Fast Generation of Broken-Symmetry States in a Large System Including Multiple Iron-Sulfur Assemblies: Investigation of QM/MM Energies, Clusters Charges, and Spin Populations

Greco, Claudio; Fantucci, Piercarlo; Ryde, Ulf; de Gioia, Luca

Published in:
International Journal of Quantum Chemistry

DOI:
10.1002/qua.22849

2011

Fast Generation of Broken-Symmetry States in a Large System including Multiple Iron–Sulfur Assemblies: Investigation of QM/MM Energies, Clusters Charges and Spin Populations

Claudio Greco*[a], Piercarlo Fantucci[a], Ulf Ryde[b], Luca De Gioia[a]

[a]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126, Milan (Italy); [b]Department of Theoretical Chemistry, Lund University, P.O. Box 124, Lund S-221 00 (Sweden).

***Corresponding Author:** claudio.greco@unimib.it, Tel +390264483473

Keywords: metalloproteins, broken-symmetry approach, antiferromagnetic coupling, QM/MM, iron–sulfur clusters

Abstract

A DFT study is presented, regarding the energetics and the Mulliken population analyses of a QM/MM system including multiple iron–sulfur clusters in the QM region. The [FeFe]-hydrogenase from *Desulfovibrio desulfuricans* was studied, and both the active site (an Fe_6S_6 assembly generally referred to as the H-cluster) and an ancillary Fe_4S_4 site were treated at the BP86-RI/TZVP level. The antiferromagnetic coupling that characterize both sites was modeled using the broken-symmetry (BS) approach. For such a QM system, 36 different BS couplings can be defined, depending on the localization of spin excess on the various spin centers. All the BS states were obtained by means of an effective and simple method for spin localization that is here described, and compared with more sophisticated approaches already available in literature. The variation of the QM/MM energy among the various geometry-optimized protein models was found to be less than 25 kJ mol^{-1}. This energy variation almost doubles if no geometry optimization is performed. A detailed analysis of the additive nature of these variations in QM/MM energy is reported. The Mulliken charges show very small variations among the 36 BS states, whereas the Mulliken spin populations were found to be somewhat more variable. The relevance of such variations is discussed in light of the available Mössbauer and EPR spectroscopic data for the enzyme. Finally, the influence of the basis set on the spin populations, charges and structural parameters of the models was investigated, by means of QM/MM computations on the same system at the BP86-RI/SVP level.

Introduction

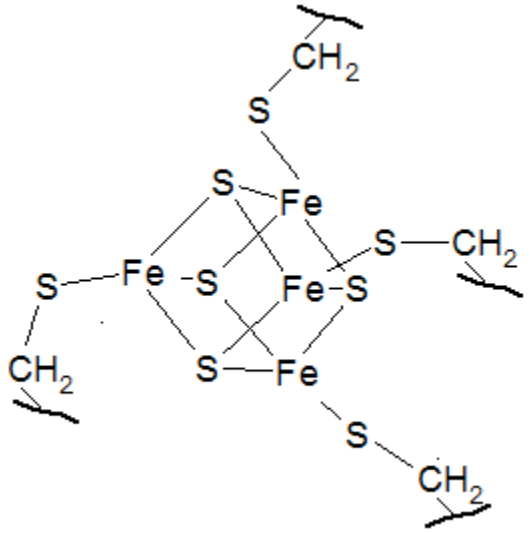
Antiferromagnetic pairing of spin centers is a common feature in many inorganic systems, both in the context of material science, and in organometallic systems, including bioinorganic species like metalloproteins featuring iron–sulfur clusters. In fact, iron–sulfur assemblies have been selected along evolution, in order to function as mediators of electron transfer in redox metabolism. Moreover, they can be directly involved in enzymatic catalysis.[1,2] Among the various forms of Fe–S complexes found in metalloproteins, the tetranuclear Fe_4S_4 assemblies represent one of the most common cases (see Scheme 1). The Fe_4S_4 assemblies usually present high-spin iron ions antiferromagnetically coupled to give rise to an overall low-spin state. Biochemical investigations have shown that they can attain at least four physiologically relevant redox states, viz. the partially oxidized $2Fe^{(2+)}2Fe^{(3+)}$ and $Fe^{(2+)}3Fe^{(3+)}$ states, the highly reduced $3Fe^{(2+)}Fe^{(3+)}$ state, and the rare, completely reduced $4Fe^{(2+)}$ state.[3,4]

Notably, the wavefunction in antiferromagnetically coupled Fe_4S_4 clusters has a multideterminantal nature. In the context of the single-determinant density functional theory (DFT) approach, such coupling can be modeled by means of the broken-symmetry (BS) scheme, in which either alpha or beta spin excess is localized on the different spin centers composing the system. The resulting unrestricted wavefunction corresponds to a linear combination of pure spin states, as shown by Noodleman.[5,6]

The BS approach has been shown to successfully describe key properties in antiferromagnetically coupled molecular systems.[7-9] In particular, Torres et al. have shown that the BS approach is useful for the computation of reliable reduction potential values in Fe_4S_4 clusters.[10] Such results proved seminal for subsequent studies by Bruschi and coworkers,[11,12] who have employed the BS approach to investigate the occurrence of protonation-coupled intramolecular electron transfers in Fe–S complexes.[12] By means of a detailed analysis of Mulliken spin populations and atomic charges, the latter authors were able to show that the active site of [FeFe]-hydrogenases – an Fe_6S_6 cluster composed by two covalently linked Fe_2S_2 and Fe_4S_4 subclusters – can undergo charge-transfer events of functional importance as a result of modifications in the second coordination sphere of the iron atoms.

[FeFe]-hydrogenases, as well as many other metalloenzymes, contain more than one Fe–S cluster of functional relevance, each possibly showing antiferromagnetic coupling and peculiar redox properties. In this regard, the increase of computational power available for quantum chemical calculations allows for extended investigations of the electronic structure of metalloproteins, as it is now feasible to include several Fe–S assemblies in a single QM calculation. Such an approach, which can in principle lead to a detailed quantum-chemical picture of the entire electron-transfer chain within a protein, presents several challenging points. First of all, the inclusion of one or more Fe_4S_4 clusters in a protein matrix leads to the fact that the various possible BS spin-localizations pattern are non-equivalent, as they can undergo different polarization effects by the environment leading to different energies, spin populations and atomic charges. Moreover, the enlargement of the QM region to include several clusters obviously leads to the necessity of fine-tuning the

level of theory used; in particular, the dimension of the basis used for the calculations has to be chosen with care, in order to be able to obtain reliable results at the lowest possible computational cost.



Scheme 1.

In view of these considerations, we have carried out a hybrid quantum mechanical/molecular mechanical (QM/MM) study of an enzyme by including several Fe–S clusters in the same BS representation, using a large basis of TZVP quality at the DFT level. We have considered the [FeFe]-hydrogenase from *Desulfovibrio desulfuricans* (DdH), including two Fe_4S_4 and one Fe_2S_2 assemblies in the QM region. We have also developed a simple and rapid spin-localization technique that allows us to study all possible BS solutions for such spin coupled system. Fluctuation of the energies, Mulliken spin population and charges for the various BS solutions will be discussed, along with the effects of reducing the basis to SVP quality.

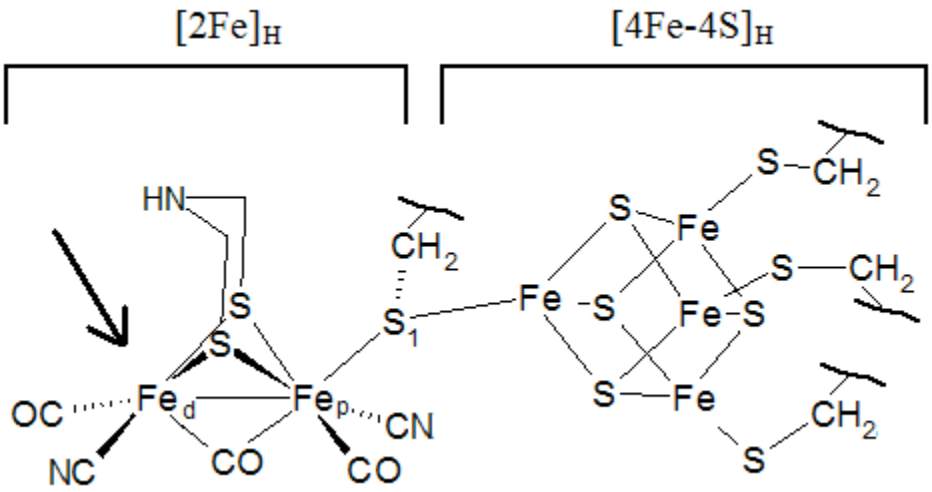
Methods

THE MODEL SYSTEM

As mentioned in the introduction section, the [FeFe]-hydrogenase from *D. desulfuricans* (DdH) has been used as the model system for the present investigation. This enzyme catalyzes the $2H^+ + 2e^- \rightarrow H_2$ reversible reaction at high efficiency. All the calculations are based on the 1.6-Å resolution structure of the enzyme

(PDB code: 1HFE), which is a heterodimer composed of a small and a large subunits.[13] The latter harbors the active site of the enzyme: the H-cluster, which is a hexanuclear complex composed of a classical ferredoxin-like Fe_4S_4 subcluster – referred to as the $[4Fe\text{-}4S]_H$ subcluster in Scheme 2 – and an Fe_2S_2, $[2Fe]_H$ portion, including carbonyl and cyanide ligands. The two Fe–S assemblies are bridged by a cysteine sulfur atom (S_1 in Scheme 2). Notably, the $[2Fe]_H$ cluster features a vacant coordination site on one of the iron atoms, ready for proton binding (Fe_d in Scheme 2). Moreover, it also includes an unusual dithiolate ligand that, according to the most recent literature,[14] is likely to correspond to a di(thiomethyl)amine residue (DTMA). Finally, the large subunit of the protein contains two additional Fe_4S_4 clusters, the F and the F' clusters, which are thought to mediate electron transfer from the bulk to the active site (see Figure 1).

Spectroscopic[15-17] and theoretical[18-20] investigations have shown that in the reduced form of [FeFe]-hydrogenases, the H-cluster is diamagnetic, with the $[2Fe]_H$ subcluster attaining the Fe(I)Fe(I) redox state, and the $[4Fe\text{-}4S]_H$ assembly attaining the $2Fe^{(2+)}2Fe^{(3+)}$ state. On the other hand, the F and F'-clusters were shown to be in the $3Fe^{(2+)}Fe^{(3+)}$ redox state.[15,16] This state of the enzyme will be considered in all the QM/MM calculations here presented.



Scheme 2. Structure of the H-cluster; the arrow indicates the vacant coordination site, available for proton binding in the $[2Fe]_H$ subcluster. The iron atoms of the latter are termed either distal (Fe_d) or proximal (Fe_p), depending on their position with respect to the tetranuclear subsite.

QM/MM CALCULATIONS

In the QM/MM calculations, the protein is divided into two subsystems: system 1 is treated at QM level, and it is allowed to relax. It contains the H-cluster atoms and the F-cluster. The remaining portion of the protein, together with water molecules surrounding it are included in system 2, is treated at MM level. The total energy is calculated as:

$$E_{QM/MM} = E_{QM1} + E_{MM12} - E_{MM1} \quad (1)$$

In eq. (1), E_{QM1} is the QM energy of the quantum system, truncated by hydrogen atoms (*vide infra*), including all the electrostatic interactions with the MM system 2. E_{MM1} is the classical, molecular mechanical energy of system 1; in E_{MM1}, system 1 is still truncated by hydrogen atoms, while the electrostatic contributions are not taken into account. Finally, E_{MM12} is the MM energy of system 1 + system 2. In both E_{MM12} and E_{MM1}, the charges of the QM system are zeroed, to avoid double-counting of the electrostatic interactions between systems 1 and 2.

However, before starting any QM/MM optimization, it is necessary to add the hydrogen atoms that are not detected by means of X-ray investigation of protein crystals. After having added the hydrogen atoms to the crystal, the protein was solvated in a sphere of water molecules with a radius of 48 Å, with the Amber tleap routine. In order to optimize the positions of hydrogen atoms and solvent water molecules, a 90-ps simulated-annealing molecular dynamics calculation was carried out, followed by 10000 steps of conjugate-gradient energy minimization. All the non-hydrogen atoms in the protein were kept at the crystallographic position, in keeping with the our previous QM/MM studies of [FeFe]-hydrogenases.[21,22] QM/MM optimizations with a relaxed MM system were described in ref. 21 and gave negligible differences in computed Mulliken atomic charges and spin populations for the $[2Fe]_H$, $[4Fe\text{-}4S]_H$ and F-cluster, compared to the corresponding calculations with the MM system fixed. The protonation state of the histidine residues was assigned as previously reported.[12,21,22] All lysine and arginine residues were considered in their positively charged state, while aspartate and glutamate side chains were included in the anionic form. Finally, the iron-bound cysteine residues (i.e. amino acids 36, 38, 41, 45, 66, 69, 72, 76, 179, 234, 378 and 382) were assumed to be deprotonated. All the metal-bound ligands found in the PDB file were included in the QM/MM model, except a water molecule bridging Fe_d and Fe_p (see Scheme 2) that was replaced by a carbonyl group.

All the QM/MM optimizations were carried out with the COMQUM program suite,[23,24] which uses TURBOMOLE[25] for the QM part and AMBER 8 [D. A. Case, et al., *Amber 8*. 2004, University of California: San Francisco, CA.] (with the Amber 1999 force field)[26] for the MM part. For the F'-cluster, which is included in the MM region, the reduced $[Fe_4S_4(SCH_3)_3]^{3-}$ redox state was considered.[15,16] The QM

calculations were carried out with density functional theory (DFT), by using the BP86[27,28] functional and, if not otherwise stated, an all-electron valence triple-ζ basis set with polarization functions on all atoms (TZVP)[29] in conjunction with the resolution-of-the-identity (RI) technique.[30,31] The antiferromagnetic coupling that characterizes the Fe_4S_4 assemblies included in the QM region of the QM/MM model was treated by means of the broken-symmetry (BS) approach.[5,6] Details on the BS scheme used and on the approach to obtain the BS wavefunctions are described in the Results section.

In practice, apart from system 1 – which is represented by a wavefunction during the QM/MM geometry optimizations – each atom is represented by a partial point charge. All such MM charges are included in the Hamiltonian of the QM calculations, and thus the quantum chemical system is polarized by the charges of system 2 in a self-consistent way. When the quantum and classical regions are connected by a chemical bond, the hydrogen link-atom approach is applied,[32] i.e. the QM system is truncated with hydrogen atoms, the positions of which are linearly related to the corresponding carbon atoms in the protein.

As for the composition of the QM region in our systems, in model **1,** the QM system (region 1) includes the iron and sulfide ions of the Fe_6S_6 H-cluster and of the Fe_4S_4 F-cluster, the DTMA ligand bridging Fe_d and Fe_p (see Scheme 2), three CO groups, two CN^- ligands, and eight CH_3S^- groups that represent the Cys residues connecting the H-cluster and the F-cluster to the rest of the enzyme large subunit (Cys179, Cys234, Cys378, Cys382, and Cys 45, Cys66, Cys69, Cys76, respectively); the total number of atoms in the QM system is 78 in this case.

In order to evaluate whether the energy variations among different BS states in model **1** are additive, we have also considered two additional models, characterized by the presence of only the H-cluster or only the F-cluster in the QM region. Thus, the QM system of model **2** includes only the H-cluster as defined in Scheme 2, which means that Cys 179, Cys 234, Cys 378 and Cys 382 side chains were included in the QM region as well, giving 50 atoms in the QM system (see Figures 1 and 3). In model **3**, the QM system includes only the sulfide and iron ions of the F-cluster, together with the side chains of the coordinating Cys residues (Cys 45, Cys66, Cys69, Cys76).

Notice that, in all models, the F'-cluster was treated at MM level (i.e. it was included in system 2). Moreover, in model **2**, the F-cluster was included in system 2 as well, whereas in model **3**, the H-cluster was included in the MM region together with the F'-cluster. The inclusion of one or more of the three Fe–S clusters in the MM region implies that their atoms were represented by point charges, that were computed using the Merz–Kollman electrostatic potential (ESP) approach.[33] These charges were obtained from QM calculations at the BP86-RI/TZVP level on truncated models of the H-cluster, the F-cluster or the F'-cluster.

Results

As mentioned in the introduction section, the broken-symmetry (BS) treatment of the QM system in our QM/MM models accounts for the non-equivalency of the iron atoms in the Fe_4S_4 assemblies. Brunold et al.[34] noted that already in the case of the isolated H-cluster, six different possible BS spin topologies can be defined. In this study, the inclusion of a further Fe_4S_4 cluster in the QM system leads to the situation that each of the six spin configurations of the H-cluster can be coupled with six possible spin configurations of the F-cluster, thus leading to 36 different BS wavefunctions.

Previous studies on the isolated H-cluster[11,35] and on QM/MM models in which the QM portion was limited to the H-cluster and a neighboring cysteine sidechain[21] showed that the energy difference among different BS spin states is small, never exceeding 10 kJ mol^{-1}. In view of this, Stiebritz and Reiher[35] decided not to use any explicit spin localization procedures for the selection of a particular BS spin topology. Instead, they always accepted the spin-density distribution calculated by the program they used for QM calculations (TURBOMOLE[25]).

However, the inclusion of an additional Fe_4S_4 cluster in the QM region of our QM/MM model rises the question of how large the QM/MM energy variation is among the larger number of BS solutions available, and of the effects of changing the BS solutions on the computed Mulliken charges and spin populations. Such points will have major relevance for future DFT investigation aiming at the characterization of the electronic structure of entire electron-transfer chains within metalloproteins. Moreover, technical aspects regarding the fast generation of the various BS solutions also acquire more and more relevance, as the number of antiferromagnetically coupled Fe_4S_4 cluster included in the QM system increases. In fact, the inclusion of additional Fe_4S_4 clusters in a quantum-mechanical region brings an exponential increase in the number of BS wavefunctions that can be defined for a given model.

In the case of QM or QM/MM models of the H-cluster, it is commonly found that open-shell SCF calculations at low spin multiplicity directly give one of the possible BS solutions. In other words, there is no need for dedicated spin-localization approaches to obtain at least one of the six BS wavefunction. The same was observed here also for the extended H-cluster + F-cluster QM region. The BS spin pattern thus obtained, which is schematically represented in Figure 2 (model **1**), was then used for the generation of the other 35 possible BS solutions, by means of the fast approach that will be described in the following.[I]

[I] Our results indicate that in none of the 36 BS states optimized at TZVP level (*vide infra*, Table 1), the Fe_p and Fe_d centers have spin populations with opposite sign and significant absolute value (i.e. equal to or larger than 0.1 a.u.). By targeted SCF computations, it was possible to obtain a guess BS wavefunction giving place to significant spin populations on Fe_p and Fe_d, with opposite signs (–1.18 and 1.05 a.u., respectively). However, when this wavefunction was used as a guess for a full SCF run using model 1 (using the QM/MM-minimized geometry of the BS state **[αααββ]-{αββα}** in Table 1) the self-consistent calculation converged to the usual solution showing small and positive spin populations on Fe_p and Fe_d. Therefore, we conclude that the $[2Fe]_H$ cluster is not antiferromagnetically coupled, in accordance with experiments **xxx ref**, and does not couple with the Fe_4S_4 assemblies.

The Mulliken spin population analysis allowed for the identification of the iron atoms characterized by alpha or beta spin excess in **1**. Typically, the spin populations of the antiferromagnetically-coupled iron ions in $2Fe^{(2+)}2Fe^{(3+)}$ and $3Fe^{(2+)}Fe^{(3+)}$ Fe_4S_4 clusters are characterized by absolute values larger or slightly smaller than 3 a.u.; values with opposite signs are observed for atoms having either alpha or beta spin excess. In order to generate the spin localization pattern of the other BS states, we simply exchange the coordinates of pairs of iron atoms that have spin populations with opposite signs. The resulting atomic coordinates were then used together with the unmodified BS wavefunction obtained for the parent model as the guess for the new BS SCF calculation. A Python program for performing such targeted atomic coordinates exchange for COMQUM and TURBOMOLE calculations is available on the website http://linux.btbs.unimib.it (C. G. webpage). As far as QM/MM calculations are concerned, the use of this script is possible only in the cases when the spin centers include metal atoms of the same element, such that the exchange of coordinates does not lead to any change in the chemical structure of the QM system. However, when the model is purely QM, the use of our approach can be easily extended to cases where the spin centers include metal atoms of different nature, as discussed in more details in the Conclusions.

This approach allowed us to easily obtain all the 36 BS states for the reduced form of the enzyme. For each of these BS states, QM/MM energy, together with Mulliken atomic charges and spin populations of the clusters are reported in Table 1. In this Table, the spin state is given as two quartets of "α" or "β", indicating the direction of the surplus spin on the iron ions, according to the numbering in Figure 2 (iron atoms of the $[4Fe\text{-}4S]_H$ subcluster in square brackets and those of the F-cluster in curly brackets, e.g. [ααββ]-{αββα} for the state in Figure 2).

First, let us discuss the Mulliken charges and spin populations for model **1** in the [ααββ]-{αββα} state: The low spin populations on the $[2Fe]_H$ and $[4Fe{-}4S]_H$ subclusters (0.32 and −0.10, respectively) fit well with the observed diamagnetic nature of the H-cluster in H_{red} (in this partitioning we assign the bridging Cys ligand entirely to the $[4Fe{-}4S]_H$ subcluster).[15,16] On the other hand, the F-cluster carries an unpaired electron (spin population 0.80), as expected for a reduced ferredoxin-like Fe_4S_4 cluster. Consistently, the Mulliken charges for the atoms in the F-cluster sum up to –2.81 *e*, which is close to –3, the value expected for a $Fe_4S_4(SCH3)_4$ cluster in the formal $3Fe^{(2+)}Fe^{(3+)}$ redox state. For the $[2Fe]_H$ and $[4Fe{-}4S]_H$ subclusters, the sum of the Mulliken charges are –2.17 and –2.02 *e*, respectively (see Table 1), which are close to what is expected if the binuclear subsite and the tetranuclear subcluster attain the $Fe_d^{+}Fe_p^{+}$ and the $2Fe^{(2+)}2Fe^{(3+)}$ redox states, respectively.

Let us now analyze in details the QM/MM energies computed for the 36 different BS couplings; in Table 1, two sets of QM/MM energy differences are listed: (i) ${}^{SP}\Delta E_{QM/MM}$, which are the differences between the QM/MM energy of the [ααββ]-{αββα} BS state at its minimum geometry and single point QM/MM energies computed for all the other 35 possible BS wavefunctions at the same geometry (the corresponding QM energies – i.e. the ΔE_{QM1} values, see eq. (1) – differ by less than 4 kJ mol^{-1} from the QM/MM energies);

(ii) $^{OPT}\Delta E_{QM/MM}$, which is the QM/MM energy differences between the [αααββ]-{αββββα} state and the other 35 BS states, each geometry-optimized as described in Methods. In addition, we also report in Table 1 the average $^{OPT}\Delta E_{QM/MM}$ values for the six subgroups of models characterized by the same BS coupling of the H-cluster (last column of the Table), and − for each subgroup − the QM/MM energy differences between the most stable model and the other five members, all geometry-optimized (third column of Table 1, energy values in parenthesis).

It turns out that the pattern of $^{OPT}\Delta E_{QM/MM}$ values is very consistent and additive. The relative energies of the six BS coupling schemes of the F-cluster are 0, 1, 5, 5, 7, and 9 kJ mol^{-1} for the {αββα}, {ββαα}, {βαβα}, {αβαβ}, {αααββ}, {βααβ} states, respectively, with variations of only ~1 kJ mol^{-1}. Likewise, for the various H-cluster BS couplings, the average energies are –4, –3, 4, 8, 5, and 12 kJ mol^{-1} for the [αβαβ], [βαβα], [αααββ], [βααβ], [ββαα], and [αββα] states, respectively. Consequently, the [βαβα]-{αββα} and [αβαβ]-{αββα} states are most stable, with almost the same energy, –8 kJ mol^{-1}. The [βαβα]-{ββαα} and [αβαβ]-{ββαα} states are also close, –7 kJ mol^{-1}. As a whole, the results in Table 1 show that the total QM/MM energy of the 36 different BS states vary by less than 25 kJ mol^{-1}, when geometry-optimized models are considered.

Notably, the maximum variation in QM/MM energies almost doubles when single-point QM/MM calculations are run for the various states, using the geometry of the [αααββ]-{αββββα} state ($^{SP}\Delta E_{QM/MM}$ values in Table 1). Moreover, the order of the states is different and the optimized [αααββ]-{αββββα} state is now predicted to be most stable. This shows that it is mandatory to optimize the geometries to decide the relative energies of the various states. The 25 kJ mol^{-1} energy variation among optimized models turned out to be additive; In fact, the energy variation among the possible BS solutions of geometry-optimized QM/MM models that include either the H-cluster or the F-cluster in the QM region is 14 and 12 kJ mol^{-1}, respectively (models **2** and **3**, see Tables 2 and 3, and Figure 3).[II]

The variation of energies among the various spin couplings for model **1** is associated with differences in computed spin populations and charges that usually are negligible. In fact, the Mulliken charges of the $[2Fe]_H$, $[4Fe{-}4S]_H$ and F clusters vary in a range of less than 0.07 a. u. (Table 1). Likewise, the spin population of the F-cluster never changes more than 0.1 a. u. among the 36 BS states. On the other hand, the $[2Fe]_H$ and $[4Fe{-}4S]_H$ spin populations depend somewhat on the BS state of the $[4Fe{-}4S]_H$ subcluster. In fact, by grouping the models according to the BS coupling in the H-cluster, one can notice that the six different BS states have non-overlapping spin population ranges for the $[2Fe]_H$ subcluster of about 0.35, 0.30, 0.27, 0.16, 0.03 and −0.36 for the [βαβα], [αααββ], [βααβ], [ββαα], [αββα] and [αβαβ] states, respectively; for the same states, the $[4Fe{-}4S]_H$ spin population ranges are around −0.12, −0.07, −0.04,

[II] In Table 2, are also reported the charges and spin population values for the $[2Fe]_H$, $[4Fe{-}4S]_H$ subclusters in model **2**. Notably, they turned out to be consistent with the charges and spin populations of model **1** (Table 1), as far as models with the same BS coupling scheme at the $[4Fe{-}4S]_H$ subsite are compared.

−0.09, −0.19 and 0.55, respectively. In other words, in the case of the first five BS coupling schemes, spin populations are small and positive (or almost zero) for the $[2Fe]_H$ subcluster, while at the same time those for the $[4Fe{-}4S]_H$ assembly are negative and small. The picture is rather different in the case of the models showing the [αβαβ] coupling, as they are characterized by small and negative spin populations of the $[2Fe]_H$ subsite, while slightly larger and positive spin populations were computed for the $[4Fe{-}4S]_H$ subcluster. However, in all cases the overall spin population for the H-cluster is always less than 0.28 a.u., a result consistent with the diamagnetic state experimentally observed for the H-cluster in H_{red}.

Finally, we evaluated the effects of using a basis set smaller than TZVP, in terms of computed charges and spin populations. The relevance of this point is clear when one considers that the computational cost of the BP86-RI/TZVP QM/MM geometry optimizations is already very large, and it would rapidly become unfeasible in the case of further extension of the QM system. However, proteins showing a larger number of organometallic assemblies compared to [FeFe]-hydrogenases are common in nature, and their theoretical investigation could imply a fine-tuning of the level of theory applied. Therefore, we carried out the QM/MM geometry optimization of model **1** for the [αααββ]-{αββα} state, as well as for the two BS states with the lowest energy ([αβαβ]-{αββα} and [βαβα]-{αββα}), using the split valence polarized basis (SVP). Computed charges and Mulliken spin populations are reported in Table 4. It can be seen that, in all three cases, the smaller SVP basis leads to computed Mulliken charges and spin populations that never differ more than 0.03 *e* from those obtained with the TZVP basis set.

Finally, we looked at the Fe_p–Fe_d and Fe_p–S_{Cys} bond distances, together with average values of the other Fe–S bond lengths for the $[2Fe]_H$, the $[4Fe{-}4S]_H$ and the F-clusters, listed in Table 5. These data show that geometry variations between the two levels of theory are very small in all cases (maximum variation: 0.03 Å for the Fe_p–Fe_d distance and 0.02 Å for the Fe–S bond distances).

Conclusions

An extended broken-symmetry QM/MM study has been presented of the electronic properties of a large spin-coupled QM region composed by a Fe–S clusters chain embedded in a protein matrix, representing the latter at the MM level. The presence of several, non-equivalent spin centers in the enzymatic system here investigated – the [FeFe]-hydrogenase from *D. desulfuricans* – led to the possibility of defining 36 different BS wavefunctions for the two Fe_4S_4 clusters included in the quantum mechanical region. A simple approach for the rapid generation of all the various BS solutions has been described, based on the exchange of atomic coordinates among spin centers. Such an approach, which is of general use if the spin centers include the same metal elements, is particularly well suited in the case of QM/MM calculations, where it is important to keep the order of atoms within the inputs of the MM and the QM routines. However, when the latter constraint does not hold, e.g. in purely QM calculations, the generation of the various BS solution from an

already converged BS wavefunction can be easily extended also to systems that include different metal centers belonging to the same period of the Table of the Elements. In fact, such metal atoms are usually characterized by superposable sets of s, p, d, f,... basis functions in a standard QM calculation, and this allows to switch from one BS solution to the other by simply exchanging the positions of this atoms in the atomic coordinates file. Actually, this procedure is equivalent to exchanging the basis function coefficients between the metal atoms in the pair(s).

When compared to the approaches already available in several QM packages for the generation of the various possible BS solutions[III] – which are based on the exchange of localized α and β spin orbitals – our approach has the advantage that no change has to be made at the level of the open-shell wavefunction. This obviously allows for a more straightforward application and may favor SCF convergence. Similar considerations hold true when a comparison is made with the approach recently described by Szilagyi and coworkers,[36] which is based on the subdivision of the system in ionic fragments that are treated separately in order to localize spins, and then joined together in a single BS system. However, all the above mentioned approaches, together with a very recent one proposed by Reiher et al. – based on a constraining of the initial local spins to the desired ideal values –[37] can be used also to generate a specific BS solution from scratch, a feature that is not shared by the approach here described. In any case, the latter turned out to be very effective for the investigation of the extended QM system composed by the [FeFe]-hydrogenases H-cluster + F-cluster, given the availability of a fully converged BS solution obtained from standard SCF calculations, and the high number of different BS couplings that could be readily generated by targeted atomic coordinates exchanges. Results show that the QM/MM energy for such an enzyme model can vary by up to 25 kJ mol^{-1} among the 36 different BS solutions, but the energies of the BS states of the two Fe–S clusters are essentially additive. Considering that a ~10 kJ mol^{-1} variation was observed in QM/MM model complexes in which only one Fe_4S_4 cluster was part of the QM region (see ref.[21] and Tables 2–3), the present result shows that the variation in energy values among QM/MM optimized models having different BS spin configurations roughly doubles when one adds a Fe_4S_4 complex to the QM portion (as expected if the energies are additive). Our results also shows that it is mandatory to optimize the geometry of each BS state to obtain comparable energies.

Notably, small variations in the computed Mulliken charges were observed among the various BS states. This shows that, for the discussion of Mulliken charges, the choice of using a single BS spin configuration is well grounded not only in the case of small QM regions,[12,21,35] but also in the case of a larger and more complex QM system including multiple Fe–S clusters. As far as Mulliken spin populations are concerned, slightly larger variations among the various BS states were observed, thus suggesting that more caution is needed for the discussion of this property in models featuring an extended BS treatment. Finally, very small

[III] See for example TURBOMOLE, JAGUAR (Schrödinger, Inc.) and ORCA (Chair of Theoretical Chemistry, Bonn University; a recent publication on the use of BS in ORCA is: Baffert, C; Orio, M; Pantazis, DA, Neese, F et al. Inorganic Chemistry (2009) 48, pp. 10281-10288).

deviations of computed charges, spin populations and structural parameters were detected as a result of reducing the basis set extension from the triple-ζ valence polarized case to the split valence polarized one. This and the previous results will serve for the fine-tuning of the level of theory to be used for the characterization of electronic properties in large biological electron-transfer chains.

Figures legends:

Figure 1. Arrangement of Fe–S clusters in DdH

Figure 2. Optimized structure of the QM region and BS spin pattern in model **1** in the [αααβ]-{αββα} state (see main text for BS states nomenclature). Upper half: iron atoms numbering in the Fe_4S_4 assemblies used in Table 1. Lower half: spin topology obtained from SCF calculations on model **1** without using any spin localization technique. Arrows pointing up or down indicate alpha or beta local spin excess, respectively. The atom colors chosen are the following: red for Fe, yellow for S, white for H, blue for N, pink for C, and light red for O

Figure 3. QM portions in model **2** in the [αααβ] state (left) and model **3** in the {αββα} state (right). Arrows pointing up or down indicate alpha or beta local spin excess, respectively. The atom colors code is the same as in Figure 2

References

1. Solomon, E. I.; Randall, D. W.; Glaser, T. Coord Chem Rev 2000, 200, 595-632.
2. Sharp, R. E.; Chapman, S. K. Biochim Biophys Acta. 1999, 1432, 143-158.
3. Watt, G. D.; Reddy, K. R. N. J Inorg Biochem 1994, 53, 281-294.
4. Angove, H. C.; Yoo, S. J.; Burgess, B. K.; Munck, E. J Am Chem Soc 1997, 119, 8730-8731; Spiro, T.G., Iron-sulfur proteins. Wiley: New York, 1982.
5. Noodleman, L.; Norman, J. G. J Chem Phys 1979, 70, 4903-4906.
6. Noodleman, L. J Chem Phys 1981, 74, 5737-5743.
7. Noodleman, L.; Peng, C. Y.; Case, D. A.; Mouesca, J. M. Coord Chem Rev 1995, 144, 199-244.
8. Mouesca, J. M.; Chen, J. L.; Noodleman, L.; Bashford, D.; Case, D. A. J Am Chem Soc 1994, 116, 11898-11914.
9. Konecny, R.; Li, J.; Fisher, C. L.; Dillet, V.; Bashford, D.; Noodleman, L. Inorg Chem 1999, 38, 940-950.
10. Torres, R. A.; Lovell, T.; Noodleman, L.; Case, D. A. J Am Chem Soc 2003, 125, 1923-1936.
11. Bruschi, M.; Greco, C.; Zampella, G.; Ryde, U.; Pickett, C. J.; De Gioia, L. C R Chimie 2008, 11, 834-841.
12. Bruschi, M.; Greco, C.; Kaukonen, M.; Fantucci, P.; Ryde, U.; De Gioia, L. Angew Chem Int Ed Engl 2009, 48, 3503-3506.
13. Nicolet, Y.; Piras, C.; Legrand, P.; Hatchikian, C. E.; Fontecilla-Camps, J. C. Structure 1999, 7, 13-23.
14. Silakov, A.; Wenk, B.; Reijerse, E.; Lubitz, W. Phys Chem Chem Phys 2009, 11, 6592-6599; Ryde, U.; Greco, C.; De Gioia, L.; J Am Chem Soc 2010, 132, 4992-4993.
15. Patil, D. S.; Moura, J. J. G.; He, S. H.; Teixeira, M.; Prickril, B. C.; Dervartanian, D. V.; Peck, H. D.; Legall, J.; Huynh, B. H. J Biol Chem 1988, 263, 18732-18738.
16. Pereira, A. S.; Tavares, P.; Moura, I.; Moura, J. J. G.; Huynh, B. H. J Am Chem Soc 2001, 123, 2771-2782.
17. Lubitz, W.; Reijerse, E.; van Gastel, M. Chem Rev 2007, 107, 4331-4365.
18. Liu, Z. P.; Hu, P. J Am Chem Soc 2002, 124, 5175-5182.
19. Cao, Z. X.; Hall, M. B. J Am Chem Soc 2001, 123, 3734-3742.
20. Bruschi, M.; Zampella, G.; Fantucci, P.; De Gioia, L. Coord Chem Rev 2005, 249, 1620-1640.
21. Greco, C.; Bruschi, M.; De Gioia, L.; Ryde, U. Inorg Chem 2007, 46, 5911-5921.
22. Greco, C.; Bruschi, M.; Heimdal, J.; Fantucci, P.; De Gioia, L.; Ryde, U. Inorg Chem 2007, 46, 7256-7258.
23. Ryde, U. J Comput Aided Mol Des 1996, 10, 153-164.
24. Ryde, U.; Olsson, M. H. M. Int J Quantum Chem 2001, 81, 335-347.
25. Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. Chem Phys Lett 1989, 162, 165-169.
26. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179-5197.
27. Becke, A. D. Phys Rev A 1988, 38, 3098-3100.
28. Perdew, J. P. Phys Rev B 1986, 33, 8822-8824.
29. Schafer, A.; Huber, C.; Ahlrichs, R. J Chem Phys 1994, 100, 5829-5835.
30. Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R. Chem Phys Lett 1995, 240, 283-289.
31. Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Theor Chem Acc 1997, 97, 119-124.
32. Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. J Physical Chem A 2000, 104, 1720-1735.
33. Besler, B. H.; Merz, K. M.; Kollman, P. A. J Comput Chem 1990, 11, 431-439.
34. Fiedler, A. T.; Brunold, T. C. Inorg Chem 2005, 44, 9322-9334.
35. Stiebritz, M. T.; Reiher, M. Inorg Chem 2009, 48, 7127-7140.
36. Szilagyi, R. K.; Winslow, M. A. J Comput Chem 2006, 27, 1385-1397.
37. Herrmann, C.; Podewitz, M.; Reiher, M. Int. J. Quantum. Chem 2009, 109, 2430-2446.