chapter nine

QM/MM APPROACH

active site geometries of copper proteins

9

CONTENTS

SUMMARY

Using hybrid QM/MM calculations to optimize the geometry of the active site of metalloproteins in the presence of the protein and a layer of solvent molecules is presented in this chapter. A new link model for use in these QM/MM studies is presented that minimizes the influence of the introduction of artificial capping atoms is presented, which is used in subsequent optimizations of the active sites of wildtype, mutated and metal-substituted azurin.

9.1 *AddRemove*

A new link model for use in QM/MM studies

Hybrid QM/MM methods[149] split a system under study up in two parts: the electronically more important one that is treated with quantum mechanics (*real* QM system) while the remaining part is treated on a classical level (MM system). The division into the two subsystems is sometimes trivial as in the case of solvation of an organic compound where one puts the organic compound in the QM system and the solvent in the MM system. However, in biochemical systems the division often results in bonds that cross the QM/MM boundary. Several approaches have been used to handle this situation (see ref. 149), like for instance localized orbitals[310], pseudopotentials[150,311-313], but the easiest and most commonly chosen solution is to use link atoms[314,315]. In the latter concept one introduces a (artificial) capping atom in the quantum system (comprising then the *capped* QM system), which is usually taken to be a hydrogen atom. Singh and Kollman[314] (who originally proposed the concept) didn't put any constraints on the position of the capping atom, thereby adding three degrees of freedom to the system. To avoid the addition of these unwanted and artificial degrees of freedom, many later developed link models impose the capping atom LC, the quantum atom LQ and the classical atom LM to have a mutual coordinate dependence (for instance to lie on a straight line), thereby removing again the additional degrees of freedom (see Figure 9.1.1).

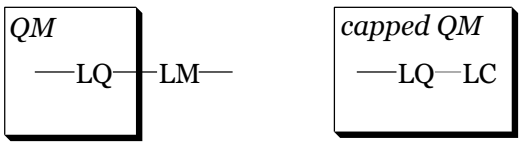

FIGURE 9.1.1. SCHEMATIC REPRESENTATION OF LINK BOND

Although there are many link models in the recent literature[151-154,316], the newly developed model as described in this section (AddRemove), will be compared primarily with the currently implemented model in the ADF[117] program, as this model is more or less representative for the other models. After introducing the new model, a comparison is made also with other (well known) models.

In the current implementation in ADF[117] (IMOMM/ADF)[155-157], the degrees of freedom of the real classical atom LM are removed: it "follows" the capping atom LC with its position depending on the positions of the LQ and LC atoms as:

$$\vec{r}_{LM} = \alpha\ \vec{r}_{LC} + (1-\alpha)\ \vec{r}_{LQ} \qquad (1)$$

In this equation, α is an arbitrary constant parameter but its value is difficult to generalize or choose. The energy depends on the value chosen for it and one can compare energies between different molecules only if the αs are chosen equal for both.

The capping atom interacts only within the *capped* QM system and has no interactions with the MM environment. Moreover, the real classical LM atom does not interact with the real QM system; its interactions with the QM system are "replaced" by the QM interactions of the capping atom LC. The real classical LM atom of course does have MM interactions with the rest of the MM system, but they

are not used for updating its position (since it is constrained to "follow" the LC atom).

Before doing the QM calculation, the MM system is optimized first. In this part, the atoms in the *real* QM system are kept fixed at their positions, and interact with atoms in the MM system through classical MM interactions. In the IMOMM/ADF implementation[155-157], also the classical LM atoms are kept fixed in this part. A schematic example for the optimization scheme is given on the left in Figure 9.1.2.

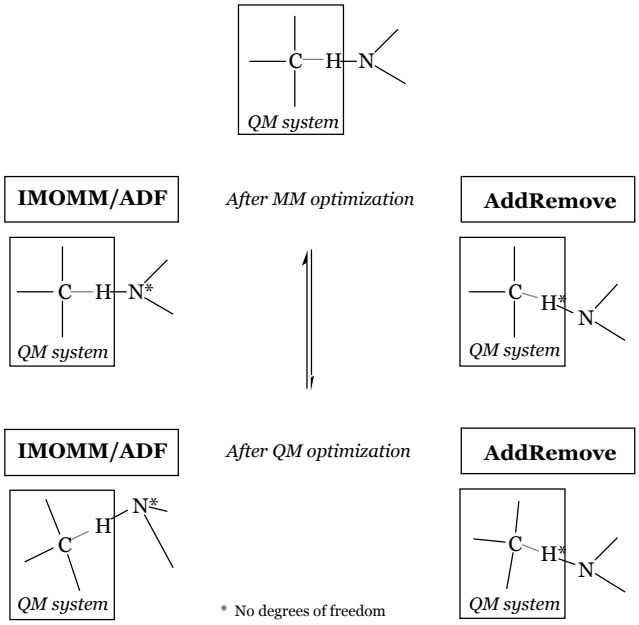

FIGURE 9.1.2. QM/MM OPTIMIZATION SCHEME

After the MM optimization has ended, the QM calculation is performed, and the QM energy, gradients and charges calculated. Then the MM system is optimized again before starting a new QM cycle, and so on, until the QM system is fully optimized.

One of the advantages of the IMOMM and AddRemove model is that the interactions between the QM and MM system are treated at the same level of theory as the interactions within the MM system itself. Then, if one uses for those interactions the AMBER95 or GROMOS96 force fields[126,127] (which is constructed for treating the interactions within proteins, and is known to perform well for them), one is certain that the interactions (in the MM system itself as well as between the QM and MM systems) are treated properly.

AddRemove methodology

In the AddRemove model, the position dependence is reversed: it is the capping atom LC that "follows" the classical atom LM, and not vice versa:

$$\vec{r}_{LC} = \vec{r}_{LQ} + R_{eq}\vec{u}_{LQ-LM} \tag{2}$$

Here, $\vec{u}$ is a unit vector pointing from atom LQ to atom LM (see Figure 9.1.1 for atom labels), and $\boldsymbol{R_{eq}}$ is the equilibrium distance for the LQ-LC bond it would have in the classical force field that is used for

the MM system. This has the major advantage that one does not remove the degrees of freedom of the real classical atom LM (as is the case in the IMOMM/ADF model[155-157]). Instead, the artificial degrees of freedom of the capping atom LC are removed. Also, unlike the IMOMM/ADF implementation[155-157], the energy does not depend anymore on some arbitrary parameter α. Furthermore, the real classical LM atoms interact normally with the *real* QM system, and are allowed to move in the optimization of the MM system (see also the right hand side of Figure 9.1.2). Then, before the QM calculation starts, the positions of the capping atoms are updated with eq. (2) from the current positions of the LM atoms. After the QM energy and gradients have been obtained, they are corrected for by removing the interactions of the capping atoms: the MM interactions, which the capping atoms would have with the atoms in the real QM system if it were a classical atom, are subtracted. That is, the capping atoms are Added and Removed. Schematically, this would result in the following energy expression for the interaction between the QM and MM systems:

$$U^{QM/MM} = \sum_{\substack{j\ MM \\ i\ realQM}} \sum_{j \neq LM} U^{MM}(i,j) + \sum_{n=1}^{Nlinks} \sum_{i\ realQM} U^{QM}(i,LC_n) - U^{MM}(i,LC_n) + U^{MM}(i,LM_n) \qquad (3)$$

The first summation runs over all classical atoms that are not involved in the link bond, which couple normally with the *real* QM system through standard MM interactions. The second part is specific for the atoms involved in the link bonds. Here, the "quantum" interactions of the capping atoms LC with the *real* QM system ($U^{QM}(i,LC_n)$) are corrected for by subtracting its MM counterparts ($U^{MM}(i,LC_n)$), and replaced by the MM interactions of the classical LM atom with the *real* QM system ($U^{MM}(i,LM_n)$). Of course, it is not possible to project out the $U^{QM}(i,LC_n)$ interaction, but schematically this is justified. Moreover, after summation and combination with the "QM energy" of the *real* QM atoms, simply the total QM energy (of the capped QM system) is recovered.

For the forces working on the atoms, the same procedure is followed. That is, the following expression is valid for the gradients of the interaction energy:

$$g^{QM/MM} = \sum_{\substack{j\ MM \\ i\ realQM}} \sum_{j \neq LM} g^{MM}(i,j) + \sum_{n=1}^{Nlinks} \sum_{i\ realQM} g^{QM}(i,LC_n) - g^{MM}(i,LC_n) + g^{MM}(i,LM_n) \qquad (4)$$

This expression follows directly from using eq. 3 for the energy expression. Note that this expression is one of the things where the AddRemove model differs from other link models (like ONIOM[151-153]); this topic is discussed in more detail in the **Comparison** part.

For convenience, the gradient of the artificial LC atoms is set to zero, since any resulting nonzero gradient would be the result of the inability of the MM force field to represent QM interactions; this effect is ignored also for the MM system. Moreover, the position of the capping atom is not a degree of freedom, and leaving the gradients of the capping atoms nonzero would affect only the convergence of the optimization procedure of the QM system, **not** the positions of the atoms.

The electrostatic interactions between the *real* QM and MM systems are obtained in a classical manner, but can be handled in an iterative way (*updated electrostatic coupling*): during the optimization procedure, the charges of the QM system are updated after each QM optimization cycle and used in the next MM optimization cycle. The charges for the QM system are obtained from a recently developed charge analysis, which is derived from an atomic multipole expansion (the Multipole Derived Charge analysis)[183]. In the ADF program[117], the Coulomb potential can be obtained from a set of atomic multipoles[118] (together with a short range function). These atomic multipoles are

reproduced *exactly* by the charges, and therefore also the molecular multipoles as well as the Coulomb potential. In this approach, we take as much data from the QM calculations as possible.

Comparison with other link models

The expression for getting the interaction energy seems similar to ones used by other link models like for instance the ONIOM model[151-153], when the latter is used with 2 layers. However, this similarity is only present in the energy expression. The specific implementation of this expression differs to a great extent.

Unlike the ONIOM model[151-153], the degrees of freedom of all real atoms are completely free, i.e. all *real* atoms are completely free to move. This is in contrast to the ONIOM method, where not only the bond between the real quantum atom LQ and the capping atom LC, but also the bond between the *real* atoms LQ and LM is frozen at a certain predefined distance. This is similar to the introduction of the IMOMM/ADF[155-157] model α-parameter, upon which the energy and geometry depends. In fact, it is found that depending on the specific value given for this parameter (g parameter in ONIOM terminology), the bond between the *real* LQ and LC atoms ranges from 1.46 to 1.67 Å for a CC bond when the parameter α is varied from 0.5 to 1.0[151].

More importantly, the gradient expressions don't follow naturally from the energy expressions; unlike our implementation, where not only the energy but also the gradient is corrected (see eqs. 3 and 4), in the ONIOM method the gradient of the capping atoms is *projected* onto the LQ and LC atoms. This not only introduces the coordinates of the capping atoms as degrees of freedom, but more importantly, introduces inconsistencies in the total picture of the system. Finally, unlike the ONIOM method, which needs several calculations for several systems on a different level that are coupled in a not always well-defined manner, the AddRemove model can be used directly within a QM scheme. The only modifications needed are the optimizations of the MM system before every QM step, followed by adding the QM/MM interaction energies and forces to the QM and MM forces to get the total energies and forces working on the atoms. In this respect, the AddRemove model is more similar to the QMPot model by Sauer et al.[154]. However, unlike their work, the interactions between the QM and MM systems are simply taken from a classical force field (AMBER95[126]) that is known to perform well for treating interactions within proteins; therefore, no additional assumptions have to be made.

In the **Results** we will compare results from using the AddRemove model with those found for the IMOMM/ADF model[155-157]; although there are differences between the IMOMM[155-157] and ONIOM[151-153] energy expressions, both use the same gradient expressions (by projecting the gradient of the artificial link atoms onto its neighboring atoms). As the geometry obtained in a geometry optimization depends only on the gradient, *not* on the energy, the IMOMM results are representative also for a 2-layered ONIOM model.

Results

In Sections 9.2-9.4, the results of using a QM/MM approach for getting the active site geometries of copper proteins are reported, where the AddRemove model has been used to couple the QM and MM systems. In those sections, the QM system consists of the complete active site that is surrounded by a MM system, which consists of the rest of the protein as well as a layer of solvent molecules surrounding the protein.

In order to check the validity of the AddRemove model for treating the interactions involved in cutting such a biochemical system into a QM and MM system, a few amino acid residues were treated either by a full QM, or a QM/MM description. As the main objective in the study on copper proteins (see Sections 9.2 to 9.4) is the active site geometry, in this section only the geometry of these residues will be discussed. Moreover, as it is the gradient expression (see equation 4) in which the AddRemove differs from other link models (like ONIOM[151-153]), and the geometry is determined completely by the gradient, not the energy, this enables a meaningful comparison between AddRemove and other models.

For the QM/MM description, both the IMOMM/ADF[155-157] and the AddRemove model have been used. The residues studied are methionine (Met), cysteine (Cys) and tyrosine (Tyr). The amino acid residues were described with a dipeptide surroundings, where on both sides of the backbone connection, the peptide bond is included explicitly. The residues were studied with either the complete molecule in the QM system (*full QM*), or with the QM system cut off at the C-alpha position with hydrogens added to complete the link bonds (*QM/MM)*. In the QM/MM description, the electrostatic coupling of the QM and MM systems was either done by a *simple* or an *update* approach. In the latter approach, the charges of the QM atoms are updated after every QM geometry optimization cycle, while in the *simple* approach the standard AMBER95 charges were used.[126]

The calculations were performed with the ADF program[117,187], using the Becke[120]-Perdew[121] exchange-correlation potential in a triple zeta valence plus polarization (TZP) basis set. In another study[317], where a number of exchange-correlation potentials available in ADF were tested in several standard basis sets, it was found that the Becke-Perdew xc-potential in that basis set predicts the bond distances of a certain set of molecules with an average absolute deviation of 1.24 pm. This value gives an indication of the accuracy of the bond distances that could be reached.

TABLE 9.1.1. LINK BOND DISTANCES (PM) FROM *FULL QM* AND *QM/MM* STUDIES

	full QM	*IMOMM simple el.st.*	*IMOMM updated el.st.*	*AddRemove simple el.st.*	*AddRemove updated el.st.*
cysteine					
Cα-N	146.0	151.7	151.1	147.1	146.7
Cα-C	155.6	157.8	158.9	154.3	155.4
Cα-Cβ	153.3	153.6	153.5	152.5	152.6
methionine					
Cα-N	145.8	152.1	151.6	147.3	145.7
Cα-C	154.2	157.1	157.7	153.9	155.2
Cα-Cβ	154.7	154.3	154.4	153.4	153.2
tyrosine					
Cα-N	146.9	151.5	151.8	147.0	147.8
Cα-C	155.9	157.7	157.6	154.4	153.4
Cα-Cβ	153.1	154.5	154.7	152.7	153.0

The most relevant property to look at in these calculations, is the distance between the *real* atoms involved in the link bonds; as the system is cut off at the C-alpha position, there are two distances per

molecule, which result from the bonds of the C-alpha to the peptide-bond on either side (C_α-$N_{backbone}$ and C_α-$C_{backbone}$ distances). For convenience, the distance between the C_α and C_β atoms are also given to be able to check the performance of the methods in that respect. The distances for the *full QM* and IMOMM/AddRemove *QM/MM* descriptions are given in Table 9.1.1. These results show a rather poor performance for the IMOMM/ADF method[155-157] to describe the link bonds; the average absolute deviation in the C_α-$C_{backbone}$ bond is 2.3 pm (*simple* electrostatic coupling) or 2.8 pm (*updated* electrostatic coupling), while it is even higher for the C_α-$N_{backbone}$ bond: 5.5 (*simple*) and 5.3 (*updated*) pm. The other two *regular* bonds of the C-alpha atoms are well described with an average absolute deviation of 0.7 pm (C_α-C_β) and 0.1 pm (C_α-H; not shown here in the Table). Taking the C_α-C_β bond as a reference for the accuracy that could be reached by putting part of the molecule in the MM system, this means an increase by a factor of 3.1-4.0 for the C_α-$C_{backbone}$ bond deviation and 7.1-7.3 for the C_α-$N_{backbone}$ bond.

The AddRemove model performs much better in that respect with average absolute deviations of 0.9 (*simple*) or 0.6 (*updated*) pm for the C_α-$N_{backbone}$ bond, 1.0 (*simple*) and 1.2 (*updated*) pm for the C_α-$C_{backbone}$ bond. The regular bonds are equally well described as indicated by the average absolute deviation of 0.8 pm for the C_α-C_β bond, and < 0.1 pm for the C_α-H bond (data not shown in Table). By comparing again the accuracies of the link bonds with the one found for the C_α-C_β bond deviation, we find factors of 1.3-1.6 for the C_α-$C_{backbone}$ bond and 0.8-1.2 for the C_α-$N_{backbone}$ bond, which is a substantial improvement over the results of the IMOMM model[155-157].

Conclusions

The division of a system under study in a quantum mechanical and a classical system in QM/MM calculations is sometimes very natural (as in the case of solvation of an organic compound, where the organic compound can be treated at the QM level and the solvent at the MM level), but a problem arises in the case of bonds crossing the QM/MM boundary. A new link model is presented that uses a capping (link) atom to satisfy the valences of the quantum chemical system, with the position of the *artificial* capping atom depending on the positions of the real atoms involved in the link bond. Therefore, the introduction of the capping atom does not lead to additional degrees of freedom; moreover, the degrees of freedom of all *real* atoms are kept (unlike other link models). Furthermore, the introduction of this artificial atom is corrected for after every QM optimization cycle by subtracting the classical molecular mechanics interactions with the real QM system it would have if it were a classical atom. Charges of the *real* QM atoms can be updated after every QM geometry cycle, and used in the electrostatic coupling between the QM/MM systems in the subsequent optimization of the MM system.

The new model is compared to other (well-known) link models, and tested in comparison with the other link model implemented in the ADF program (IMOMM/ADF) on a few amino acid residues. In these test calculations the residue is described either by a complete QM description (*full QM*) or by a QM/MM description, where the QM system of the residue has been cut off at the C-alpha position. For the QM/MM description, both the AddRemove and IMOMM model has been used, either with *simple* or *updated* electrostatic coupling between the QM and MM systems. The distances between the C-alpha and the other two real atoms involved in the link bonds, differ only marginally when using either one of these two electrostatic couplings. This is a positive result, especially in view of the systems where the QM/MM method will be applied to (metalloproteins). For those systems, there are no predefined charges in the force field, and one has no other option than to use the updated ones.

A large difference is found between the performance of the IMOMM/ADF and AddRemove models. Taking the deviation for predicting the distance of the regular C_α-C_β bond (0.7 pm) as a reference value when part of the system has been put in the MM system, the IMOMM model gives a rather poor performance: the deviation increases by a factor of 3.1-4.0 for the C_α-$C_{backbone}$ bond and 7.1-7.3 for the C_α-$N_{backbone}$ bond. The AddRemove model gives a much better description; the deviation changes only slightly, by a factor of 1.3-1.6 for the C_α-$C_{backbone}$ bond and 0.8-1.2 for the C_α-$N_{backbone}$ bond.

9.2 *Wildtype azurin*

Optimization of the active site geometry in both the reduced and oxidized state

There have been many discussions over the last decades about the influence of the protein environment on the active site of copper proteins.[55,108,109] It has been argued that the protein forces the active site in a strained conformation, which is a compromise between the favored geometry of the reduced and oxidized state. Recently, this view has been challenged by Ryde et al.[11-14,17,20], who performed extensive studies on small model systems of the active site of type-1 copper proteins. They found active site geometries that are remarkably similar to the ones observed in plastocyanin, a typical type-1 copper protein that lacks the axial Gly residue observed in azurin. However, one feature that could not be reproduced by these model systems is the distance of the axial Met ligand, which was nearly always observed at too close a distance from copper. In another investigation by de Kerpel et al.,[104] the same procedure has been applied to study *wildtype* and Co-substituted azurin. Again problems were encountered for the axial ligands of the *wildtype* azurin model, with a distance of 7.18 Å for Met121 and 2.34 Å for Gly45. The Met121 residue has moved away completely and formed a hydrogen bond to the H_ε-atom of one of the imidazole groups.[104]

TABLE 9.2.1. OPTIMIZED DISTANCES (Å) OF *WILDTYPE* AZURIN ACTIVE SITE *IN VACUO I*

	Starting structure	*X-ray*[280] *reduced state*	*in vacuo I reduced state*	*X-ray*[7] *oxidized state*	*in vacuo I oxidized state*
Cu-Gly45[a]	2.955	3.021 ± 0.077	3.196	2.967 ± 0.093	2.981
Cu-His46[a]	2.064	2.133 ± 0.093	2.138	2.076 ± 0.060	2.038
Cu-Cys112[a]	2.267	2.293 ± 0.013	2.229	2.237 ± 0.044	2.177
Cu-His117[a]	1.978	2.095 ± 0.091	2.066	2.011 ± 0.069	2.012
Cu-Met121[a]	3.164	3.247 ± 0.069	3.459	3.149 ± 0.070	3.214
Cu-N2S[b]	0.043	0.092 ± 0.013	0.098	0.083 ± 0.051	0.100

a) distance of Cu to closest atom in this residue;
b) distance of copper from plane made by the S_γ(Cys112), N_δ(His46) and N_δ(His117) atoms

In Section 6.3, it was shown that with the Becke[120]-Perdew[121] xc-potential in a larger basis set, the potential energy surface of the axial copper-Met121 distance is very flat in the range from 3.0 to 4.0 Å. Therefore, just like observed by de Kerpel et al.[104], the Met121 residue has a large freedom to move and no well-defined energy minimum is observed. In contrast, for the other axial group, Gly45, a well defined energy minimum was observed around 2.9 Å. However, the Gly45 residue is connected to one of the in-plane ligands (His46) and as only the Gly45 residue was allowed to move in Section 6.3, the localization of such an energy minimum may have been forced by the connection to His46. A geometry optimization of the whole active site should give an indication of whether the Gly45 energy minimum is forced by the His46 connection, or not. Moreover, it can provide some more insight on the role of the protein on the active site geometry. Therefore, starting from the structure that was used in Section 6.3, a full geometry optimization was performed without any constraints on the atoms. The optimized copper-ligand distances are given in Table 9.2.1. Given in Table 9.2.1 are also the statistical errors from

the X-ray data[7,280] obtained after averaging the distances over the four molecules in the asymmetric unit. A remarkable similarity in these numbers is observed, apart from the distances of the axial groups in the reduced state. The difference with the X-ray data is at the edge of the uncertainty in the X-ray data, which is estimated to be 0.1-0.2 Å. Still, a satisfactory agreement between experimental and calculated distances is obtained. The difference between the two Histidine distances is properly predicted and the decrease in distance in going from the reduced to the oxidized state is observed as well.

Performing the geometry optimization with another starting structure for the active site (also taken from X-ray data[a 5]) however can have a marked influence on the obtained geometry (see Table 9.2.2), which indicates that for the active site *in vacuo* local minima exist with virtually the same energy.

TABLE 9.2.2. OPTIMIZED DISTANCES (Å) OF *WILDTYPE* AZURIN ACTIVE SITE *IN VACUO II*

	Starting structure	*in vacuo II reduced state*	*in vacuo II oxidized state*
Cu-Gly45[a]	3.066	3.007	2.480
Cu-His46[a]	2.208	2.050	2.040
Cu-Cys112[a]	2.216	2.217	2.180
Cu-His117[a]	2.362	2.126	2.023
Cu-Met121[a]	3.007	4.738	4.317
Cu-N2S[a]	0.056	-0.053	-0.130

a) see subscript Table 9.2.1 for definition

In this case, the Met121 residue has moved away from the copper completely, and is found at 4.7 Å in the reduced state and at 4.3 Å in the oxidized state. The flat potential energy surface between 3.0 and 4.0 Å for the copper-Met121 distance may be a large contributor to this drifting away of Met121. At the same time, Gly45 moves closer to the copper, in the oxidized state it is even found at 2.5 Å. The differences between calculated and experimental copper-residue distances are now too large to be "accounted for" by the experimental uncertainty. Therefore, it has to be concluded that if one wants to obtain good active site geometries, one can not ignore the influence of the protein: it should be included in the optimizations.

Hybrid QM/MM calculations

In principle, if one wants to include the whole protein in the optimizations, one could treat it completely by quantum chemical methods. However, even with the linear scaling techniques[118] that are currently in use and despite the enormous amount of computing power, performing a geometry optimization on a system of some 2000 atoms in an acceptable basis set is probably still awkward to do. Moreover, precious CPU time would be wasted as a detailed and therefore expensive description is used for large regions in space, which could be very well described by classical interactions by using specially designed protein force fields like AMBER95[126] or GROMOS96[127]. Therefore, a more appropriate and efficient way to include the protein in the geometry optimizations is by performing

[a] One of the molecules in the asymmetric unit of F114A-azurin, which shows a large similarity with *wildtype* azurin, apart from the elongated Cu-His distances

hybrid QM/MM calculations (see Section 2.3). In this case, one puts the place of interest (active site) in the QM region while the rest of the protein and solvent molecules are put in the MM system. In Section 9.1, a new link model (AddRemove[158]) to connect the two systems has been described, which uses capping atoms to satisfy the valency of the QM system. These capping atoms are necessary in the case of proteins, when covalent bonds are crossing the QM/MM boundary. The introduction of these capping atoms is artificial, and if not accounted for, would lead to an artificial increase of the degrees of freedom, as well as QM interactions that should not be present. Therefore, after every QM cycle, the QM interactions of these artificial capping atoms with the real QM atoms are "removed" by subtracting their counteracting MM interactions (see Section 9.1). The positions of the capping atoms are constrained to lie in the direction of the covalent bond crossing the QM/MM boundary, and therefore no artificial degrees of freedom are added to the system. Therefore, the capping atoms are "added" and "removed".

A hybrid QM/MM calculation has been performed using the QM/MM code[155] in the ADF program[117,187], which has been adapted and optimized in order to be able to treat proteins[a]. A conjugate gradient optimization scheme[1] has been added and used, the calculation of the forces has been parallellized and the memory allocation scheme of the van der Waals interactions adjusted to make it use the available memory more efficiently. The Becke[120]-Perdew[121] exchange-correlation potential has been used in a TZP basis set, with an integration accuracy of 6.0. For the construction of the ADF[117,187] inputfile, the pdb2adf[318] program has been used. After inclusion of the protein, the total number of QM atoms was 61, the total number of protein atoms 1930 and total size of the system ranging from 2200 to 14000, depending on whether a shell of water molecules surrounding the whole protein has been added. In order to keep the molecules from evaporating from the system, a (spherical) wall force has been introduced that acts to keep the atoms in an imaginary spherical box with a radius of 33 Å. The AddRemove[158] model (Section 9.1) has been used to couple the QM and MM system, with "updated" electrostatic coupling. This means that after a QM cycle, the charges on the QM atoms have been updated to their MDC-q values, which come from the Multipole Derived Charge analysis[183]. These charges are subsequently used in a classical electrostatic fashion to couple the QM and MM system.

The inclusion of the protein has a positive effect on the optimized copper-residue distances (see Table 9.2.3). Not only is again a good agreement observed between calculated and experimental[7] copper-residue distances, the same active site geometry is observed when started from another starting structure. Apart from the axial Met121, all residues are found closer to copper than in the X-ray data[7,280], and agree well with EXAFS data[319] that put the in-plane ligands closer to copper than X-ray does. The axial groups are too far away from copper to be "seen" by EXAFS[319].

The decrease in these distances in going from the reduced to the oxidized state is well predicted by the QM/MM calculations. Unlike the X-ray data[7,280], which place the two in-plane Histidines at distances that differ by some 0.04-0.06 Å, the QM/MM calculations predict that they are at much more similar distances from copper, with differences of 0.02 Å or less. Cys112 is positioned at 2.24 Å from copper in the reduced state and 2.17 Å in the oxidized state, which is at still somewhat larger distances than in other type-1 copper proteins like plastocyanin, but compares very well to EXAFS studies that place it at around 2.15 Å in the oxidized state and at 2.20 Å in the reduced state[25,319]. The decrease in this distance by about 0.05 in going from the reduced to the oxidized state is a feature preserved in the EXAFS and X-ray data, as well as in the QM/MM calculations.

[a] Adaptation and optimization of QMMM code by M. Swart for this thesis

TABLE 9.2.3. QM/MM OPTIMIZED DISTANCES (Å) OF *WILDTYPE* AZURIN ACTIVE SITE

	X-ray[280] *reduced state*	*QM/MM reduced state*	*X-ray*[7] *oxidized state*	*QM/MM oxidized state*
Cu-Gly45[a]	3.021 ± 0.077	3.003	2.967 ± 0.093	2.855
Cu-His46[a]	2.133 ± 0.093	2.069	2.076 ± 0.060	2.017
Cu-Cys112[a]	2.293 ± 0.013	2.238	2.237 ± 0.044	2.169
Cu-His117[a]	2.095 ± 0.091	2.076	2.011 ± 0.069	1.995
Cu-Met121[a]	3.247 ± 0.069	3.393	3.149 ± 0.070	3.178
Cu-N2S[a]	0.092 ± 0.013	0.065	0.083 ± 0.051	0.094

a) see subscript Table 9.2.1 for definition

Also for the angles a good agreement is observed between calculated and experimental[7,280] (X-ray) data (see Table 9.2.4). The difference between the two is at maximum some 6 degrees, which is well within the limits posed on the angles by the experimental uncertainty of the X-ray coordinates. Again a good agreement between calculated and experimental results is observed when looking at the difference in angles in going from the reduced to the oxidized state. In nearly all cases when an angle is found by X-ray data to decrease (or increase) upon oxidation, the same trend is observed by the QM/MM calculations.

TABLE 9.2.4. QM/MM OPTIMIZED ANGLES (°) OF *WILDTYPE* AZURIN ACTIVE SITE

	X-ray reduced state	*QM/MM reduced state*	*X-ray oxidized state*	*QM/MM oxidized state*
S_{112}-Cu-N_{46}	133	137	132	130
S_{112}-Cu-N_{117}	122	123	122	127
N_{46}-Cu-N_{117}	105	99	105	102
S_{112}-Cu-S_{121}	113	115	110	111
S_{112}-Cu-O_{45}	99	95	99	94
S_{121}-Cu-O_{45}	145	149	148	151
N_{117}-Cu-O_{45}	87	84	89	85
C_{45}-O_{45}-Cu	128	125	130	130
C_{112}-S_{112}-Cu	106	104	109	110
C_{121}-S_{121}-Cu	137	128	140	131
$C^{\gamma}{}_{46}$-N_{46}-Cu	130	129	131	129
$C^{\varepsilon}{}_{46}$-N_{46}-Cu	126	125	125	124
$C^{\gamma}{}_{117}$-N_{117}-Cu	128	125	130	125
$C^{\varepsilon}{}_{117}$-N_{117}-Cu	127	129	124	128

Despite their extensive work on model systems[11-14,20,103,104], also Ryde et al. came to the conclusion that in order to get good active site geometries, the protein has to be included by using QM/MM calculations (for which they use the ComQum label)[17]. Especially if one wants to obtain the axial groups at the correct position, this seems to be of vital importance. Although their ComQum results perform better than their model studies in this respect for oxidized plastocyanin, it still does

not place the axial methionine at the right position in the reduced state. However, was the distance underestimated in the model system by some 0.5 Å, in the ComQum calculation it is overestimated by almost the same amount. The same problems with the axial ligands is observed in their study on azurin, where the copper-Met121 distance is too short by 0.3 Å in the reduced state and too long in the oxidized state by 0.5 Å. Also the decrease in copper-Cys112 distance by about 0.05 Å upon oxidation is less pronounced in their ComQum calculations. As they also use the AMBER95[126] force field for describing the interactions in the MM system, the difference with our results can only be coming from either the use of the B3LYP[122] xc-potential or the smaller basis sets.

Spin density distribution

Atomic spin densities can be obtained by taking the difference between the atomic charges resulting from either the spin-up or spin-down density. Like mentioned in Section 3.1, the Multipole Derived Charges[183] are obtained from both densities, and therefore in a spin-unrestricted calculation provide not only the atomic charge but also the atomic spin density.

The spin density present in the oxidized state is delocalized in the active site, with large amounts present on copper (20.7 %), the sulphur of Cys112 (52.7 %), the sulphur of Met121 (9.3 %) and the nitrogens of His46 (4.8 %) and His117 (5.8 %), while on the oxygen of Gly45 only a tiny amount is observed (0.8 %). These results therefore indicate that although the Met121 is at a long distance from the copper atom, it is weakly covalently bonded to it, while Gly45, which is located closer to copper than Met121, is not covalently bonded or, at best, only very weakly. Also some tiny amounts of spin density are found on nearby protons, as in the case of Met121-Hε (~0.5 %) and Cys112-Hβ (~0.8 %), His46-Hα/β (0.3-0.4 %), His117-Hβ (~0.5 %) and His117/46-Hδ/ε (~0.1 %).

Hydrogen bonds

One of the more important residues in the nearby surroundings of the active site seems to be the Asn47 residue, which is preserved at this position in all azurin molecules and other type-1 copper proteins found so far. It makes five hydrogen bonds, two of which are to residues not directly related to the copper site, two to Ser/Thr113 with its sidechain oxygen and nitrogen (at 2.95/2.87 Å) and one to the Cys112 sulphur with its backbone amide nitrogen at a N-S distance of 3.44 Å. These are all conserved in the QM/MM optimizations, as is another hydrogen bond to the Cys112 sulphur, by the backbone nitrogen of Phe114 with a N-S distance of 3.79 Å. Also, hydrogen bonds are observed connecting the backbone nitrogen of Cys112 and the backbone oxygen of Met121 at 2.96 Å, the backbone nitrogen of Met121 and the backbone oxygen of His117 at 2.75 Å, and the backbone nitrogen of His117 and the backbone oxygen of Phe114 at 3.03 Å.

Conclusions

The optimization of the geometry of an active site *in vacuo* does not always lead to structures corresponding to the active site as it is observed in the protein. The influence of the protein on the geometry is found to be a factor that can not be simply ignored. This is especially valid for the axial groups in type-1 copper proteins, especially Met121 that has an enhanced freedom to move *in vacuo* (see also Section 6.3). Without the inclusion of the protein, the active site geometry may sometimes be reasonably represented, but in many cases the site deforms. After inclusion of the protein in the geometry optimization by hybrid QM/MM calculations, a very good agreement is observed between

calculated and experimental copper-residue distances and angles. In these QM/MM calculations, the system of interest (active site) is put in the QM system while the rest of the protein and solvent molecules is placed in the MM system. The results presented here give a better description than the ComQum results by Ryde et al.

9.3 *Metal substitution and mutants*

The effect of metal substitution and mutations on the active site geometry

In the previous section, a hybrid QM/MM calculation was used to optimize the active site geometry of *wildtype* azurin. However, many experimental studies have focused not only on *wildtype* azurin, but also on metal-substituted azurin[4,33,36,54,59,80,104,320-322] or mutants of it[3-5,23,27,30,31,35-38,42,48,61,63,75,77,80-84,88,319,322-326]. The metal-substitution is either found by accident, as in the case of zinc-azurin[59] that has been found as a colorless by-product in the heterologous expression of *wildtype* azurin, or made on purpose. In the latter case, copper is substituted for either zinc, cobalt, nickel, manganese, cadmium or mercury, for a structural or spectroscopic characterization by different techniques like paramagnetic NMR spectroscopy, X-ray diffraction or UV-VIS absorption spectroscopy. Another widely used technique to study the function of amino acid residues in metal proteins is site-directed mutagenesis, in which one or more residues are substituted in order to study the influence of this mutation on the functions and properties of the metal protein, like reduction potential[3] and UV-VIS[36,54,81,319,322], EPR[23,38,74,302-304], and NMR[33,36,47,56,76,78,80,84,319-321,327-329] spectra.

In this section, the active site structure of a few metal substituted azurin molecules or azurin mutants for which a crystal structure is available, is optimized using QM/MM calculations. This enables a further check on the reliability of using the QM/MM method for obtaining active site structures, by comparing the calculated and experimental structures. Furthermore, the computed active site structure is used to obtain UV-VIS spectra and EPR data that can be checked with experimental data.

Zn-substituted azurin

In the preparation of azurin samples for use in spectroscopic studies, it was shown that the samples may contain another protein that co-purified with azurin and could not easily be distinguished from the latter.[59] In later studies, using among others new isolation procedures, it was possible to isolate the contaminating by-product, which was called azurin*. It was shown that this protein was highly similar to azurin, but was not the oxidized, reduced or apo-form of it, nor could it readily incorporate copper ions. Using several techniques, it was established that a “heavy” atom was bound in the active site, which was proposed and confirmed to be zinc based on, among other evidence, the absence of any features in the visible absorption spectrum. X-ray crystallographic data[59] showed a distinctly different binding of the metal to the active site residues. Compared to wildtype azurin, the zinc atom moves 0.3 Å down towards Gly45, which moves 0.3 Å upwards itself. Unlike copper, which is found at 0.08 Å above the NNS-plane of His46, His117 and Cys112, zinc is located at 0.16 Å below this plane. As a result, the metal-Gly45 distance is decreased from 2.97 Å for copper to 2.32 Å for zinc.

A hybrid QM/MM calculation was performed to optimize the active site geometry of zinc-azurin, where the starting structure was taken from wildtype copper-azurin, with copper replaced by zinc. The remaining details of the QM/MM setup are similar to the one used for wildtype azurin as described in Section 9.2. The optimized zinc-residue distances are given in Table 9.3.1. Even though in the starting structure the zinc atom is at some 2.85 Å from Gly45, it moves towards it in the QM/MM calculation and at the end is located at a distance of only 2.17 Å from it. The zinc atom is then found 0.16 Å below the NNS plane, just like found by the X-ray data. The distances to the in-plane ligands are very well described, and are found to be slightly larger than for *wildtype* azurin. This feature is observed both in

the X-ray crystallographic studies and the QM/MM calculation for the metal-Cys112 distance, which increases by 0.063 Å (X-ray) or 0.071 Å (QM/MM) in going from copper to zinc.

TABLE 9.3.1. QM/MM OPTIMIZED DISTANCES (Å) OF ZN-AZURIN

	X-ray	*QM/MM*
Zn-Gly45 [a]	2.319 ± 0.091	2.173
Zn-His46 [a]	2.069 ± 0.050	2.085
Zn-Cys112 [a]	2.300 ± 0.022	2.240
Zn-His117 [a]	2.007 ± 0.036	2.056
Zn-Met121 [a]	3.383 ± 0.094	3.484
Zn-N2S [a]	-0.164 ± 0.042	-0.158

a) see subscript Table 9.2.1 for definition

Co/Ni substituted azurin

Recently, also the crystal structures of cobalt[73] and nickel[330] substituted azurin have become available, which showed an active site structure similar to zinc-azurin[59]. The metal atom is found in these three cases in a distorted tetrahedral coordination, where the metal atom is located below the NNS plane. The Met121 residue is found at even larger distances than in *wildtype* azurin. Some remarkable features of the cobalt and nickel azurin structures are however the distances of the metal to the in-plane ligands. Bonander et al.[73] performed a search in the Cambridge Data Base (CDB), looking for metal compounds with the metal coordinated to at least two nitrogens and a sulphur. The average metal-nitrogen distance is for both cobalt and nickel 2.02 Å, with a maximum of 2.11/2.14 Å. However, for cobalt azurin the His residues are found at 2.32 and 2.25 Å, while for nickel azurin they are located at 2.23 and 2.22 Å. The search in the CDB also gave an average nickel-oxygen distance of 2.10 Å, with a maximum of 2.16 Å, and an average cobalt-oxygen distance of 1.98 Å with a maximum of 2.16 Å.

In this case, two separate QM/MM calculations have been performed for both cobalt and nickel azurin. The QM/MM setup was the same as described before for wildtype azurin (Section 9.2). In the first try, the starting structure was taken from *wildtype* azurin and copper was replaced by either cobalt or nickel, while in a second run the starting structure was taken from the cobalt azurin crystal structure. The fully optimized active site structure is in both cases the same, whether started from the *wildtype* or the cobalt-azurin starting structure. The QM/MM optimized metal-residue distances are reported in Table 9.3.2.

Unlike the QM/MM calculations on *wildtype* and zinc azurin, the difference between the calculated and experimental metal-ligand distances is not within the experimental uncertainty anymore. The in-plane ligands are found in all cases closer to the metal atom in the QM/MM calculations than in the X-ray data. The same trend has been observed by de Kerpel et al.,[104] who performed calculations on small model systems of the active site of cobalt-azurin. Their cobalt-His distances are somewhat longer (~0.05 Å) than the ones observed here, but are still significantly smaller than the reported X-ray values[73]. Moreover, the QM/MM cobalt-His distances reported here agree very well with the average Co-N distance (2.02 Å) found in the Cambridge Data Base (CDB), as described above[73]. Since the X-ray Co-N distances are 0.12-0.28 Å longer than the *maximum* observed in the CDB, this may suggest that the cobalt-nitrogen distances observed in the crystal structure of Co-azurin[73] are too long.

TABLE 9.3.2. QM/MM OPTIMIZED DISTANCES (Å) OF CO/NI-AZURIN ACTIVE SITE

	X-ray Co-az.[73]	*QM/MM Co-az.*	*X-ray Ni-az.*[330]	*QM/MM Ni-az.*
Co/Ni-Gly45[a]	2.230 ± 0.077	2.089	2.46 ± 0.06	2.123
Co/Ni-His46[a]	2.243 ± 0.045	2.038	2.23 ± 0.09	2.059
Co/Ni-Cys112[a]	2.337 ± 0.024	2.196	2.39 ± 0.07	2.172
Co/Ni-His117[a]	2.421 ± 0.099	2.013	2.22 ± 0.12	2.055
Co/Ni-Met121[a]	3.561 ± 0.126	3.552	3.30 ± 0.05	3.209
Co/Ni-N2S[a]	-0.207 ± 0.071	-0.230	-0.18	-0.082

a) see Table 9.2.1 for definition

The distance of cobalt from the Cys112 and Gly45 residues is much smaller in the QM/MM calculations than in the X-ray data. The QM/MM calculated distance to Gly45 of 2.09 Å falls again well within the range observed in the CDB, while the X-ray distance of 2.23 Å is 0.07 Å larger than the maximum observed in CDB. Also for Cys112 a much smaller distance is observed in the QM/MM calculations than in the X-ray data. On the other hand, for Met121 a good agreement between calculated and experimental distances is found. In both cases, this residue is located at some 3.55 Å from the metal and can therefore have at most a very weak bonding with the metal atom. Another feature preserved in both QM/MM and X-ray data is the position of the metal with respect to the NNS plane. The cobalt atom is located in both cases below this plane at some 0.21/0.23 Å.

The spin density[a] of cobalt-azurin (amounting a total of three electrons due to the quartet ground state) is delocalized in the active site also. On the metal itself resides a total spin of 1.830 (61.0 %), while a total of 0.498 (16.6 %) is located on the Cys112 sulphur. This is a different situation then for *wildtype* azurin, where the spin density was mainly found on Cys112 and only roughly 20 % on copper. The remaining spin density is located on the Gly45 oxygen (0.133; 4.4 %), the His46 nitrogen (0.092; 3.1 %), the His117 nitrogen (0.121, 4.0 %) and the Met121 sulphur (0.031; 1.0 %). Therefore, even though the cobalt-sulphur distance is rather large (3.55 Å), there is still a weak covalent interaction present. This agrees well with some peaks observed in the 1D ^{1}H NMR spectrum of Co(II)-azurin, which have been assigned to protons in the methyl group of the Met121 residue. These hydrogens carry a small spin density of ~0.006 or 0.2 %. Just like the relative amount of spin density on the sulphur of Cys112 is smaller than in wildtype azurin, also the relative amount of spin density on the β-hydrogens is smaller by a factor of approximately three. Compared to wildtype azurin, a considerable increase of spin density is observed for His46-H$_\alpha$ (1.1 %), while it stays roughly the same for the other hydrogens in His46 and His117 (0.1-0.4 %).

For nickel-azurin, the same pattern is observed as for cobalt-azurin. The nickel-His distances observed in the X-ray studies are ~0.12 Å larger than the *maximum* observed in the CDB, while the QM/MM optimized nickel-His distances of ~2.05 Å agree perfectly with the average distance observed in the database. The same feature is observed for the nickel-Gly45 distance, which is found at 2.46 Å in the X-ray data and at 2.12 Å in the QM/MM calculations. Again, the distance from X-ray data is larger (0.26 Å) than the maximum observed in the database, while the QM/MM distance fits perfectly with the average Ni-O distance in the database. However, for the nickel-Cys112 distance the opposite is found. The distance from X-ray data fits very well with the average distance in the CDB, while the QM/MM optimized distance is found at almost 0.2 Å smaller than the minimum. Still, as there is only

[a] From Multipole Derived Charge analysis; see sections 3.1 and 9.2 for more details

a small difference of less than 0.03 Å with the metal-Cys distance in either wildtype or cobalt-azurin, this preserved metal-Cys112 distance may be typical for azurin. The Met121 is found in both the calculated and experimental structure at a shorter distance from nickel than from cobalt by some 0.2-0.3 Å, while the nickel atom itself is located, just like zinc and cobalt, below the NNS plane of His46, His117 and Cys112, by some 0.1/0.2 Å.

For nickel-azurin, the spin density[a], which amounts to two electrons for the triplet ground state, is found mainly on the nickel (1.01, 50.5 %), Cys112 sulphur (0.46, 23 %), Gly45 oxygen (0.111, 5.5 %), the His46 (0.072, 3.6 %) and His117 (0.094, 4.7 %) nitrogens and the Met121 sulphur (0.060, 3 %). This delocalization of spin on Met121 agrees, just like cobalt-azurin, well with some peaks in NMR spectra, which were assigned to protons in the Met121 residue.

Asn47Leu mutant

The Asn47 residue is preserved in all type 1 copper proteins for which the structure has been resolved so far, which makes it likely that it has a special function in the protein. In fact, it is found that the sidechain of the residue forms hydrogen bonds to Tyr113, which link the loops containing the copper ligands, while the backbone amide nitrogen forms a hydrogen bond towards the sulphur of Cys112. The former hydrogen bonds are presumably important for the stability and rigidity of the protein, while the hydrogen bond to the Cys112 sulphur may be involved in the peculiar spectroscopic features of the type 1 copper proteins.

The Asn47Leu mutant of *Alcaligenes denitrificans* azurin has been made in order to study the importance of this residue for the structure and function of azurin. Characterization of the mutant by UV-visible, electron spin resonance, and NMR spectroscopy revealed that the spectroscopic features are very similar to the ones of wildtype azurin. The only significantly different features are the midpoint reduction potential that changes from 286 mV (*wildtype*) to 396 mV (*N47L mutant*), and the temperature dependence of the midpoint potential, showing distinctly different entropic and enthalpic effects upon reduction. It was shown that the change in reduction potential could be reasonably well reproduced by the electrostatic interaction of the Asn47 sidechain atoms with the *extra* charge that "enters" on copper upon reduction, under the assumption that this *extra* charge resides at the copper atom. However, it has been shown in Section 6.1 that this extra charge delocalizes in the active sites, and resides on copper for only roughly 13 %.

The QM/MM optimized structure of the N47L-azurin mutant has been obtained like described previously, and results in copper-residue distances that are given in Table 9.3.3. The distances to the in-plane ligands are very similar to the *Pseudomonas aeruginosa* wildtype azurin values, but some differences are observed for the axial residues. While the distance of copper from either Met121 or the N2S-plane in the reduced state is similar to the wildtype values, the distance from Gly45 is smaller by some 0.16 Å. This difference is increased in the oxidized state, where not only the copper-Gly45 but also the copper-Met121 distance differs by some 0.3 Å from the wildtype structure. The Gly45 oxygen has moved upwards considerably (~0.2 Å), while the copper moved downwards by 0.12 Å. At the same time, the Met121 sulphur has moved away from the copper by some 0.2 Å. These changes with respect to the *wildtype* structure may have been caused by using the structure from *Alcaligenes denitrificans* azurin instead of *Pseudomonas aeruginosa* azurin; making the N47L change in the *Pa* azurin structure and optimizing it with QM/MM in the oxidized state leads to copper-ligand distances that are virtually identical to the *wildtype Pa* values, with values of 2.803 (Gly45), 2.023 (His46), 2.176 (Cys112), 2.000 (His117) and 3.002 (Met121). The only significant change in comparison with *wildtype* azurin is found

[a] From Multipole Derived Charge analysis; see sections 3.1 and 9.2 for more details.

for the axial groups with changes in the copper-ligand distance of –0.052 (Gly45) and –0.176 (Met121).

TABLE 9.3.3. QM/MM OPTIMIZED DISTANCES (Å) OF ASN47LEU AZURIN ACTIVE SITE

	X-ray[280] *reduced state*	*QM/MM reduced state*	*X-ray*[280] *oxidized state*	*QM/MM oxidized state*
Cu-Gly45 [a]	3.264 ± 0.107	2.840	3.075 ± 0.175	2.507
Cu-His46 [a]	2.047 ± 0.097	2.035	1.947 ± 0.039	1.999
Cu-Cys112 [a]	2.153 ± 0.102	2.227	2.139 ± 0.013	2.168
Cu-His117 [a]	2.211 ± 0.084	2.069	2.047 ± 0.147	1.989
Cu-Met121 [a]	3.091 ± 0.067	3.342	3.087 ± 0.047	3.479
Cu-N2S [a]	0.088 ± 0.041	0.071	0.052 ± 0.002	0.036

a) see subscript Table 9.2.1 for definition

Still, in the QM/MM optimized *Alcaligenes denitrificans* structure, the coordination of the strong in-plane ligands remains preserved, including the hydrogen bond of the amide nitrogen of Leu47 to the Cys112 sulphur (at 3.39 Å), as well as the hydrogen bond of the Phe114 amide nitrogen to the sulphur (3.91 Å). Therefore, although the axial structure gets distorted considerably, the more important in-plane ligands remain close to their wildtype positions. This would account for the almost negligible effect of the mutation on the spectroscopic features, while the difference in the axial residue positions might account for the difference in reduction potential and the entropic and enthalpic effects on these.

Asn47Asp mutant

Whereas the Asn47Leu mutant described above did not change the charge of residue 47, the Asn47Asp (N47D) mutation[4] of *Pseudomonas aeruginosa* (*Pa*) azurin introduces a negative charge close to the copper site. The influence of this introduced charge on the stability and functions of the protein has been investigated by X-ray crystallographic, spectroscopic and kinetic studies[4]. The crystal structure was obtained for the zinc mutant[4], and was found to be very similar to the one of wildtype zinc-azurin[59]. Unlike the N47L-mutant that had been purified to get rid of the zinc-substituted form, the N47D-mutant was obtained with a copper content of only 3 %. However, zinc is spectroscopically silent in the visible region of the absorption spectra as well as in EPR spectra, and because of its electronic properties does not show any reduction potential. Therefore, the spectroscopic properties of the copper-mutant could be determined. The spectroscopic and kinetic properties are similar to wildtype *Pa* azurin, while a reduction potential is found that is only 20 mV higher.

This mutant poses a challenge for the QM/MM method, as the N47D-structure for copper has not yet been determined. Therefore, the active site structure was optimized not only for the zinc N47D mutant but also for the copper-protein. The QM/MM optimized metal-residue distances are given in Table 9.3.4. Just like observed for the N47L mutant, the coordination of the in-plane ligands around the metal atom is very similar to *wildtype* azurin, thereby accounting for the negligible effect of the mutation on the spectroscopic features. The coordination of the axial residues is for the zinc-N47D structure rather similar to the wildtype zinc structure, although the metal has shifted downwards by ~0.1 Å.

TABLE 9.3.4. QM/MM OPTIMIZED DISTANCES (Å) OF ASN47ASP ZN/CU AZURIN ACTIVE SITE

	X-ray[4] *Zn-N47D*	*QM/MM Zn-N47D*	*X-ray*[7] *Cu-wildtype*	*QM/MM Cu-N47D*
Zn/Cu-Gly45 [a]	2.358 ± 0.055	2.167	2.967 ± 0.093	2.536
Zn/Cu-His46 [a]	2.096 ± 0.054	2.100	2.076 ± 0.060	2.003
Zn/Cu-Cys112 [a]	2.271 ± 0.068	2.266	2.237 ± 0.044	2.160
Zn/Cu-His117 [a]	2.047 ± 0.144	2.056	2.011 ± 0.069	2.004
Zn/Cu-Met121 [a]	3.439 ± 0.060	3.557	3.149 ± 0.070	3.307
Zn/Cu-N2S [a]	-0.136 ± 0.035	-0.264	0.083 ± 0.051	0.028

a) see subscript Table 9.2.1 for definition

The axial coordination in the copper N47D structure is, like was observed for the N47L structure, different from the wildtype structure. The distance of the Gly45 oxygen to copper is decreased by roughly 0.3 Å to 2.54 Å, while the distance of the Met121 sulphur from copper is increased by some 0.13 Å to 3.31 Å. This is accompanied by a downward move of copper itself of some 0.06 Å.

The hydrogen bonds towards the Cys112 sulphur are either longer (Asp47 backbone nitrogen) or shorter (Phe114 backbone nitrogen) by roughly 0.2 Å in this mutant. The mutation doubles the number of hydrogen bonds between the 47 and 113 residues. The backbone nitrogen of Thr113, which is hydrogen bonded to the sidechain oxygen of Asn47 in the wildtype structure, is now connected to both sidechain oxygens of Asp47. This is accomplished by a rotation of the Asp47 sidechain to reduce the distance of the second oxygen from the nitrogen. A fourth hydrogen bond is observed between the second oxygen and the backbone nitrogen of Phe114.

Phe114Ala mutant

The Phe114 residue is part of a hydrophobic patch at the northern end of azurin near the copper site and exhibits a π-electron overlap with the copper ligand His117, which has been suggested to be involved in the electron self exchange (e.s.e.) rate of the protein[6,7,95]. To study this suggested role the Phe114Ala mutant of *Pa* azurin had been made[5], which creates a hole in the immediate surroundings of the copper site. This mutation was found to have a non-negligible effect on the spectroscopic features of the protein; the wavelength of maximum absorption in the UV-vis spectra decreases by ~10 nm with an intensity that is only 78 % compared to wildtype *Pa* azurin, the hyperfine splitting parameter increases by 6 % and the reduction potential increases by ~50 mV. The latter may be not that important, as also the N47D and N47L mutants showed reduction potentials that differ from the wildtype one. The crystal structure (at 2.6 Å resolution) showed an active site structure that is similar to wildtype azurin, except for the copper-His distances which are elongated by a small amount in the case His46 (2.13 Å) and by a large amount for His117 (2.46 Å). It is uncertain whether this is due to the mutation or to the uncertainty imposed by the resolution of the crystal.

Therefore, the geometry of the active site was optimized in the presence of the protein using hybrid QM/MM calculations. The QM/MM optimized copper-residue distances are given in Table 9.3.5, which differ quite a bit from the wildtype values.

TABLE 9.3.5. QM/MM OPTIMIZED DISTANCES (Å) OF PHE114ALA AZURIN ACTIVE SITE

	X-ray reduced state	*QM/MM reduced state*	*X-ray*[5] *oxidized state*	*QM/MM oxidized state*
Cu-Gly45 [a]	-	2.867	2.983 ± 0.138	2.555
Cu-His46 [a]	-	2.075	2.127 ± 0.071	2.037
Cu-Cys112 [a]	-	2.319	2.228 ± 0.012	2.185
Cu-His117 [a]	-	2.134	2.463 ± 0.128	2.003
Cu-Met121 [a]	-	3.061	3.022 ± 0.121	3.506
Cu-N2S [a]	-	0.173	0.431 ± 0.289	-0.022

a) see subscript Table 9.2.1 for definition

The in-plane coordination is rather similar, but just like found for the N47L and N47D mutants, the Gly45 oxygen moves closer to copper by some 0.3 Å. At the same time, the distance of the Met121 sulphur from the copper is increased by almost the same amount. A rather peculiar conformation for the Met121 residue is observed in this mutant. The distance of the Cβ of Met121 from the Cβ of Cys112 (normally found between 3.65 and 3.8 Å) decreases to 3.41 Å and is accompanied by a $C\gamma_{121}$-$C\beta_{112}$ distance of only 3.98 Å (normally 4.45-4.55 Å). This is not observed in either the N47D or the N47L mutants, which showed normal values for these distances.

The hydrogen bond from the backbone nitrogen of residue 114 to the Cys112 sulphur is found at a shorter distance of 3.48 Å, while the Asn47 backbone nitrogen to Cys112 sulphur hydrogen bond is elongated at 3.67 Å. This feature was also present in the N47D mutant and can therefore not be responsible for the change in spectroscopic features.

Met121Gln mutant

The Met121Gln mutant has received some special attention[36-38,42,86,90,103], as it has been used as model for stellacyanin, a type 1 copper protein for which the structure has been elucidated only recently[96]. It had been proposed that the fourth ligand in stellacyanin might be glutamine, and the spectroscopic features of the M121Q azurin mutant[42] were indeed very similar to the ones observed in stellacyanin. The crystal structure of the oxidized M121Q mutant shows a tetrahedral coordination around copper, with the usual three in-plane ligands (His46, Cys112, and His117) and Gln121 as the fourth ligand at 2.26 Å. The fifth coordinating residue that is normally found in azurin (Gly45) is located at 3.37 Å in the crystal structure at a weakly bonding position. Upon reduction, a peculiar crystal structure is observed; the distances of copper to Gly45, His46 and Cys112 remain rather the same, but the Gln121 and His117 residues move away from copper to 2.73 and 2.67 Å respectively. Also the angles around the copper change dramatically, thereby changing the coordination from approximately tetrahedral to almost linear with two strong ligands (His46 and Cys112).

The active site structure of both the reduced and oxidized form of the M121Q-mutant of *Alcaligenes denitrificans* (*Ad*) azurin has been optimized in the presence of the protein and a layer of solvent molecules surrounding the protein, as described for wildtype *Pa* azurin (Section 9.2). The QM/MM optimized copper-residue distances are given in Table 9.3.6.

TABLE 9.3.6. QM/MM OPTIMIZED DISTANCES (Å) OF MET121GLN AZURIN ACTIVE SITE

	X-ray[42] *reduced state*	*QM/MM reduced [p/dp]*	*X-ray*[42] *oxidized state*	*QM/MM oxidized state*
Cu-Gly45 [a]	3.332 ± 0.089	3.002 / 2.862	3.372 ± 0.136	3.341
Cu-His46 [a]	1.975 ± 0.062	1.885 / 2.004	1.936 ± 0.041	1.950
Cu-Cys112 [a]	2.089 ± 0.030	2.114 / 2.221	2.121 ± 0.019	2.148
Cu-His117 [a]	2.675 ± 0.078	3.086 / 2.074	2.046 ± 0.025	2.022
Cu-Gln121 [a]	2.726 ± 0.051	3.086 / 3.327	2.263 ± 0.017	2.300
Cu-N2S [a]	0.200 ± 0.004	0.039 / -0.057	0.260 ± 0.007	0.270

p) His117 protonated at δ position

dp) His117 deprotonated at δ position

a) see Table 9.2.1 for definition

At first, the geometry of the reduced protein was optimized using the normal protonation states of the residues in the site, e.g. both His46 and His117 were protonated at the ε position and bonded to copper at the δ position (*dp* entries in Table 9.3.6). However, this results in an active site structure similar to the reduced wildtype structure, with a distance of the copper to His117 that is more than 0.6 Å smaller than in the crystal structure. This deviation is too large to be accounted for by the experimental uncertainty, especially since the resolution of the crystal structure is 1.94 Å. Therefore, some other effect must be responsible for this peculiar structure. As a second attempt it was checked whether His117 could have been protonated at both nitrogen positions. The results are also given in Table 9.3.6 (*p* entries), which give a somewhat better agreement with the crystal structure regarding the distances of copper to Cys112 and His46, but not for the distances to His117 and Gln121 that are 0.41 and 0.36 Å too long respectively. Given the resolution of 1.94 Å, a superposition of the *p* and *dp* results is also ruled out. Finally, it has been checked whether the His117 residue might have been rotated as to orient with the ε nitrogen towards the copper. Both the deprotonated and protonated nitrogen structure (now at the ε position) has been optimized, which resulted for the protonated case in an almost identical structure as the one where both nitrogens were protonated and the copper was connected to the δ nitrogen. The structure of the deprotonated ε-nitrogen was found to give even longer copper-His117 distances. As these four QM/MM calculations all resulted in structures that disagree to a large extent with the crystal structure, it seems that the crystal structure does not correspond to either one of the systems optimized by the QM/MM calculations.

The oxidized protein structure is very well reproduced by the QM/MM calculations. The difference between the crystal and QM/MM optimized structure is in all cases very small.

Co/Ni substituted Met121Gln mutant

Just like for wildtype azurin, also for the M121Q mutant has copper been replaced by cobalt and nickel to study the spectroscopic features of the mutant with paramagnetic proton NMR[36]. However, it is shown earlier in this section that the metals may coordinate in different manners. The extent with which they do in the M121Q mutant has also been studied using hybrid QM/MM geometry optimizations of the active site in presence of the protein. The optimized copper-residue distances are given in Table 9.3.7, where for reference also the QM/MM optimized and crystal structure data of the copper M121Q protein are given.

TABLE 9.3.7. QM/MM OPTIMIZED DISTANCES (Å) OF CO/NI MET121GLN AZURIN ACTIVE SITE

	X-ray[42] *Cu*	*QM/MM Cu*	*QM/MM Co*	*QM/MM Ni*
Co/Ni-Gly45 [a]	3.372 ± 0.136	3.341	2.792	3.079
Co/Ni-His46 [a]	1.936 ± 0.041	1.950	2.015	2.002
Co/Ni-Cys112 [a]	2.121 ± 0.019	2.148	2.189	2.158
Co/Ni-His117 [a]	2.046 ± 0.025	2.022	1.998	2.009
Co/Ni-Gln121 [a]	2.263 ± 0.017	2.300	2.158	2.068
Co/Ni-N2S [a]	0.260 ± 0.007	0.270	0.247	0.258

a) see subscript Table 9.2.1 for definition

For both cobalt and nickel M121Q azurin, the position of the metal relative to the N2S plane is reversed relative to wildtype cobalt/nickel substituted (*Co-Ni wt*) azurin; this is due to the axial bonding of the metals that takes place in M121Q azurin for both metals with Gln121 instead of Gly45, as observed in *Co-Ni wt* azurin. The distances to the in-plane ligands remain almost the same as in *Co-Ni wt* azurin, while the bond to oxygen is shorter than in *Co-Ni wt* azurin for nickel and longer for cobalt by some 0.05-0.07 Å. The distance to the other axial ligand, in this case Gly45, is shorter than in *Co-Ni wt* azurin for both metals by 0.76 Å (Co) and 0.13 Å (Ni) respectively.

Compared with the copper-M121Q structure a larger similarity is observed between the three metals than for *Co-Ni wt* azurin. All three metal sites show more or less a tetrahedral coordination, with only a weakly bonding interaction with Gly45.

Conclusions

The effect of metal substitution and and/or residue mutations can be very well reproduced by hybrid QM/MM geometry optimizations, in which the active site structure is optimized in the presence of the protein and solvent molecules. Even when starting the QM/MM calculations with structures that are not biased for the metal in the site, the metal atom moves away towards its favored position.

The zinc/cobalt/nickel substituted azurin structures show a tetrahedral coordination around the metal, as shown also by the crystal structures. In these metal sites, the metal atom moves below the N2S-plane of the strong in-plane ligands (His46, Cys112 and His117), thereby getting closer to the Gly45 oxygen. The latter residue moves upwards itself, leading to a metal-oxygen distance in a range of 2.09-2.17 Å for the three metals. The distance of the His46, His117 and Gly45 residues from either cobalt or nickel is shorter than observed in the crystal structure, at distances that are normal for these kind of bonds.

The function of the Asn47 residue, which is conserved in all type 1 copper proteins for which the structure is known, had been studied by site directed mutagenesis by mutating it to either Leu (N47L) or Asp (N47D). For both these mutants, the active site structure has been optimized with QM/MM calculations; for the N47D mutant, not only the copper-protein has been used but also the zinc-structure, as the crystal structure of it was obtained for the latter. In the N47L mutant, two hydrogen bonds connecting residue 47 with Thr113 are removed, which may have impact on the stability of the protein. However, the hydrogen bond of the backbone nitrogen towards the Cys112 sulphur is preserved in both the N47L and the N47D mutant. In the latter, instead of removing two hydrogen

bonds, two are added between the residues 47 and 113. This is brought about by a rotation of the Asp47 sidechain to facilitate these two extra hydrogen bonds. In both cases, the Gly45 oxygen moves closer towards the copper at a distance of ~2.5 Å. This may have an effect on the reduction potential, but as the coordination of the strong in-plane ligands and the hydrogen bonds towards the Cys112 sulphur are preserved, the spectroscopic features hardly change.

The Met121Gln mutant can in the oxidized state be very well described by the QM/MM calculations; the differences between the QM/MM optimized and crystal structure are small. In the reduced state however, the crystal structure can not be reproduced by the QM/MM calculations. In this peculiar crystal structure, the His117 residue is found at some 2.7 Å from the copper. Neither the normal situation, with His117 connected to copper with its *deprotonated* Nδ atom, nor an unusual one, with His117 connected to copper with the *protonated* Nδ atom, provides an active site similar to the crystal structure. In the first, the His117 residue is found at its normal position at some 2.05 Å from the copper, while in the second it is found at a too large distance of 3.1 Å. It has also been checked whether the residue might have rotated 180°, thereby connecting to the copper with its Nε atom, but this resulted in a structure similar to the previous *protonated* His117 one, therefore different from the crystal structure. Therefore, it seems that the crystal structure does not correspond to either one of the systems optimized by the QM/MM calculations.

9.4 *H117G/N42C azurin*

An engineered double mutant exhibiting unexpected spectroscopic features[a]

The His117Gly mutant[75,319,324,325] creates a hole in the protein surface through which exogenous ligands can enter the active site and bind to the copper. Depending on the kind of ligand, different types of copper sites are obtained; binding of chloride, bromide, imidazole (and derivatives) or pyridine (and derivatives) restores the characteristic spectroscopic features of a type 1 copper site[319], whereas the binding of for instance water or histidine creates novel azurin copper sites, which are more similar to "normal" (or type 2) copper sites. For instance, without external ligands, His117Gly azurin is green with absorption maxima at 420 and 628 nm, but addition of imidazole restores the characteristic blue color of wildtype azurin.

In an attempt to construct dimers of *Pseudomonas aeruginosa* azurin, the double mutant H117G/N42C has been made[61] that would enable a dimer by the formation of a sulphur bond between both Cys42's, which has been successfully applied in the N42C single mutant[327] and was shown to be catalyzed by the presence of copper. Adding Cu(II) to the colorless monomeric form of apo-H117G/N42C results in the immediate formation of a yellow color, with a strong absorption at 385 nm[61]. The addition of external ligands like imidazole or chloride does not influence the spectroscopic features, therefore the open binding site on copper due to the His117Gly must be occupied by another ligand. The most likely candidate for this new ligand is then Cys42, which was proven by several experiments; for instance, adding Cu(II) to the apo-H117G/N42C dimer results in the expected green color of the H117G mutant, while the blocking of Cys42 by iodoacetamide also restores the characteristic properties of the H117G mutant.

The absorption at 385 nm is characteristic of a type-2 copper site and can be assigned as a cysteine sulphur to copper charge transfer band. In the EPR spectrum of the double mutant two species can be distinguished, both with tetragonal type-2 copper site characteristics, which can be attributed to intermediates that are formed during the uptake of copper. After three days, only the most stable species remains with typical type-2 site characteristics: a g_z value of 2.23 with a hyperfine splitting of 0.0160 cm^{-1}.

Molecular Dynamics simulations

As it has been suggested that the Cys42 residue is picking up the copper from the solution and bringing it into the active site, the same process has been investigated with Molecular Dynamics simulations. At the beginning the copper was bonded to Cys42 and then brought into the active site by applying distance restraints in six steps. Inbetween any of these steps, the protein was allowed to react to the new situation and relax. As the experimental results indicated a tetragonal coordination around copper, in the last step an improper dihedral potential has been added to ensure a tetragonal coordination. At the end of the process, the copper was present in the active site at 2.35 Å from His46, 2.45 Å from Cys112, 2.48 Å from Cys42 and 2.81 Å from Met121. These copper residue distances show a reasonable amount of strain on the residues, which might have been build up in the process of bringing the copper into the active site.

[a] Joint study by M. van den Bosch (MD simulations) and M. Swart (QMMM optimizations)

QM/MM geometry optimization

Starting from the structure obtained in the MD simulation, the active site geometry has been optimized in the presence of the protein and a surrounding layer of water molecules using hybrid QM/MM calculations, which were setup as described in Section 9.2. The optimized copper-residue distances are given in Table 9.4.1.

TABLE 9.4.1. DISTANCES (Å) IN H117G/N42C-AZURIN ACTIVE SITE

	Molecular Dynamics	*QM/MM*	*QM/MM (+H2O)*
Cu-Cys42	2.48	2.29	2.31
Cu-Gly45	5.38	5.32	4.99
Cu-His46	2.35	2.04	2.05
Cu-Cys112	2.45	2.18	2.21
Cu-Met121	2.81	4.01	3.91
Cu-H2O	-	-	2.89

The strain in the copper site has been removed, as indicated by the copper-His46 and copper-Cys112 distances that are at normal values again, while also Cys42 has moved closer to the copper at 2.29 Å. The Met121 on the other hand has completely moved away from the copper site and is not involved as the proposed fourth tetragonal ligand, nor does any other ligand move closer to take its place.

A possible fourth ligand for the copper in this site might be a water molecule that could have stayed connected to copper during the picking up and subsequently bringing in into the active site of copper, or it may already been present in the active site. Therefore, a water molecule was placed in the active site close to the space occupied by Met121 before it moved away, at roughly 2.0 Å from the copper, which is a normal distance for a copper-oxygen bond. In the QM/MM geometry optimization however, the water moves away from the copper just as Met121 did, and is located at 2.9 Å in the end. The ligands remain approximately at the same position, while Gly45 and Met121 move a little closer to the copper again. Still, the coordination is far from tetragonal and therefore also the water molecule can not be the fourth ligand.

A confirmation that the strain in the protein has been removed in the QM/MM optimization is given by the fact that the protein remained stable in a Molecular Dynamics simulation, that was performed on the QM/MM optimized structure without a water molecule in the active site, and resulted in normal copper residue distances again. As this strain was not released in the intermediate relaxation parts of the initial MD simulations, it indicates that the bringing in of the copper may have been done too fast and with too hard pulling. Currently, this process is reinvestigated with MD simulations where the pulling in of the copper is performed at a lower speed and with less brute force.

Computation of EPR parameters

The g-tensor and hyperfine splitting for the active site obtained in the QM/MM optimization were computed using the experience obtained in *Chapter* 7 (copper nuclear charge of 28.6, in a DZV basis set). The g-tensor values of the double mutant without a water molecule in the active site are 2.025, 2.030, 2.132 with a hyperfine splitting (A_Z) of –398 MHz, while they are 2.027, 2.030, 2.121 with a hyperfine splitting (A_Z) of –360 with a water molecule in the site. These are all considerably lower than the experimental values of 2.23 (g_Z) and -480 MHz (A_Z).[61]

Conclusions

The engineered N42C/H117G double mutant of azurin, which has some remarkable spectroscopic features, has been modeled using a combination of Molecular Dynamics simulations and QM/MM geometry optimizations.

In the simulations, the copper has been attached to Cys42 and pulled inside the active site using distance restraints. Subsequently, the active site structure has been optimized in the presence of the protein, where the presumed four copper ligands (Cys42, His46, Cys112 and Met121) have been put in the QM system. However, in the QM/MM optimization the Met121 immediately drifts away from the copper and ends up at nearly 4 Å, therefore at a non-bonding position.

As the EPR spectra obtained experimentally for the protein seem to indicate that the protein contains a type 2 copper site with four ligands, a water molecule has explicitly been included in the QM system, approximately at the position left open by the Met121 residue, and a new QM/MM geometry optimization performed. However, also the water molecule moves away from the copper and can not be the purported fourth ligand.

The computed g-tensor values of the optimized active site structures are too small in comparison with the experimental values, thereby indicating that the optimized structures, with copper effectively bonded to three ligands, can not be responsible for the spectroscopic features.