), the calculations were performed with the Becke 1988−Perdew 1986 (BP86) density functional[26,27] together with the 6-31G* basis set[28] for H, C, N, O, and S and the DZP basis set for Fe and Ni.[29,30] The calculations were sped up by expanding the Coulomb interactions in auxiliary basis sets, the resolu-

tion-of-identity approximation.[31,32] For nitrite reductase (NIR), we instead used the three-parameter hybrid B3LYP method, as implemented in the Turbomole package.[33–35] These calculations employed the 6-31G* basis set for all atoms,[9] except for copper, for which we used the DZP basis sets of Schäfer et al.,[29,30] enhanced with *p*-, *d*-, and *f*-type functions with exponents of 0.174, 0.132, and 0.39 (DZpdf). All QM calculations were performed with the Turbomole 5.8 and 5.9 software.[35]

To obtain more accurate energies, a single-point B3LYP calculation was performed for the optimized structures. In these calculations, the 6-311+G(2d,2p) basis set was used for the light atoms,[9] whereas the DZP basis set for the Fe and Ni was enhanced with *s*-, *p*-, *d*-, and two *f*-type functions with the following exponents: 0.013 772, 0.041 843, 0.1244, 2.5, and 0.8 for Fe and 0.145 763, 0.146 588, 0.044 447, 0.1458, 6.74, and 1.04 for Ni, and the basis set for Cu was enhanced with *s*, *p*, and *f* functions with exponents of 0.0155, 0.046 199, and 3.55.

Vibrational frequencies were obtained with the same method and basis sets as for the geometry optimizations for H2ase (BP86/6-31G*). The QM systems were first reoptimized in a vacuum, and then vibrational frequencies were calculated for those structures. From these frequencies, zero-point energies and thermal corrections to the Gibbs free energy (at 298 K and 1 atm pressure) were obtained, using an ideal-gas approximation.[36] In nitrite reductase, a water molecule hydrogen-bonded to the glutamate model in the QM system had to be added to prevent the QM system from relaxing to the Both state.

In some cases, solvation effects were estimated by single-point calculations using the continuum conductor-like screening model (COSMO).[37,38] These calculations were performed at the same level of theory as the geometry optimization and with default values for all parameters (implying a water-like probe molecule) and a dielectric constant (ϵ) of 4. For the generation of the cavity, a set of atomic radii has to be defined. We used the optimized COSMO radii in Turbomole (1.30, 2.00, 1.83, and 1.72 Å for H, C, N, and O, respectively, and 2.223 Å for the metals).[39] These energies are called $E_{QM1+\epsilon=4}$ in the following.

MM Calculations. All MM calculations were run with the sander module in the AMBER 8 software,[40] using the Amber 1999 force field.[41,42] The QM system was represented by charges fitted to the electrostatic potential, calculated in 115 000−260 000 points selected at random around the QM system up to a distance of 8 Å and outside the van der Waals envelope of the QM system. The fit used singular-value decomposition methods to ensure that all of the fitted charges are significant. The charges were constrained to exactly reproduce the QM dipole moment and quadrupole moments, as well as the Boltzmann-weighted electrostatic potential, the CHELP-BOW method.[43] The charge on each 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.[22] 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 containing junctions have noninteger 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.

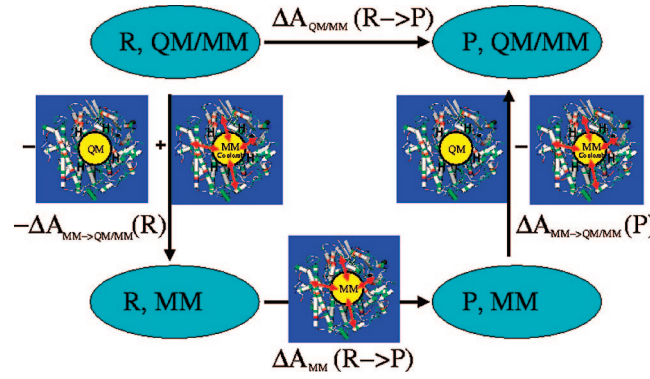

Figure 1. The QTCP cycle and energies.[18,19] We want to obtain the QM/MM free-energy difference between the reactant (R) and the product (P, upper horizontal line, $\Delta A_{QM/MM}(R \rightarrow P)$). It is obtained by calculating the computationally much cheaper MM free-energy difference ($\Delta A_{MM}(R \rightarrow P)$, lower horizontal line), as well as the two QM → QM/MM correction terms ($\Delta A_{MM \rightarrow QM/MM}$, the two vertical lines). All energies can be obtained by MD at the MM level only.

All bond lengths involving hydrogen atoms were constrained by the SHAKE algorithm.[44] The water solvent was described explicitly using the TIP3P model.[45] The electrostatics were treated with the particle-mesh Ewald method[46,47] with a grid size of 80^3 Å, a fourth-order B-spline interpolation, a tolerance of 10^{-5}, and a real-space cutoff of 8 Å. The temperature was kept constant at 300 K using the Berendsen weak-coupling algorithm[48] with a time constant of 1 ps. The molecular dynamics (MD) time step was 2 fs, and the nonbonded pair list was updated every 50 fs.

QTCP and QM/MM-FE. QTCP (QM/MM thermodynamic cycle perturbation) is a method to estimate free-energy differences at the QM/MM level.[18,19] It is based on the thermodynamic cycle in Figure 1. Thus, we estimate the QM/MM free-energy difference between the reactant (R) and the product (P), $\Delta A_{QM/MM}(R \rightarrow P)$, by performing three free-energy perturbations. The first one (the horizontal line in Figure 1) is the corresponding free-energy difference at the MM level, $\Delta A_{MM}(R \rightarrow P)$ (if the difference in free energy between the reactant and product is too large, this perturbation may be done in several smaller steps). The second and third perturbations change the description of the R and P states from the MM level to the QM/MM level (the vertical lines in Figure 1). Summing up these three energies, we obtain the free-energy change from the reactant to product at the QM/MM level:

$$\Delta A_{QTCP}(R \rightarrow P) = \Delta A_{MM}(R \rightarrow P) + \Delta A_{MM \rightarrow QM/MM}(P) - \Delta A_{MM \rightarrow QM/MM}(R) \quad (2)$$

The advantage with this approach is that the phase space needs to be sampled only at the MM level, thereby avoiding the extremely time-consuming sampling at the QM/MM level. A similar approach has been used by Warshel and co-workers,[12,13] but it led to severe convergence problems. We solved these by keeping the QM system fixed during

the simulations.[18,19] Test calculations at the semiempirical level have shown that this does not affect the final free energies significantly.[49] By this procedure, we also avoid the need of bonded MM parameters for the QM system.

In practice, the QTCP calculations were performed in the following way (see http://www.teokem.lu.se/~ulf/Methods/qtcp.html for a more detailed description).[18,19] We start from the structures obtained by the QM/MM (for H2ase,[25] keeping system 2 fixed) or quantum-refinement (nitrite reductase[8]) optimizations. Then, the proteins were better solvated by using an octahedral water box, extending at least 9 Å from the protein. This system was then minimized by 100 steps of MM minimization, keeping the QM atoms fixed and restraining the heavy atoms in the protein by a harmonic force constant of 100 kcal/(mol·Å^2). This was followed by a MD equilibration in the NPT ensemble (1 atm pressure and 300 K temperature) for 70 ps, restraining the QM atoms by a harmonic force constant of 500 kcal/(mol·Å^2) to the QM/MM structure. For the truncated H2ase, heavy atoms near the cut were restrained by harmonic forces of 10 kcal/(mol·Å^2). After this equilibration of the volume of the simulation cell, the QM system was moved back to the exact position in the QM/MM calculations (the QM system moves slightly during the constant-pressure simulation) and was fixed at this position for the remaining simulations. Finally, the system was equilibrated in the NVT ensemble for 200 ps, and snapshots were saved every 10 ps for an additional time of at least 200 ps.

Finally, free energies were calculated from these snapshots (strictly speaking, Helmholtz free energies are calculated, rather than Gibbs free energies, but the difference between these two quantities is small for proteins).[50] The free energy at the MM level ($\Delta A_{MM}(R \rightarrow P)$ in eq 2) was calculated by a free-energy perturbation (FEP), in which the coordinates and the charges of the QM atoms (with C junction atoms) were changed from the R state to the P state (or vice versa) and the energy difference was calculated:

$$\Delta A_{MM}(R \rightarrow P) = -k_B T \ln\left\langle e^{-\frac{E(P)-E(R)}{k_B T}} \right\rangle_R \qquad (3)$$

where k_B is the Boltzmann constant, T is the temperature, and the angular brackets indicate an average over a MD ensemble, sampled for the R state (Figure 1). An estimate of the accuracy of this free energy is obtained by reversing the R and P states, using snapshots from the simulation of the P state instead (the hysteresis).

Likewise, the vertical MM → QM/MM free energies are obtained by a FEP, in this case for the same state described at two different levels of theory (Figure 1). As has been described before,[19] this is accomplished by the following equation:

$$\Delta A_{MM \rightarrow QM/MM}(X) = -k_B T \ln\left\langle e^{-\frac{E'_{QM1+ptch}(X)-E_{el12}(X)}{k_B T}} \right\rangle_X \qquad (4)$$

where X is either the R or P state and $E'_{QM1+ptch}$ is essentially the same energy as $E_{QM1+ptch}$ in eq 1 (i.e., the QM energy of the QM system embedded in a point-charge model of the surrounding protein, but excluding the self-energy of the point charges), with the small exception that, in the QM/MM calculations, a few point charges near the junction atoms were deleted, which is not the case for $E'_{QM1+ptch}$, in which all point charges of the MM system were included. E_{el12} is the electrostatic interaction energy between the QM (with H junction atoms) and MM systems (excluding interactions within the two systems), calculated at the MM level (i.e., by Coulomb's law), using the same point charges for the MM system as in $E'_{QM1+ptch}$ and the QM fitted point charges for the QM system (i.e., the same point charges used in the MD simulation), with the exception for the junction atoms, for which the original CHELP-BOW point charges were used (in the MD simulations, the point charges on the junction atoms were adapted so that the total charge of the protein was an integer; in the calculation of E_{el12}, the sum of the point charges of the QM system is an integer, whereas that of the MM system is not an integer). Thereby, we obtain an energy that is as similar as possible to that in the QM calculation ($E'_{QM1+ptch}$).

The vertical MM → QM/MM free energies ($\Delta A_{MM \rightarrow QM/MM}(P)$ and $\Delta A_{MM \rightarrow QM/MM}(R)$) are computationally expensive to obtain. Therefore, approaches have been suggested that employ only a single QM calculation for each state (i.e., $\Delta A_{MM \rightarrow QM/MM}(P) - \Delta A_{MM \rightarrow QM/MM}(R)$ is approximated by $E_{QM}(P) - E_{QM}(R)$). This corresponds to the QM/MM-FE approach[14,15] (called QM-FE if the QM system is optimized in a vacuum, rather than by QM/MM[10,11]). We have shown that this was a rather good approximation to QTCP for our test case, the methyl transfer by catechol *O*-methyltransferase, and we have also discussed how the $E_{QM}(P) - E_{QM}(R)$ energy difference is best calculated.[18,19] In this paper, we also test the QM/MM-FE approach. The $E_{QM}(P) - E_{QM}(R)$ energy difference is calculated by optimizing the QM wave functions with a point-charge model of the surroundings. Then, the point charges are removed and the energy is estimated by a single self-consistent field iteration (i.e., without reoptimizing the wave function and using a final grid size of 3 in Turbomole). Thereby, we obtain the energy of the QM system polarized by the surroundings, but excluding the electrostatic interactions between the QM system and the point charges, E_{QM1pol}. However, essentially the same results are obtained by using vacuum QM calculations (i.e., without any point-charge model, E_{QM1}) directly on the QM/MM structures.

In some QTCP calculations, solvent-exposed charged amino acids were neutralized (not in the MD simulations, but only in the final FEPs). Several methods to neutralize the residues were tried, for example, changing them to the corresponding neutral residue in the Amber libraries or scaling down all charges by a factor (dielectric constant) of 80, but the results were very similar. Therefore, we use in the presented results the simplest approach, namely, zeroing all charges in the neutralized residues. The solvent exposure of the residues was determined by counting the number of (non-hydrogen and nonwater) atoms within 15.5 Å of the center of the charged group (the side-chain nitrogen atom of Lys and the amino terminal; C^ζ of Arg; the average position of the two carboxylate oxygen atoms in Asp, Glu,

and the carboxy terminal; and the center of the imidazole ring of doubly protonated His). Groups with less than 575 protein atoms within this radius were considered solvent-exposed.[51]

Poisson−Boltzmann Solvation Energies. For comparison, the solvation energy of the QM system in the surrounding protein and water was estimated by solving the Poisson−Boltzmann equation using the solvinprot module of the software Mead 2.2,[52] following the same procedure used before for both H2ase and NIR.[8,25] Similar procedures have been used for several other proteins.[5−7,9,53−55] The QM system was modeled with the same QM point charges as in the QTCP calculations. The dielectric constants of the QM system, the protein, and the water solvent were set to 1, 4, and 80, respectively. No explicit water molecules were included in the calculations (except the three H_2O ligands of Mg in the H2ase). The positions of the atoms were taken from the QM/MM structures (H2ase) or from the ComQum-X structures (NIR). The Mead calculations were performed with 351^3 grid points and a spacing of 0.25 Å, centring the grid on the QM system. The reported energies are the average of seven calculations in which the grid origin was moved 0.1 Å in the positive and negative directions along each Cartesian axis. The maximum difference among the seven calculations was 2 kJ/mol. Parse radii[56] were used for all atoms. These values have previously been shown to provide reasonable agreement with experimental results.[9,57,58] The MEAD solvation energies (ΔG_{PB}) were added to the polarized QM energy to get a full free-energy estimate:

$$\Delta E_{PBtot} = \Delta G_{PB} + \Delta E_{QM1pol} \quad (5)$$

[Ni,Fe] Hydrogenase. The calculations on H2ase were based on the $P2_12_12_1$ crystal structure of the Ser499Ala mutant from *D. fructosovorans*,[59] which was the crystal structure with the best resolution (1.81 Å) at the start of our investigation. 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.[59]

Two different sizes of the simulated system were tested. One set included the full protein: 818 protein residues and either 5797 (QM/MM) or 14 790 (QTCP) water molecules, giving a total of 29 689 or 56 668 atoms. In the second set of calculations, all residues more than 27 Å from the Ni ion were deleted, and solvation water molecules were added to the protein, forming a sphere with a radius of 33 Å (602 protein residues and 1042 water molecules, giving a total of 12 178 atoms; this is the system used in our previous QM/MM study of this enzyme[25]). In the QTCP calculations, this system was embedded in an octahedral box with 6701 water molecules, yielding a total of 29 149 atoms in the system.

For metal sites outside the active site (the two [4Fe−4S] clusters, the [3Fe−4S] cluster, and a six-coordinated Mg ion), we used Merz−Kollman electrostatic potential (ESP) charges, taken from QM calculations of truncated models of each site. Force constants for the bond, angle, and dihedral terms were estimated from the Hessian matrix, obtained from a QM frequency calculation of the optimized models, using the approach by Seminario.[60,61]

The QM system, depicted in Figure 2, consisted of the Ni and Fe ions and their first-sphere ligands (two CN^- ions, 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 modeled by CH_3S^-, whereas His was modeled by imidazole and Glu-25 by acetic acid. 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).

In some calculations, the QM system was enlarged with some surrounding amino acids, as is illustrated in Figure 2. The seven different QM systems, N, NR, NACG, NACG′, NACGR, NCHACG, and NCHACGR, contained 46, 59, 66, 86, 79, 91, and 104 atoms, respectively.

Hydrogen atoms were added to the protein using the leap module in AMBER.[40] The protonation status of histidine residues was determined by an inspection of the local surroundings and hydrogen-bond structure. This gave protonation of the N^1 atom for residues S5 (an initial S refers to the small subunit, whereas residue numbers without S refer to the large subunit), S92, S160, S243, 367, 481, and 549; protonation of the $N^{\epsilon 2}$ atom for S13, S184, 27, 66, 113, 118, 121, 123, 188, 210, 228, 349, and 419; and protonation of both of these atoms for S61, S192, 79, 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.

The positions of the hydrogen atoms and solvation water molecules were optimized by a simulated-annealing molecular dynamics calculation followed by a MM minimization, as has been described before,[62] with the QM system in the HIP state. Test calculations in which this optimization was performed for the HID state instead changed the final QTCP free energies by up to 8 kJ/mol (Table 1, the last row).

The oxidation states of metals were Fe^{II} [63,64] and Ni^{III}. This gives a neutral QM system when Arg-476 is included and a -1 *e* charge otherwise. The QM system was considered in the closed-shell singlet state.

Nitrite Reductase. The calculations on copper NIR were based on the crystal structure of oxidized NIR at pH 6.0 (Protein Data Bank file 1NIC at 1.9 Å resolution).[65] 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[8] using the ComQum-X quantum refinement approach. It is essentially a QM/MM minimization, in which the structure is restrained to remain close to crystallographic raw data (the structure factors).[8,66]

Hydrogen atoms and water molecules were added to the protein and were optimized in the same way as for H2ase. His-100, -135, and -306 were protonated on the $N^{\delta 1}$ atom and His-60, -95, and -145 on the $N^{\epsilon 2}$ atom, and the other

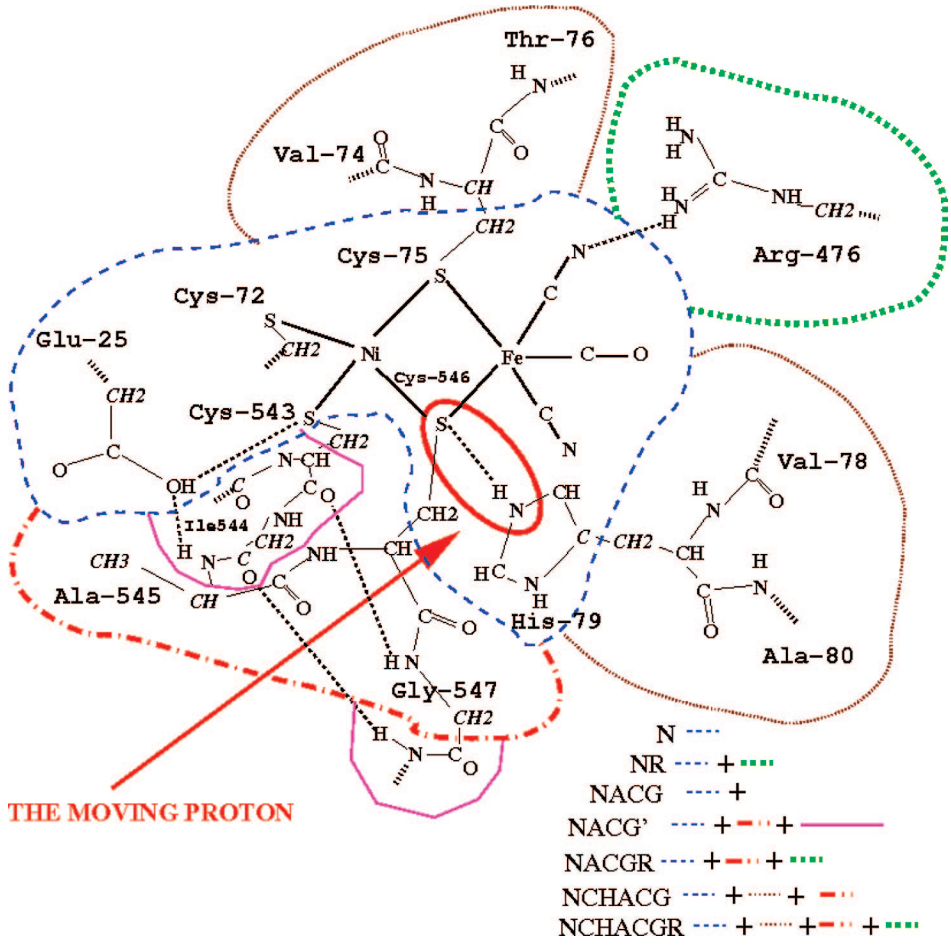

Figure 2. The seven QM systems used for the [Ni,Fe] hydrogenase (N, NR, NACG, NACG′, NACGR, NCHACG, and NCHACGR). In reality, Arg-476 is located above and His-79 below the plane of the paper. Cut bonds that are truncated by hydrogen atoms in the calculations are indicated by dotted lines and hydrogen bonds with dashed lines. The moving proton is indicated by the red circle.

His residues were protonated on both nitrogen atoms. The only Cys residue is a copper ligand and was therefore assumed to be deprotonated.

The QM system consisted of the Cu ion, three imidazole groups, as models of the His ligands, two acetate ions, as models of Asp-98 and Glu-279, and a water molecule. The QM system is shown schematically in Figure 3. The copper ion was studied in the oxidized Cu^{II} doublet state.[67] Therefore, the QM system was neutral. In some calculations, the QM system was enlarged by His-255 and Lys-262, resulting in a charge of the QM system of +2 *e*. The other copper site was treated by pure MM methods, using our previously developed and calibrated parameters (from QM calculations of the oxidized state).[68]

In the QTCP calculations, the whole trimeric protein (1005 amino acids) was considered. This was accomplished by copying the structure of the QM system from the ComQum-X calculation[8] to the other two subunits of the protein. For the QM energies in QTCP, 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).

The QM/MM calculations for NIR were carried out on the quantum-refined structure, to which hydrogen atoms and a solvation layer of ~4000 water molecules were added and optimized by a 90-ps simulated-annealing calculation, followed by a MM minimization. The latter calculations were performed in the Both state to obtain the same environment for all four states. QM/MM_free structures were then obtained by optimizing system 2 for each of the four states separately by a MM minimization until the root-mean-square force was below 10^{-3} au.

Results and Discussion

[Ni,Fe] Hydrogenase: QTCP. Hydrogenases are enzymes that catalyze the seemingly simple reversible reaction[69]

$$H_2 \rightleftarrows 2H^+ + 2e^- \qquad (6)$$

There are several types of these enzymes. The H2ases have an active site with one Ni ion coordinated to four Cys residues and an Fe ion, which coordinates to two of the same Cys ligands and also two CN^- ions and one CO molecule (Figure 2).[70] Many crystal structures of H2ases have been presented,[59,70] but several details of the reaction are still controversial because the reactants and products of eq 6 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.

Table 1. QTCP Results for [Ni,Fe] Hydrogenase (kJ/mol)[a]

Protein	QM system	ΔA_{MM}	hysteresis	$\Delta A_{MM \rightarrow QM/MM}$	ΔA_{QTCP}
truncated	N[b]	−38.0	1.2	−1.2	−39.1
	N	−36.6	1.8	−0.7	−37.3
	N[c]	−33.9	2.8	−0.8	−34.7
	NR	0.9	0.4	−31.5	−30.6
	NACG	−18.1	1.5	−22.4	−40.5
	NACG′	−20.3	0.9	−23.3	−43.5
	NACGR	6.8	2.7	−47.9	−41.1
	NCHACG	−8.0	1.9	−40.7	−48.7
	NCHACGR	20.3	0.9	−64.2	−43.9
	NCHACGR[d]	51.0	0.6	−63.6	−12.6
full	N	−48.4	0.8	−2.3	−50.6
	NR	−14.4	4.0	−37.8	−52.2
	NACG	−46.0	1.5	−20.0	−66.0
	NACGR	−9.8	0.5	−53.1	−62.9
	NCHACG	−38.2	0.9	−39.7	−77.9
	NCHACGR	−6.7	2.5	−75.0	−81.8
	NCHACGR[d]	13.2	2.5	−59.2	−46.0
	NCHACGR[e]	−1.6	3.8	−72.8	−74.3

[a] A negative energy means that the HIP state is most stable. The calculations were performed both on the full and on the truncated protein and with seven different sizes of the QM system (cf. Figure 2). ΔA_{MM} and $\Delta A_{MM \rightarrow QM/MM}$ are the MM and MM → QM/MM parts of the QTCP free-energy difference, respectively (cf. Figure 1). The second column gives the hysteresis in the ΔA_{MM} term. The sum of the ΔA_{MM} and $\Delta\Delta A_{MM \rightarrow QM/MM}$ terms yields the QTCP free energy, ΔA_{QTCP}. [b] Six-step perturbation. [c] Nine Na^+ ions were added to the MM system to neutralize the protein. [d] The QTCP calculations were carried out without the water solvent. [e] The hydrogen atoms were equilibrated for the HID state, rather than for the HIP state.

Many theoretical investigations of the H2ases have also been presented.[25,71–76] In principle, it should be possible to deduce the energetically most favorable protonation sites by theoretical methods, but it is hard to obtain converged energies in such calculations.[25] In this investigation, we use a somewhat simpler test case, namely, the transfer of a single proton between the Ni ligand Cys-546 and His-79 along a hydrogen bond (Figure 2). We try to estimate the relative free energies 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).

We started to study this proton transfer in the truncated protein with the normal (N) QM system (cf. Figure 2) by dividing it into six separate steps, with H−N distances of 1.08 (HIP state), 1.14, 1.26, 1.44, 1.62, 1.80, and 1.98 Å (HID state). The corresponding QTCP free-energy profile is shown in Figure 4 (upper, black curve). It can be seen that the QTCP results indicate that the HIP state is 39 kJ/mol more stable than the HID state and that there is no barrier (activation energy) for the proton transfer. The hysteresis in each step is less than 1 kJ/mol, indicating that the step size is small enough. In fact, the proton transfer can be performed in a single step, giving a similar result (37 kJ/mol; cf. Table 1, first two rows), with a hysteresis of only 2 kJ/mol. Therefore, we have used single-step perturbations in the following investigations, where we study the effect of increasing the quantum system, with both the truncated and full proteins.

The Size of the QM System. An advantage with the MM free energies in the QTCP calculations is that it is straightforward to calculate approximate contributions to the free-energy difference from each QM atom or from each MM residue. Moreover, the interactions can be divided into contributions from the various types of MM energy terms, that is, bonds, angles, dihedrals, electrostatics, and van der Waals interactions.

The results of such an analysis are shown in Table 2. It can be seen that the energies are completely dominated by the electrostatic interactions, with only minimal contributions from the bonded and van der Waals interactions. It can also be noted that there are many residues with quite large contributions, but they to a large extent cancel. The largest contribution comes from Arg-476 (−28 kJ/mol), which is hydrogen-bonded to the CN ligand of the iron ion (cf. Figure 2). Two aspartate residues, Asp-114 and Asp-541, which are hydrogen-bonded to Arg-476, more than cancel the effect of Arg-476 (+15 and +19 kJ/mol, respectively). The MM parts of three QM residues also have significant contributions: Cys-75 (−9 kJ/mol), His-79 (+9 kJ/mol), and Cys-546 (−10 kJ/mol). In addition, the neighboring Gly-547 residue contributes by −12 kJ/mol, and His-481 (which forms a hydrogen bond to the proton acceptor His-79) favors the HIP state by 11 kJ/mol. Residues Arg-70 and Ala-71, which are close to the QM system, favor the HID state by ∼10 kJ/mol each.

These results indicate that we may systematically improve the results by enlarging the QM system by the residues with the largest MM contributions. To this end, the QM system was extended with the rest of residues, Cys-75 and His-79 (including the CO and NH groups of the preceding and succeeding residues, called CH), Ala-545, Cys-546, and Gly-547 (ACG), as well as the guanidine group of Arg-476 (R) in various combinations, as illustrated in Figure 2. In addition, we also tested a larger variant of the NACG system, in which the backbone atoms of Ile-544 and the amide linkage to Val-548 were also included (called NACG′).

The results of the QTCP calculations with the extended QM system calculations are listed in Table 1 for both the truncated and full-protein simulations. It can be seen that including Arg-476 in the QM system (R) has a small and varying effect on the QTCP energy (−4 to +9 kJ/mol, with an average of +1 kJ/mol), which probably mostly reflects the accuracy of the method (note that the hysteresis is 1−4 kJ/mol). On the other hand, there seems to be a more consistent effect of adding both Cys-546 (ACG) and Cys-75 plus His-79 (CH) to the QM system (average effects of −10 kJ/mol for both residues). These two contributions are reasonably additive. Thus, the results suggest that, if residues covalently connected to the QM system have large MM components, it is advisable to include them in the QM system, whereas other interactions (e.g., hydrogen bonds) are sufficiently well described by the MM force field.

The Size of the Simulated System. As mentioned above, we have performed QTCP calculations both with a spherically truncated system and for the full H2ase protein. The latter calculations are ∼5 times more expensive in terms of computer time, and similar truncations are frequently used in theoretical simulations of large proteins. As discussed above, calculations with the two systems give similar trends for the enlarged QM systems, and the HIP state is clearly favored in both calculations. However, the results in Table

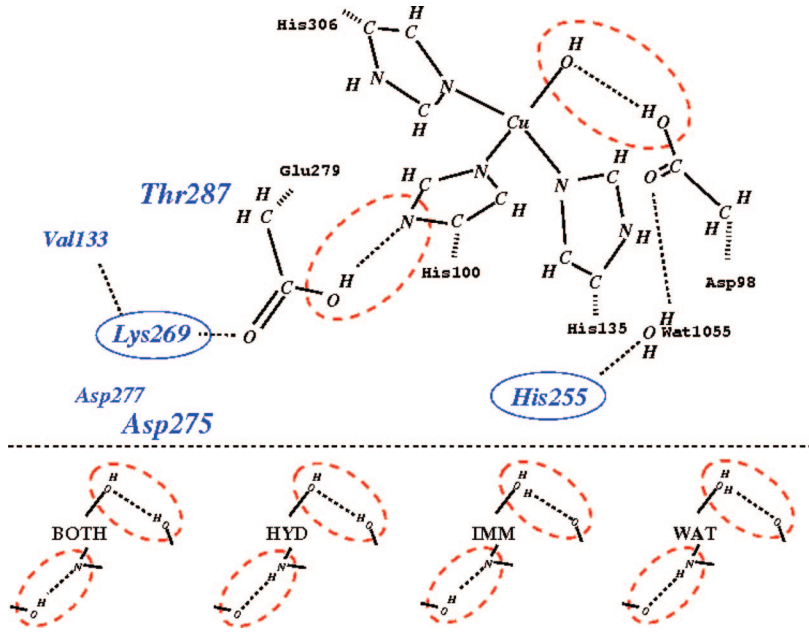

Figure 3. The normal (N) QM system for copper nitrite reductase. The moving protons are indicated as dashed ellipses. The resulting four geometries of the protonation of Glu-279, His-100, and Asp-98 and copper-bound water are sketched and labelled in the lower part of the figure. In some calculations, this QM system is extended with Lys-269 and His-255 (these residues are in ellipses in the figure; NHK). The location of the energetically important residues Val-133, Thr-287, Asp-275, and Asp-277 are also indicated.

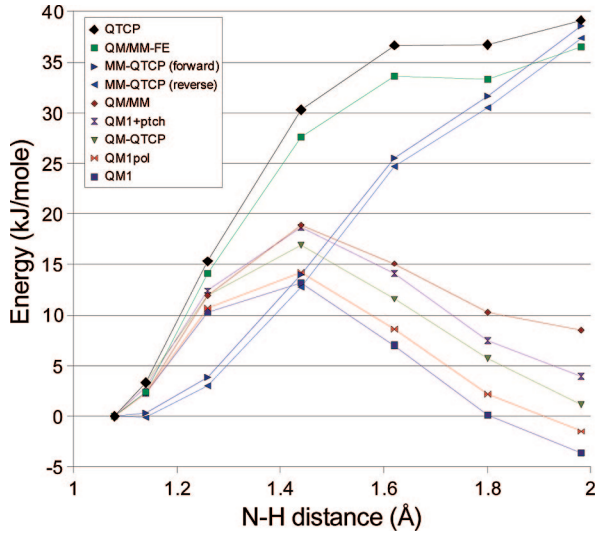

Figure 4. Various energy terms (defined in the Methods section) for the six-step proton transfer between His-79 (the HIP state at 1.08 Å) and Cys-546 (the HID state at 1.98 Å) in [Ni,Fe] hydrogenase. The data are from the truncated protein and the normal (N) QM system.

1 show that there is a shift in the free-energy difference when going from the truncated to the full protein. This shift is 13−38 kJ/mol (average 26 kJ/mol), favoring the HIP state in the full protein. The shift cannot be attributed to any single residue left out in the truncated protein; instead, it is spread over a large number of residues. The individual contributions of residues common to the full and truncated protein are similar (within 2 kJ/mol), as can be seen in Table 3. Thus, the truncation of a protein, even as far as 27 Å from the active site, can have a significant influence on the energetics of the protein.

Table 2. The Largest (Approximate) Contributions to the MM Free-Energy Difference (ΔA_{MM}, kJ/mol) between the HIP and HID States from the Surrounding Protein in the QTCP Simulation of Truncated [Ni,Fe] Hydrogenase with the Normal (N) QM System[a]

residue	bonded terms	van der Waals	electrostatics	total
Arg-70	0	0	10	10
Ala-71	0	−1	12	11
Cys-75	−1	0	−8	−9
His-79	0	0	9	9
Asp-114	0	0	15	15
Arg-476	0	−2	−26	−28
His-481	0	−3	−8	−11
Asp-541	0	0	19	19
Cys-546	0	1	−11	−10
Gly-547	0	0	−12	−12
Mg ion	0	0	9	9
water (sum)	0	0	−11	−11

[a] A negative value indicates that the HIP state is favored. The total QTCP free-energy difference for this QM system is −37 kJ/mol.

To verify that the shift really is caused by the truncated parts of the protein, and not by the water solvation, the calculation with the largest QM system (NCHACGR) was repeated without any water molecules in the system (except the three water ligands of the Mg^{2+} ion). This was achieved by stripping off the water molecules after the MD simulation. This led to a difference of 33 kJ/mol between the truncated and full simulations (cf. Table 1), clearly showing that the shift comes from the truncated part of the protein. Furthermore, it can be seen from the table that the water solvation favors the HIP state by ∼30 kJ/mol in both the full and truncated proteins.

Naturally, the full protein plus the solvent is larger than the truncated protein plus the solvent, and it is conceivable

Table 3. The Largest (Approximate) Contributions to the MM Free-Energy Difference (ΔA_{MM}, kJ/mol) between the HIP and HID State from the Surrounding Protein in the QTCP Simulation of the Full or Truncated [Ni,Fe] Hydrogenase with Various QM Systems[a]

residue	N	N	NR	NACG	NACGR	NACGR	NCHACG	NCHACGR
protein	full	truncated	full	full	full	truncated	full	full
Arg-70	10	10	10	12	11	12	11	11
Ala-71	10	11	12	10	13	14	13	11
Cys-75	−10	−9	−15	4	−4	−3	QM	QM
His-79	11	9	10	12	11	11	QM	QM
Asp-114	16	15	19	15	18	17	16	18
Hip-115	−7	−7	−8	−7	−7	−8	−7	−7
Arg-428	−9	−9	−9	−8	−9	−9	−9	−9
Arg-476	−29	−28	−3	−27	−4	−4	−27	−4
His-481	−10	−11	−11	−12	−11	−13	−9	−10
Asp-541	20	19	23	19	22	21	19	22
Cys-543	7	5	16	5	8	7	12	12
Ala-545	8	7	6	−11	−10	−9	−10	−11
Cys-546	−9	−10	−5	QM	QM	QM	QM	QM
Gly-547	−14	−12	−12	9	4	4	6	7
Mg	8	9	10	8	8	8	7	7
water (sum)	−10	−11	−20	−12	−15	−13	−12	−15

[a] A negative value indicates that the HIP state is energetically favored by the residue. QM indicates that the whole residue was included in the QM system.

that the shift is caused by the extra size in the full system. Therefore, we estimated this effect by calculating the bulk solvation energy from the change in dipole moment (the HID state of the isolated QM system in a vacuum has a larger dipole than the HIP state, 19 and 15 D for NCHACGR and 16 and 14 D in NACGR, respectively), using the Onsager formula.[77] However, this effect is only ∼5 kJ/mol, and it actually increases the shift.

Next, nine Na^+ ions were added to the simulated system to compensate for the different total charge between the truncated and whole protein. The QTCP free energies (Table 1) show that the Na^+ ions favor the HIP state, but their effect is only marginal (∼5 kJ/mol) and cannot explain the energy difference between the full and truncated proteins.

Finally, we studied the MM part of the free energies in more detail. In Figure 5, the cumulative sum of ΔA_{MM} from each residue is plotted as a function of its closest distance to the QM system. It can be seen that the curves from the simulation of the full and truncated proteins differ by 3 kJ/mol already at zero distance (i.e., for the residues that are partly in the QM system). The curves then run roughly in parallel, with differences of less than 6 kJ/mol up to 16.5 Å, at which point the first missing residue in the truncated system is found (these distances are measured from the whole QM system, whereas the protein was truncated on the basis of the distance from the Ni atom alone; moreover, the surrounding protein is flexible). From there on, the difference increases and ends up at 40 kJ/mol (note that neither $\Delta A_{MM \rightarrow QM/MM}$ nor ΔA_{MM} from the water molecules are included in Figure 5).

These differences are caused by at least three effects. First, the charges on the QM atoms differ in the two simulations, because they are estimated from the QM ESP, on the basis of the QM/MM structures, which are slightly different for the full and truncated proteins (both the QM and MM coordinates affect the QM charges, because we used wave functions polarized by the point charges of the protein in the charge calculations). This effect can be estimated by using

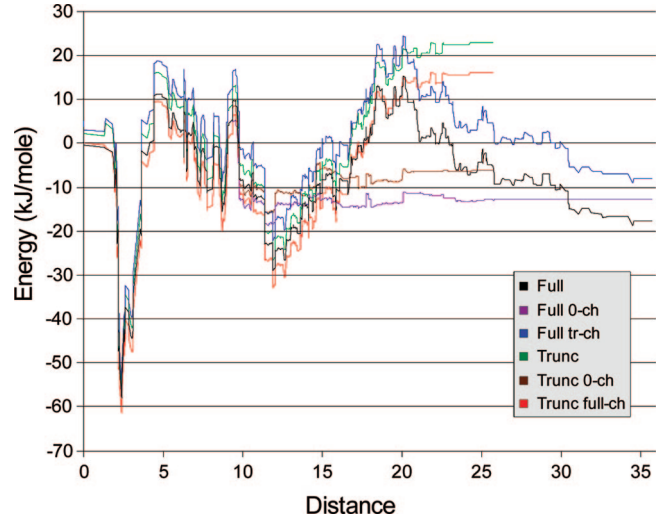

Figure 5. The cumulative residue components of ΔA_{MM} for the NCHACG QM system of [Ni,Fe] hydrogenase as a function of the distance of the residue from the QM system. Six different simulations are shown, three with the full protein (Full) and three with the truncated protein (Trunc). Two used the normal charges, whereas in two, all of the charged residues with a distance greater than 20 Å from the QM system were neutralized (0-ch; note that those curves are identical to the standard-charge curve up to approximately this distance). In the last two simulations, the charges of the QM system were taken from the other simulation (Full trunc-ch and Trunc full-ch).

the QM charges from the truncated protein for the full-protein QTCP calculation, and vice versa (the charges were switched only in the FEP, not in the MD simulations). Those results are also included in Figure 5, and it can be seen that the charges change the free energies by 7−10 kJ/mol, most of which arises from interactions with residues within 5 Å of the active site. This changes the difference between the simulations with the full and truncated protein somewhat, but the difference for the total QTCP free energy is still −27 and −26 kJ/mol, using the full-protein or truncated-protein

Table 4. The Protein Part of the MM Free-Energy Difference (ΔA_{MM}, kJ/mol) between the HIP and HID State in [Ni,Fe] Hydrogenase (NCHACG QM System) As a Function of How Far from the Active Site Solvent-Exposed Charges Are Scaled down by an Effective Dielectric Constant of 80[a]

distance	full	truncated	difference
∞	−17.7	22.8	40.5
25	−4.2	22.9	27.1
20	12.5	19.8	7.3
15	−6.1	−1.4	4.8
10	−7.2	−3.5	3.7
5	−11.6	−8.0	3.6

[a] At 5 Å, all solvent-exposed charges are scaled down (the closest one is at 9.4 Å). The results for simulations with the full and the truncated protein are given, as well as their difference.

charges in both simulations, respectively). Thus, the differences in the QM charges cannot explain the difference between the full and truncated protein.

The second effect is the difference in the coordinates of the two systems, both for the QM and MM atoms (which also indirectly affect the charges, as we saw above). This effect explains the difference between the simulations using the same charges, at distances up to 16.5 Å in Figure 5. It can be seen that the effect is modest, up to 5 kJ/mol.

Consequently, the difference between the full- and truncated-protein simulations actually comes directly from the truncation of the protein, that is, from the long-range electrostatics of deleted residues in the truncated simulations. This is an alarming and problematic observation for theoretical simulations of proteins.

However, the effect comes almost entirely from charged residues: From Figure 5, it can be seen that, if we neutralize all solvent-exposed charged residues, the difference between the full and truncated system is reduced to 6 kJ/mol (10 kJ/mol for the total QTCP free energy). This shows that the difference between the full and truncated protein is almost entirely caused by charged groups at large distances from the active site. Such effects have been much discussed. It is know experimentally that charged solvent-exposed groups have little influence on acid constants, redox potentials, and ligand-binding affinities.[78,79] It has also been observed in theoretical studies of such properties that more accurate and stable results are obtained if interactions with charged solvent-exposed groups are ignored or scaled down with a large effective dielectric constant.[79–82] Therefore, it is likely that this large effect of distant charged groups is an artifact.

The only remaining question is how to treat this problem in practical simulations. Apparently, interactions with solvent-exposed charges should be scaled down, but it remains to be settled how these residues should be selected (e.g., if all solvent-exposed charged residues should be neutralized or only those outside a certain distance from the quantum system). From Table 4, it can be seen that the free-energy difference between the full and truncated protein calculations varies by less than 4 kJ/mol if the distance between the QM system and the down-scaled residues (by $\epsilon = 80$) is varied between 0 and 20 Å. However, the absolute free-energy difference changes drastically. Therefore, we suggest that the only reasonable choice is to scale down all solvent-exposed charged residues, as was done in Figure 5.

Finally, we need also to decide how strongly the charges should be scaled down. Values between ∼20 and infinity have been discussed.[79] However, from Table 4, it is clear that the total effect is only up to 40 kJ/mol, meaning that the various choices differ by less than 2 kJ/mol (from 40/20 to 40/∞ kJ/mol). Therefore, we have decided to simply ignore these interactions, to avoid introducing another, quite arbitrary constant. The results of such QTCP calculations are given in the first column of Table 5 ($\Delta A_{QTCP\_0\text{-}ch}$). The energies are 4−14 kJ/mol less negative than the QTCP energies of the full protein with no charges removed.

Extrapolation. All QM calculations described up to now were performed with the BP86 functional and the DZP/6-31G* basis sets. This basis set is only medium-sized and not fully converged. For accurate results, the QTCP free energies should therefore be extrapolated to calculations employing larger basis sets and perhaps also a more-accurate DFT functional.[18,19] Consequently, we have extrapolated the QTCP results using calculations with the B3LYP functional and the 6-311+G(2d,2p) basis set. In addition, we have performed frequency calculations for the HID and HIP states, optimized in a vacuum to correct the results for the zero-point energies and the entropy and enthalpy of the QM system (the QM system is fixed in the QTCP calculations).[18,19,49]

The results of these calculations are collected in Table 5. It can be seen that the results depend only slightly on the type of the functional and the size of the basis set. The effect of the basis set is 3−8 kJ/mol and that of the functional is 1−6 kJ/mol, both favoring the HIP state (together by 5−14 kJ/mol). On the other hand, the zero-point energy favors the HID state by +5 kJ/mol, whereas the entropy correction is +4 kJ/mol, thus roughly canceling the functional and basis set dependency. If these corrections are combined with our best QTCP estimates ($\Delta A_{QTCP\_0\text{-}ch}$), we arrive at our final estimate that the HIP state is ∼75 kJ/mol more stable than the HID state (with the two largest QM systems).

Comparison with Other Methods. In this section, we will compare the QTCP results with results obtained by other methods, both to get a deeper understanding of the energies and to see if results of similar quality can be obtained with cheaper methods. Let us first go back to the six-step perturbation of the normal QM system in truncated H2ase, which gave a QTCP free-energy difference of 39 kJ/mol in favor of the HIP state. It can be seen from Figure 4 that the original QM/MM energies are quite different, indicating that the HIP state is only 9 kJ/mol more stable than the HID state. The QM/MM curve follows rather closely the QM with point charges (QM+ptch) curve, with a largest difference of 5 kJ/mol. This shows that the difference in the MM correction ($E_{MM123} - E_{MM1}$; cf. eq 1) is less than 5 kJ/mol.

In the figure, it can also be seen that the effect of the point charges ($E_{QM1+ptch} - E_{QM1}$) is less than 8 kJ/mol. Finally, the effect of the polarization of the wave function by the point charges ($E_{QM1pol} - E_{QM1}$) is 2 kJ/mol in all of the intermediates. Interestingly, all three QM energies (E_{QM1}, E_{QM1pol}, and $\Delta A_{MM \rightarrow QM/MM}$) are close to zero, and the E_{QM1} energy actually indicates that the HID state is 4 kJ/mol more

Table 5. Extrapolated Results for the Relative Energy (kJ/mol) of the HID and HIP States in the full [Ni,Fe] Hydrogenase (A Negative Sign Indicates That the HIP State Is More Stable)[a]

QM system	$\Delta A_{QTCP\_0\text{-}ch}$	BP86/B1	BP86/B2	B3LYP/B2	ΔA_{corr}	extrapolated
N	−37.4	3.6	−0.1	−1.1	9.4	−32.7
NR	−47.9	−28.7	−31.6	−33.6		−43.4
NACG	−52.2	−13.2	−20.4	−25.0		−54.5
NACGR	−54.0	−45.2	−50.5	−55.1		−54.5
NCHACG	−71.8	−31.4	−39.4	−45.5		−76.5
NCHACGR	−71.8	−62.2	−68.0	−73.5		−73.7

[a] The table shows vacuum QM energies of the isolated QM system (ΔE_{QM}), calculated with the BP86 and B3LYP methods, and with either the 6-31G* or 6-311+G(2d,2p) basis sets (called B1 and B2). In addition, the zero-point energy difference between the HID and HIP states, as well as the entropy and thermal corrections to ΔA is included, obtained from a harmonic analysis of the vibrational frequencies of the QM system, optimized in a vacuum (ΔA_{corr}). Only the smallest QM system (N) was used in those calculations, because for larger systems, major rearrangements of the second-sphere ligands can be expected. Finally, we have added all corrections to the QTCP free-energy estimate for the full protein with solvent-exposed charges neutralized ($\Delta A_{QTCP\_0\text{-}ch}$), to obtain the extrapolated results in the last column.

Table 6. Relative QM Energies (kJ/mol) of the HIP and HID States (A Negative Energy Indicates That HIP State Is Most Stable) in the Truncated and the Full [Ni,Fe] Hydrogenase with Different Sizes of the QM System[a]

protein	QM system	ΔE_{QM1}	ΔE_{QM1pol}	$\Delta E_{QM1+ptch}$	$\Delta E_{QM1+\epsilon=4}$	ΔE_{PBtot}	$\Delta E_{QM/MM}$	$\Delta E_{QM/MM\_free}$	$\Delta A_{QM/MM\text{-}FE}$	ΔA_{QTCP}
truncated	N	3.6	1.5	−3.8	−54.2	−68.2	−8.5	35.6	−28.4	−37.3
	NR	−28.7	−28.5	−3.4	−70.3	−76.4	−8.7	−14.5	−27.6	−30.6
	NACG	−13.2	−16.3	−17.6	−70.1	−95.4	−4.0	−58.0	−34.4	−40.5
	NACG′	−18.9	−21.0	−16.9	−72.8	−59.1	−19.3	−29.9	−41.3	−43.5
	NACGR	−45.2	−45.7	−18.2	−85.5	−71.9	−1.8	−21.7	−38.8	−41.1
	NCHACG	−31.4	−35.5	−37.7	−78.2	−66.9	−21.3	−41.6	−43.5	−48.7
	NCHACGR	−62.2	−63.7	−34.0	−91.5	−80.4	−34.6	2.4	−43.3	−43.9
full	N	2.3	2.4	−11.8	−56.5	−82.9	−18.2	−28.9	−46.0	−50.6
	NR	−29.9	−30.0	−12.7	−72.0	−85.1	−19.2	−23.5	−44.4	−52.2
	NACG	−15.5	−17.6	−22.7	−72.2	−102.4	−24.5	−18.6	−63.6	−66.0
	NACGR	−46.2	−46.9	−23.1	−85.1	−101.4	−25.6	−37.8	−56.7	−62.9
	NCHACG	−30.7	−33.1	−38.5	−78.1	−101.9	−38.9	−106.3	−71.3	−77.9
	NCHACGR	−63.1	−64.6	−42.3	−93.9	−105.9	−42.6	−54.4	−71.4	−81.8
both	MAD	26.4	25.1	30.3	23.3	32.4	31.5	28.6	5.1	0.0

[a] All energy terms are defined in the Methods section (they are, from the left, the QM energy of the isolated QM system in a vacuum, the QM system polarized by the point-charge model, the QM system with a point-charge model or dissolved in a continuum solvent with $\epsilon = 4$, the QM + Poisson-Boltzmann solvation energy, the QM/MM energy with system 2 fixed at the starting structure or relaxed, as well as the QM/MM-FE and QTCP free energies). The mean absolute difference (MAD) between the various energies and ΔA_{QTCP} is given on the last line.

stable than the HIP state in a vacuum. Thus, the stabilization of the HIP state comes entirely from the surrounding protein (as the ΔA_{MM} energies show), and neither of these energies gives any good approximation to the full QTCP free-energy difference.

However, the QM/MM-FE results are much better. This method uses the same MD simulations and ΔA_{MM} values as QTCP; the only difference between these two methods is in the MM → QM/MM correction ($\Delta A_{MM \rightarrow QM/MM}$). In the more-accurate QTCP approach, this term is estimated by FEP, using ∼200 QM calculations of the same QM system, but with different coordinates of the surrounding point charges. In the QM/MM-FE approach, this term is approximated by a single QM calculation for the reactant and the product (we have used the ΔE_{QM1pol} energy, but the results in Figure 4 show that similar results would have been obtained also with the vacuum ΔE_{QM1} energy). From Figure 4, it can be seen that the E_{QM1pol} energy is within 4 kJ/mol of the $\Delta A_{MM \rightarrow QM/MM}$ term throughout the reaction. Therefore, the QM/MM-FE results are very similar to the QTCP results, as has been seen also before for three other proteins.[19,43,83]

Similar results are obtained in all of the other simulations, as shown in Table 6. The ΔE_{QM1pol} results are always close to the vacuum values (ΔE_{QM1}; maximum deviation 4 kJ/mol). Thus, the polarization of the QM system has a small effect on the relative energies of the HIP and HID states.

All QM-only energies are poor approximations of the QTCP results: The mean absolute deviation (MAD) between ΔE_{QM1} or ΔE_{QM1pol} and QTCP is 26 or 25 kJ/mol, respectively. This is interesting, because most theoretical groups try to estimate reaction energies in proteins using small models of the active site.[1] In fact, ΔE_{QM1} does not converge to QTCP when the system is enlarged (for the largest QM system with 104 atoms, the difference is still ∼20 kJ/mol), and there is no way to decide when convergence is reached. Moreover, we have seen that this is a quite inefficient technique, because hydrogen-bond interactions are well-described by MM.

If the QM system is embedded in a continuum solvent with a dielectric constant of 4 (as is done in many QM-only studies of proteins;[1] $\Delta E_{QM1+\epsilon=4}$), the result is somewhat improved (MAD = 23 kJ/mol, or 20 kJ/mol if compared to ΔA_{QTCP_0ch}). On the other hand, if the protein is modeled by a seemingly more accurate point-charge model ($\Delta E_{QM1+ptch}$), the results get worse (MAD = 30 kJ/mol).

Somewhat unexpectedly, including the solvation of both the surrounding protein and solvent (by solving the Poisson−Boltzmann equation, ΔE_{PBtot} from eq 5), as in

several earlier studies,[5–7,9,53–55] does not improve the results; the MAD from the QTCP results is still 32 kJ/mol. Moreover, this method consistently favors the HIP state.

The original QM/MM results are of similar quality: They underestimate the free-energy difference by 9−42 kJ/mol (average 32 kJ/mol; $\Delta E_{QM/MM}$ in Table 6). This is quite unexpected, because QM/MM typically gives exaggerated electrostatic interactions.[83] The results were not improved if all protein atoms within 15 Å of the QM system and all water molecules were relaxed in the QM/MM optimization ($\Delta E_{QM/MM\_free}$): The energies change much, but they fluctuate more, and the results differ by −28 to +73 kJ/mol from the QTCP free energies, with a similar MAD (29 kJ/mol). This shows that QM/MM energies are sensitive to the conformation of the surrounding protein and solvent, and that a proper sampling of conformations is essential in obtaining stable and reliable energies.

Thus, the only method that gives reasonable results for all models is QM/MM-FE: It closely parallels the QTCP results, with a MAD of only 5 kJ/mol, although the QM/MM-FE results are always more positive than QTCP. Thus, we can conclude that, for H2ase, QM/MM-FE is a reasonable approximation to QTCP that saves much computer time.

Nitrite Reductase. Copper NIR is a bacterial enzyme that catalyzes the one-electron reduction of nitrite to gaseous NO:[65]

$$NO_2^- + e^- + 2H^+ \rightarrow NO + H_2O \quad (7)$$

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. 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.[8] Therefore, we have in a previous article studied the protonation of this residue and the His-100−Glu-279 pair with vacuum QM and QM/MM calculations, as well as quantum refinements of crystal structures.[8] 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, which we called Both (protons on Asp-98 and Glu-279), Hyd (protons on Asp-98 and His-100), Imm (protons on water and Glu-279), and Wat (protons on water and His-100), cf. Figure 3. On the contrary, different treatments of the electrostatics and solvation gave estimates that differed by up to 93 kJ/mol.[8] Thus, this is an ideal test case for methods that aim at estimating accurate energies in proteins.

The results of the QTCP calculations for NIR are listed in the Table 7. Two sets of calculations were performed, because the protonation state of the His-255 residue is not clear. His-255 is proposed to be part of the proton conducting chain from the active site to the solvent, and its protonation state may vary during the catalytic cycle of NIR[65,84] (cf. Figure 3). Therefore, we run calculations with His-255 protonated either on $N^{\epsilon 2}$ (Hie) or on both nitrogen atoms (Hip). With the Hie-255 state, the calculations predict that the Hyd state is 37, 42, and 73 kJ/mol more stable than the Both, Wat, and Imm states, respectively. If His-255 is protonated on both nitrogen atoms (Hip state), these numbers change by up to 18 kJ/mol to 26, 47, and 55 kJ/mol. The results are rather well-converged for most of the reactions—the hysteresis is less than 10 kJ/mol for ΔA_{MM} in all steps, except for the Both → Imm perturbation in the Hie-255 state (14 kJ/mol). The hysteresis in this step could be reduced to 4 kJ/mol by using a time step of 1 fs and a cutoff of 10 Å for the nonbonded interactions in the MD simulation (instead of the usual values of 2 fs and 8 Å), but the actual QTCP result remained the same within 1 kJ/mol. This hysteresis is caused mainly by the water molecules.

Table 7. QTCP Results for the Four Protonation States in Nitrite Reductase (kJ/mol)[a]

reaction	ΔA_{MM}	hysteresis	$\Delta A_{MM \rightarrow QM/MM}$	ΔA_{QTCP}
		N, Hie-255		
Hyd → Both	87.3	6.1	−50.3	37.0
Imm → Wat	−78.8	9.9	47.3	−31.5
Hyd → Wat	2.5	3.5	39.7	42.2
Both → Imm	−7.5	13.8	42.7	35.3
		N, Hip-255		
Hyd → Both	71.8	1.8	−45.7	26.2
Imm → Wat	−64.1	9.7	58.8	−9.7
Hyd → Wat	12.5	2.8	34.2	46.8
Both → Imm	6.5	5.1	20.1	26.6
		NHK, Hip-255		
Hyd → Both	27.1	2.1	7.6	34.7
Imm → Wat	−16.3	5.7	−2.6	−19.0
Hyd → Wat	2.6	1.7	63.2	65.8
Both → Imm	−8.9	5.5	58.2	49.3

[a] ΔA_{MM} and $\Delta A_{MM \rightarrow QM/MM}$ are the MM and MM → QM parts of the QTCP free-energy difference, respectively (cf. Figure 1). The second column gives the hysteresis in the ΔA_{MM} term. The sum of the ΔA_{MM} and $\Delta A_{MM \rightarrow QM/MM}$ terms yields the final QTCP free energy, ΔA_{QTCP}.

Table 8. The Largest (Approximate) Contributions from Various Residues to the MM Free-Energy Difference (ΔA_{MM}) between the Four States in Nitrite Reductase in the Hip-255 State (kJ/mol) and the Normal QM System[a]

residue	Hyd → Both	Imm → Wat	Hyd → Wat	Both → Imm
Val-133	−6	−10	0	0
His-255	−33	26	22	29
Lys-269	68	−72	6	10
Asp-275	−9	13	0	−4
Asp-277	−13	13	0	−3
Thr-287	16	−18	0	0
Wat-1055	0	5	−7	−15

[a] A negative energy indicates that the product is stabilized.

The residues with the largest QTCP ΔA_{MM} free-energy difference contributions are shown in Table 8. In general, the reactions involving proton transfer between Glu-279 and His-100 (Hyd → Both and Imm → Wat) are much more affected by the surrounding protein than the other two reactions. This can be attributed to the fact that there are only two charged residues within 7 Å of the water ligand, Lys-269 and His-255. The results in Table 8 show that these two residues give the largest contributions to all four reaction energies. The distance from the $N^{\epsilon 2}$ of His-255 to the active-site Cu ion is only ∼3.5 Å, and it forms a hydrogen bond to the water molecule (Wat-1055 in the Figure 3) hydrogen-

bonded to Asp-98. By electrostatics, the doubly protonated His-255 favors states in which the moving protons are as far as possible from His-255, that is, Both and Hyd. Lys-269 forms an ion pair with Glu-279, the acceptor of the proton from His-100, involved in the Hyd → Both and Imm → Wat reactions (Figure 3). As expected, Lys-262 has a strong effect on the protonation state of Glu-279: the deprotonated forms of Glu-279 (Hyd and Wat) are favored by ∼70 kJ/mol over the protonated ones by this residue.

Interestingly, two neutral residues also have large contributions: Val-133 and Thr-287, both favoring a protonated His-100 (the Hyd and Wat states). For Val-133, this is caused by the hydrogen bond from its backbone NH to the other carboxylate oxygen atom of Glu-279. For Thr-287, the Coulomb interaction between the CB atom and the carboxylate oxygen atoms of Glu is the largest factor favoring the protonation of His-100. The negatively charged residues Asp-275 and Asp-277 both favor a proton on Glu-279 (the Both and Imm states).

If His-255 (protonated) and Lys-262 are included in the QM system, the QTCP results change by 9−23 kJ/mol (cf. Table 7). The energetic ordering of the states remains the same as that for the smaller QM system: the Hyd state is lowest, followed by Both (35 kJ/mol), Wat (66 kJ/mol), and Imm (84 kJ/mol). The hysteresis with the larger QM system is somewhat smaller than in the previous calculations, 2−6 kJ/mol.

Finally, we have also extrapolated the QTCP free energies with the 6-311+G(2d,2p) basis set and calculated zero-point and thermal corrections for the QM system from frequency calculations. The results in Table 9 show that the larger basis set changes the energies by up to 6 kJ/mol, whereas the zero-point and thermal effects are less than 3 kJ/mol for the relative energies. We also examined the effect of the solvent-exposed charges by repeating the QTCP calculations with those charges neutralized. These results are also included in Table 9, and they show that the charges change the energy difference by up to 20 kJ/mol. Thus, surface charges have a pronounced effect on the energies also in NIR.

Comparison with Other Methods. Finally, we tried to estimate the relative energies of the four protonation states of NIR with other methods. The results in Table 10 are quite similar to those for H2ase: The vacuum (ΔE_{QM1}) and vacuum-polarized (ΔE_{QM1pol}) energies are poor approximations of the QTCP free energies with MADs of 28 and 36 kJ/mol, respectively (and therefore also the original ComQum-X energies, which differ from ΔE_{QM1} by less than 5 kJ/mol[8]). The same applies to the continuum ($\Delta E_{QM1+\epsilon=4}$) or Poisson−Boltzmann solvated energies (ΔE_{PBtot}), with MADs of 23 and 39 kJ/mol, respectively. However, the results for the former improve considerably if they are compared to ΔA_{QTCP_0ch} instead (MAD = 12 kJ/mol). For all except ΔE_{PBtot}, the results are significantly better for the larger QM system with MADs of 15−20 kJ/mol, and for $\Delta E_{QM1+\epsilon=4}$, the results coincide with ΔA_{QTCP_0ch} within 4 kJ/mol. However, this good performance is probably mainly coincidental, considering that the corresponding MAD for H2ase was 19 kJ/mol and no improvement for the larger QM systems was noticed.

Table 9. Extrapolated Results for the Relative Energies (kJ/mol) of the Four Protonation States in Nitrite Reductase[a]

reaction	ΔA_{QTCP_0ch}	B3LYP/B1	B3LYP/B2	ΔA_{corr}	extrapolated
		N, Hie-255			
Hyd → Both	23.7	−30.9	−34.8	−1.0	18.7
Imm → Wat	−25.0	21.8	26.2	0.1	−20.6
Hyd → Wat	27.8	18.7	13.0	1.8	23.9
Both → Imm	26.0	27.7	21.6	2.7	22.6
		N, Hip-255			
Hyd → Both	8.0	−30.9	−34.8	−1.0	3.0
Imm → Wat	−9.4	21.8	26.2	0.1	−5.0
Hyd → Wat	35.6	18.7	13.0	1.8	31.7
Both → Imm	18.9	27.7	21.6	2.7	15.5
		NHK, Hip-255			
Hyd → Both	15.7	5.6	7.2	−1.0	16.3
Imm → Wat	−4.9	−2.6	−2.3	0.1	−4.5
Hyd → Wat	48.2	68.7	67.8	1.8	49.0
Both → Imm	33.6	62.9	62.9	2.7	33.5

[a] The table shows vacuum QM energies of the isolated QM system (ΔE_{QM}), calculated with the B3LYP method and with either the 6-31G* or 6-311+G(2d,2p) basis sets (called B1 and B2). In addition, the zero-point energy difference between the various states, as well as the entropy and thermal corrections to ΔA, is included, obtained from a harmonic analysis of the vibrational frequencies of the QM system, optimized in a vacuum (ΔA_{corr}). Only the smallest QM system (N) was used in those calculations, because for larger systems, major rearrangements of the second-sphere ligands can be expected. Finally, we have added all corrections to the QTCP free-energy estimate for the full protein with solvent-exposed charges neutralized (ΔA_{QTCP_0-ch}), to obtain the extrapolated results in the last column.

On the other hand, the simple point-charge model ($\Delta E_{QM1+ptch}$) gives unexpectedly good results with a MAD of only 12 kJ/mol for all 12 reactions and as low as 7 kJ/mol for the large QM system. Thus, in contrast to H2ase, a simple point-charge model seems to work very well for NIR.

The QM/MM energies ($E_{QM/MM}$) give a MAD of 18 kJ/mol, which is somewhat better than that for $E_{QM/MM}$ in H2ase but inferior to that for the point-charge model ($E_{QM1+ptch}$). Relaxing system 2 for each of the four states (E_{QM/MM_free}) gives much worse results. This is because the environment ends up in different local minima. Thus, it is clear that a proper sampling of the surroundings is required.

As usual, the QM/MM-FE results are similar to the QTCP ones, but the deviations are larger than for H2ase and other systems,[18,19,49,83] up to 20 kJ/mol (MAD = 8 kJ/mol). These larger deviations probably come from the large effect of polarization in NIR: In fact, ΔE_{QM1} and ΔE_{QM1pol} show larger differences than we have seen for any system before, up to 45 kJ/mol, but only for the small QM system. This indicates that the QM system is strongly polarized by Lys-269 (there is little difference between the Hip-255 and Hie-255 calculations). In fact, $\Delta A_{MM \rightarrow QM/MM}$ lies in between ΔE_{QM1} and ΔE_{QM1pol}, but QM/MM-FE based on ΔE_{QM1} is significantly worse (MAD = 14 kJ/mol), with a maximum difference of 33 kJ/mol. This shows the advantage of using ΔE_{QM1pol} in the QM/MM-FE approximation.

Conclusions

In this paper, we have studied two similar proton-transfer reactions between a metal ligand and a second-sphere group

Table 10. Relative Energies (kJ/mol) of the Four Protonation States in Nitrite Reductase, Using the Normal (N) or the Enlarged (NHK) QM System[a]

state	ΔE_{QM1}	ΔE_{QM1pol}	$\Delta E_{QM1+ptch}$	$\Delta E_{QM1+\epsilon=4}$	ΔE_{PBtot}	$\Delta E_{QM/MM}$	$\Delta E_{QM/MM\_free}$	$\Delta A_{QM/MM-FE}$	ΔA_{QTCP}
				N, Hie-255					
Hyd → Both	−30.8	−54.9	39.7	−10.4	17.1	39.1	128.6	32.4	37.0
Imm → Wat	21.8	67.0	−31.4	19.0	−5.4	8.4	0.2	−11.8	−31.5
Hyd → Wat	18.5	41.4	46.6	27.9	35.8	49.4	28.9	43.9	42.2
Both → Imm	27.5	29.2	38.3	19.2	24.1	1.9	−100.0	21.7	35.3
				N, Hip-255					
Hyd → Both	−30.8	−47.4	2.2	−10.4	−42.7	32.4	−115.4	24.5	26.2
Imm → Wat	21.8	66.9	1.8	19.0	75.9	−1.6	−7.0	2.9	−9.7
Hyd → Wat	18.5	45.6	82.0	27.9	90.0	66.2	−83.9	58.1	46.8
Both → Imm	27.5	26.0	78.0	19.2	56.9	35.3	38.5	32.5	26.6
				NHK, Hip-255					
Hyd → Both	5.5	5.6	44.9	19.3	−40.4	39.1	−73.9	32.6	34.7
Imm → Wat	−2.5	1.4	−25.7	−7.1	−31.1	8.4	−28.0	−14.9	−19.0
Hyd → Wat	68.7	76.9	60.9	48.1	40.4	49.4	−70.0	79.5	65.8
Both → Imm	65.7	69.9	41.7	35.8	111.9	1.9	31.8	61.0	49.3
MAD	27.9	35.9	13.5	23.2	38.9	18.4	69.1	8.6	0.0

[a] All energy terms are defined in the Methods section (they are, from the left, the QM energy of the isolated QM system in a vacuum, the QM system polarized by the point-charge model, the QM system with the point-charge model or dissolved in a continuum solvent with $\epsilon = 4$, the QM + Poisson-Boltzmann solvation energy, the QM/MM energy with the same or different MM environments for each state, as well as the QM/MM-FE and QTCP free energies). The mean absolute difference (MAD) between the various energies and the QTCP free energy is given on the last line.

with QM/MM free-energy perturbations. Previous studies have shown that these reactions strongly depend on the surrounding protein and that it is very hard to obtain converged and reliable energies for them.[8,25] This is of course a major problem in the theoretical modeling of enzyme mechanisms, because similar second-sphere interactions are found in most metalloenzymes and the reactivity of the metal site strongly depends on the exact position of this proton.[85–88] Therefore, it is of great interest to develop methods that can accurately predict the position of shared protons.

Our results show that it is mandatory to model the surrounding protein and the solvent in order to obtain reliable results. The proton-transfer energies strongly depend on the electrostatic interactions from the surroundings, and the effect is very long-range. In fact, Figure 5 shows that the cumulative energy does not start to converge until residues up to 10−15 Å from the active site have been included. This shows that QM calculations on isolated models of the active site will never give the correct results[1] (for the right reason), not even if the surroundings are modeled as a continuum solvent with a low dielectric constant (Tables 6 and 10 show that the MAD is 23−28 kJ/mol for ΔE_{QM1} and $\Delta E_{QM1+\epsilon=4}$).

A point-charge model seems to give reasonable results for NIR (Table 10), but not for H2ase. The same applies to the original QM/MM results for H2ase, which are quite similar to the $\Delta E_{QM1+ptch}$ results. This shows that averaging over many protein and solvent geometries also is important. Poisson−Boltzmann solvation models with an effective dielectric of ∼4 have previously been applied to many systems with rather good results,[5–7,9,53,54,56] but the results in Tables 6 and 10 show that such a procedure works unexpectedly poorly both for H2ase and NIR—in fact, it does not give any significant improvement compared to the vacuum calculations. Thus, this popular method seems to be useless for reactions of the type studied in this paper.

The only methods that give reliable results are full QM/MM free-energy perturbation (QTCP) and QM/MM-FE, which both give a full detailed account of the surroundings and sample their dynamic effects at the MM level. For H2ase, QM/MM-FE gives similar good results to those obtained previously for three other proteins,[19,49,83] with differences of 10 kJ/mol or less. However, for NIR, the QM/MM-FE results are slightly worse (errors of up to 20 kJ/mol). This problem can be attributed to an unusually strong polarization of the QM system, which can be identified by an unusually large difference between the ΔE_{QM1} and ΔE_{QM1pol} energies, and it can be solved by enlarging the QM system.

In conclusion, FEP is needed to obtain accurate and reliable energies in proteins, even at a QM/MM level. An advantage with FEP at the MM level is that it allows us to determine approximate contributions to the free energy, both from the various components in the MM energy function (bonded, van der Waals, and electrostatics) and from the various residues in the protein. The latter can be used to identify in an unbiased way important groups in the protein, which may be included in the QM system. Our results indicate that, as a rule of thumb, electrostatic interactions and hydrogen bonds are rather well-described by MM (and therefore need not be included in the QM system), whereas if MM atoms covalently bound to the QM system give large energy contributions, they should be included in the QM system.

A natural question is how accurate the QTCP results are. We have run many calculations with variations in the QM system, the MM system, the simulation protocol, and various parameters. These allow us to obtain an estimate of the final energy. First, the hysteresis in the calculation of ΔA_{MM} gives an estimate of the convergence of this FEP. We obtained hystereses of up to 4 kJ/mol for H2ase and up to 14 kJ/mol for NIR for single-step perturbations. Of course, the hysteresis can be reduced by dividing the reaction into more steps, as was done for H2ase (Figure 4).

Moreover, Tables 1 and 7 show that the size of the QM system strongly affects the results. However, if the QM

system is enlarged according to our rules above, the variations in the final free energies are 5 kJ/mol for H2ase and up to 23 kJ/mol for NIR, that is, similar to the hysteresis. Likewise, reasonable variations in the simulation protocol have a similar influence on the final energies: The QM charges affect the energies by 5 kJ/mol (Figure 5), and the state, for which the MM system was equilibrated, affects the energies by 8 kJ/mol.

However, the largest effect was observed for the surface charges. Table 4 shows that distant charges may change the energy by up to 37 kJ/mol in H2ase. For NIR, the effect is somewhat smaller, up to 20 kJ/mol. We attribute this effect to an incomplete solvation of these groups, in line with previous experimental and theoretical observations and the common use of neutralizing solvent-exposed charges.[78–82] If these charges are neutralized or scaled down by an effective dielectric constant of 40−80, the uncertainty from surface charges is reduced to ∼5 kJ/mol, as was seen in Table 4. Thus, the inherent uncertainty of QTCP seems to be ∼10 kJ/mol, to which should be added the hysteresis of the ΔA_{MM} term, which in our case gives an uncertainty of the final QTCP results of 10 kJ/mol for H2ase and 15 kJ/mol for NIR.

With this estimate, we can finally conclude that the shared proton in H2ase certainly prefers to bind to His-79, rather than to Cys-546 (by ∼75 kJ/mol), although the two states are almost degenerate in vacuum calculations with our normal QM model N (cf. Table 6). In most previous theoretical investigations of H2ase, His-79 was not considered, because the most common QM system includes only the first-sphere metal ligands (i.e., our N model without Glu-25 and His-79).[76,89] However, Stadler et al.[90] considered all possible protonation states of His-79 and concluded that a complex with a deprotonated Cys-546 and His-79 singly protonated on the $N^{\epsilon 2}$ atom (i.e., with one proton less than the states considered in this paper) reproduces experimental EPR parameters best for the oxidized unready A state of the protein. Siegbahn has studied even larger QM models (N + Arg-476 and an Asp formyl group forming hydrogen bonds with Arg-476) and considered the full reaction mechanism for both a neutral and a positively charged His-79.[71,76] He found quite large changes in the energetics, especially for the proton- and electron-transfer steps, with a doubly protonated His-79 giving the best agreement with experiments. However, he assumed that the proton stays on His-79 throughout the reaction and does not present any results for the HID state.

In our previous investigation,[25] we also considered all possible protonation states of His-79 and concluded that the HIP and HID states are close in energy and that protonation of the Cys-546−His-79 system is competitive compared to the protonation of the other three Cys ligands of the catalytic nickel ion, although the energies were sensitive to the surroundings and to the theoretical method. In this paper, we have developed a reliable method that may allow us to continue this investigation of how protons are transported to and from the active site in H2ase.

Likewise, we can conclude that the catalytic copper ion in NIR is in the Hyd state, that is, that the proton shared by His-100 and Glu-279 resides on His and that the proton shared by Asp-98 and the copper-bound water resides on Asp-98. This state is ∼16 kJ/mol more stable than the Both state (with Glu-279 protonated), although the latter state is 31 kJ/mol more stable in a vacuum with the normal size of the QM system (Table 10) and also more stable with several other methods.[8] However, the stability of the Hyd state is confirmed by the absence of any Cu-to-imidazolate charge-transfer transitions in the electronic spectra of the catalytic copper ion.[67] This conclusion is independent of the protonation state of His-255 (although the exact energy estimate will vary).

Again, this result has strong bearings on the study of the reaction mechanism of this enzyme. Most importantly, it shows that the copper ligand is deprotonated in the oxidized resting state, but it also shows that Asp-98 may serve as a reservoir and relay for the protons involved in the reaction (cf. eq 7). All previous theoretical studies of NIR have omitted Glu-279 and Lys-269 in the QM system,[91] although some have included Asp-98 and sometimes also His-255[92] and even Ile-257.[93,94] Our results show that a neutral imidazole group is a reasonable model for His-100, but it is likely that the negatively charged Glu-279 may tune the properties of the copper site, partly neutralizing the charge of the copper ion.

Finally, it seems appropriate to ask why proton-transfer reactions of this type are so sensitive to the surrounding protein, although the geometric difference between the two studied states is minimal (the proton moves 0.9 Å, some atoms in the His-79 ring move up to 0.3 Å, whereas only the Cys-546 S^{γ} atom and the oxygen atom of CO move more than 0.1 in the QM system in H2ase; for NIR, the difference is even smaller—the proton moves by 0.5 Å, whereas only the acceptor oxygen and $H^{\epsilon 1}$ in the His-100 moves more than 0.1 Å in the Hyd → Both transition). The answer is clearly the large change in dipole moment during the proton transfer: We have seen that, for H2ase, the HID state has a 2−4 D higher dipole moment, whereas in NIR, the Hyd state has a 6 D higher dipole moment than the Both state. Thus, it is clear that, if such large changes in the dipole moment are encountered, the results will strongly depend on the surroundings, and free-energy perturbation techniques combined with QM/MM are needed for accurate energies.

Acknowledgment. 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) Siegbahn, P. E. M.; Blomberg, M. R. A. *Chem. Rev.* **2000**, *100*, 421–437.

(2) Mulholland, A. J. In *Theoretical biochemistry - processes and properties of biological systems*; Eriksson, L. A., Ed.; Elsevier Science: Amsterdam, The Netherlands, 2001; Vol. 9, pp 597−653.

(3) Ryde, U. *Curr. Opin. Chem. Biol.* **2003**, *7*, 136–142.

(4) Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173–290.

(5) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochem.* **1993**, *32*, 8045–8056.

(6) Noodleman, L.; Lowell, T.; Han, W.-G.; Li, J.; Himo, F. *Rev. Chem.* **2004**, *104*, 459–508.

(7) Ullmann, G. M.; Noodleman, L.; Case, D. A. J. *Biol. Inorg. Chem.* **2002**, *7*, 632–639.

(8) Källrot, N.; Nilsson, K.; Rasmussen, T.; Ryde, U. *Int. J. Quantum Chem.* **2005**, *102*, 520–541.

(9) Olsen, L.; Rasmussen, T.; Hemmingsen, L.; Ryde, U. *J. Phys. Chem. B* **2004**, *108*, 17639–17648.

(10) Chandrasekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974–2975.

(11) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72–79.

(12) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516–17524.

(13) Olsson, M. H. M.; Hong, G.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.

(14) Zhang, Y.; Liu, H.; Yang, W. J. *Chem. Phys.* **2000**, *112*, 3483–3492.

(15) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2003**, *125*, 12035–12048.

(16) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715–17733.

(17) Li, G.; Zhang, X.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 8643–8653.

(18) Rod, T. H.; Ryde, U. *Phys. Rev. Lett.* **2005**, *94*, 138302.

(19) Rod, T. H.; Ryde, U. *J. Chem. Theory Comput.* **2005**, *1*, 1240–1251.

(20) Fisher, C. L.; Chen, J.-L.; Li, J.; Bashford, D.; Noodleman, L. *J. Phys. Chem.* **1996**, *100*, 13498–13505.

(21) Ryde, U. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 153–164.

(22) Ryde, U.; Olsson, M. H. M. *Int. J. Quantum Chem.* **2001**, *81*, 335–347.

(23) Reuter, N. I.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem.* **2000**, *104*, 1720–1735.

(24) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357–19363.

(25) Söderhjelm, P.; Ryde, U. *J. Mol. Struct. (Theochem)* **2006**, *770*, 199–219.

(26) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098–3100.

(27) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822–8824.

(28) Hehre, W. J.; Radom L.; Schleyer P. v. R.; Pople, J. A. In *Ab initio molecular orbital theory*; Wiley-Interscience: New York, 1986; pp 65−88.

(29) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(30) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

(31) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283–290.

(32) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119–126.

(33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(34) Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345–351.

(35) Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346–354.

(36) Jensen, F. In *Introduction to Computational Chemistry*; John Wiley & Sons: New York, 1993; pp. 298−308.

(37) Klamt, A.; Schüürmann, J. *J. Chem. Soc., Perkin Trans. 2* **1993**, *5*, 799–805.

(38) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187–2193.

(39) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem.* **1998**, *102*, 5074–5085.

(40) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, 2004.

(41) Cornell, W. D.; Cieplak, P. I.; 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.

(42) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(43) Sigfridsson, E.; Ryde, U. *J. Comput. Chem.* **1998**, *19*, 377–395.

(44) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(46) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(47) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8592.

(48) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(49) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452–461.

(50) Chan, H. S. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 543–571.

(51) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704–721.

(52) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473–486.

(53) Li, J.; Nelson, M. R.; Peng, C. Y.; Bashford, D.; Noodleman, L. J. *Phys. Chem. A* **1998**, *102*, 6311–6324.

(54) Rabenstein, B.; Ullman, G. M.; Knapp, E.-W. *Biochemistry* **2000**, *39*, 10487–10496.

(55) Popovic, D. M.; Quenneville, J.; Stuchebrukhov, A. A. *J. Phys. Chem. B* **2005**, *109*, 3616–3626.

(56) Sitkoff, T.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(57) Ullmann, G. M.; Noodleman, L.; Case, D. A. *J. Biol. Inorg. Chem.* **2002**, *7*, 632–639.

(58) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochemistry* **1993**, *32*, 8045–8056.

(59) Volbeda, A.; Montet, Y.; Vernède, X.; Hatchikian, E. C.; Fontecilla-Camps, J. C. *Int. J. Hydrogen Energy* **2002**, *27*, 1449–1461.

(60) Seminario, J. M. *Int. J. Quantum Chem., Quant. Chem. Sympos.* **1996**, *30*, 59–65.

(61) Nilsson, K.; Lecerof, D.; Sigfridsson, E.; Ryde, U. *Acta Crystallogr., Sect. D* **2003**, *59*, 274–289.

(62) Pierloot, K.; De Kerpel, J. O. A.; Ryde, U.; Olsson, M. H. M.; Roos, B. O. *J. Am. Chem. Soc.* **1998**, *120*, 13156–13166.

(63) Frey, M.; Fontecilla-Camps, J. C. A.; Volbeda, A. In *Handbook of metalloproteins*; Messerschmidt, A., Huber, R., Poulos, T., Wieghardt, K., Eds.; Wiley: New York, 2001; pp 880−896.

(64) Huyett, J. E.; Carepo, M.; Pamplona, A.; Franco, R.; Moura, I.; Moura, J. J. G.; Hoffman, B. M. *J. Am. Chem. Soc.* **1997**, *119*, 9291–9292.

(65) Adman, E. T.; Godden, J. W.; Turley, S. *J. Biol. Chem.* **1995**, *270*, 27458–27474.

(66) Ryde, U.; Olsen, L.; Nilsson, K. *J. Comput. Chem.* **2002**, *23*, 1058–1070.

(67) Adman, E. T.; Murphy, M. E. P. In *Handbook of Metalloproteins*; Messerschmidt, A., Huber, R., Poulos, T., Wieghart, K., Eds.; J. Wiley & Sons: Chichester, U.K., 2001; pp 1381−1390.

(68) De Kerpel, J. O. A.; Ryde, U. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 157–174.

(69) Vignais, P. M.; Billoud, B.; Meyer, J. *FEMS Microbiol. Rev.* **2001**, *25*, 455–501.

(70) Volbeda, A.; Fontecilla-Camps, J. C. *Coord. Chem. Rev.* **2005**, *249*, 1609–1619.

(71) Siegbahn, P. E. M. *Adv. Inorg. Chem.* **2004**, *56*, 101–125.

(72) Stein, M.; Lubitz, W. *Curr. Opin. Chem. Biol.* **2002**, *6*, 243–249.

(73) Fan, H.-J.; Hall, M. B. *J. Biol. Inorg. Chem.* **2001**, *6*, 467–473.

(74) Bruschi, M.; Zampella, G.; Fantucci, P.; De Gioia, L. *Coord. Chem. Rev.* **2005**, *249*, 1620–1640.

(75) Amara, P.; Volbeda, A.; Fontecilla-Camps, J. C.; Field, M. J. *J. Am. Chem. Soc.* **1999**, *121*, 4468–4477.

(76) Siegbahn, P. E. M.; Tye, J. W.; Hall, M. B. *Chem. Rev.* **2007**, *107*, 4414–4435.

(77) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.

(78) André, I.; Kesvatera, T.; Jönsson, B.; Åkerfeldt, K. S.; Linse, S. *Biophys. J.* **2004**, *87*, 1929–1938.

(79) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400–417.

(80) Mehler, E. L.; Eichele, G. *Biochemistry* **1984**, *23*, 3887–3891.

(81) Penford, R.; Warwicker, J.; Jönsson, B. *J. Phys. Chem. B* **1998**, *108*, 8599–8610.

(82) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786–3791.

(83) Heimdal, J.; Rydberg, P.; Ryde, U. *J. Phys. Chem. B* **2008**, *112*, 2501–2510.

(84) Boulanger, M. J.; Kukimoto, M.; Nishiyama, M.; Horinouchi, S.; Murphy, M. E. P. *J. Biol. Chem.* **2000**, *275*, 23957–23964.

(85) Poulos, T. L. *J. Biol. Inorg. Chem.* **1996**, *1*, 356–359.

(86) De Santis, L.; Carloni, P. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 611–618.

(87) Jensen, K. P.; Ryde, U. *J. Mol. Struct. (Theochem)* **2002**, *585*, 239–255.

(88) Jensen, K. P.; Ryde, U. *Mol. Phys.* **2003**, *101*, 2003–2018.

(89) Pardo, A.; De Lacey, A. L.; Fernández, V. M.; Fan, H.-J.; Fan, Y.; Hall, M. B. *J. Biol. Inorg. Chem.* **2006**, *11*, 286–306.

(90) Stadler, C.; De Lacey, A. L.; Montet, Y.; Volbeda, A.; Fontecilla-Camps, J. C.; Conesa, J. C.; Fernández, V. M. *Inorg. Chem.* **2002**, *41*, 4424–4434.

(91) Silaghi-Dumitrescu, R. *J. Inorg. Biochem.* **2006**, *100*, 396–402.

(92) De Marothy, S. A.; Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Comput. Chem.* **2007**, *28*, 528–539.

(93) Sundarajan, M.; Surendran, R.; Hillier, I. H. *Chem. Phys. Lett.* **2006**, *418*, 96–99.

(94) Periyasamy, G.; Sundarajan, M.; Hillier, I. H.; Burton, N. A.; McDouall, J. J. W. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2498–2506.

CT700347H