Contents

Chapter 1

Glutathione S-Transferase: First-Principles Study of the Role of the Active Site

1.1 Introduction

Ab-initio study of molecules of biological relevance is a rapidly growing field since the accuracy and efficiency reached by electronic structure calculations allow to tackle such complex systems. Understanding the general mechanisms of biological reactivity involves the use of numerical techniques able to deal with large systems where substantial charge transfer and rearrangement among atoms takes place along the reaction pathway.

In the present work we report the study of glutathione (GSH), an ubiquitous tripeptide (γ-Glu-Cys-Gly) found in eukariotic cells where it is implicated in many cellular functions that protect cells against toxicity and stress induced by chemical agents. These chemicals are deactivated by attaching the SH group of glutathione to the hydrophilic moiety of the toxifying agent, thus preventing the chemical reactivity. GSH is a substrate of a variety of enzymes such as the glutathione S-transferases (GST), that constitute a very important class of proteins that catalyze the process of deactivation of the reactive agents.

Crystallographic and site-directed mutagenesis studies have shown that a number of conserved aminoacid residues of the GST enzyme are involved

in the binding process [1]. In particular Tyr^7, the tyrosine residue close to the N-terminal end of the enzyme, is conserved in all known mammalian cytosolic glutathione transferases and it is considered to play a crucial role in catalysis. Before binding to the chemical agent, the GSH tripeptide is partially embedded in the glutathione S-transferase, and with the OH group of the tyrosine residue of GST facing the SH group of GSH.

Site directed mutagenesis experiments have revealed that, by mutating the tyrosine residue into a phenylalanine (Phe^7) aminoacid (an aminoacid almost identical to Tyr^7, except for an hydrogen atom saturating the carbon of the aromatic ring in lieu of the OH group of tyrosine) the activity of the enzyme is sensibly lowered but it does not disappear completely [2].

In particular it has been found that the pK of GSH is 9.0 for GSH in solution, while in the complex with the native complex (GST(Tyr)) it is 6.2 and for the Phe^7 mutant (GST(Phe)) the pK of GSH is equal to 7.8, i.e. the GSH group has a much larger tendency to deprotonate when bound to the native (not mutated) GST protein.

These data have been interpreted in terms of a hydrogen bond formed between Tyr^7 and GSH, that lowers the pK of GSH in the binary complex and represents the leading factor in the activation of the sulfur atom to chemical bonding with external groups. Such a low pK for the GSH-enzyme complex suggests that at physiological pH glutathione shows a deprotonated S atom in the native complex.

The spatial arrangement of the protons in the GSH-GST complex is an important open question. Understanding the protons location can help elucidating the role of the Tyr^7 group and the environment around the sulfur atom in the enzymatic reaction. Three different hypothesis have been proposed for the active site configuration once that one of the two hydrogen atoms of the active site (one belonging to GSH and one to Tyr^7) has been removed. The first suggestion is that Tyr^7 retains its proton on the OH group and acts as a H-bond donor for glutathione, GS$\cdots$HO-Tyr, thus resulting in a stabilization of the S^- anion. In this configuration the Tyr^7 group resembles the effect of a water molecule in lowering the deprotonation barrier. This configuration is consistent with the X-ray structure although crystallographic resolution is not sufficient to draw a final conclusion about the proton location. On the other hand the Tyr^7 residue may lose its proton and act as a H-bond acceptor, GSH$\cdots$O-Tyr, a configuration that assists the proton extraction from the thiol group but it does not produce a substantial negative charge surplus

on the S atom. Finally a possible proton arrangement is proposed to be a mid-way hydrogen bond, GS$\cdots$H$\cdots$O-Tyr. The first two suggestions consider the H atom moving in a double well H-bond with a large S-O distance, while for the third one the H atom sits at the center of a single well potential and a compressed S-O distance [3, 4]. Each one of the three possible configurations results in a different charge arrangement in the enzyme active site, and in particular on the sulfur atom. Thus for each of these cases interaction and positioning of the S atom in the active site electrostatic environment, e.g. due to the electric dipole field generated by the helix α_1 of the enzyme, plays a different role in the deprotonation and enzymatic process.

A previous electronic structure investigation on the complex has considered a reduced representation of GSH-GST active site at a MP2/MP4 level of accuracy [5]. In this previous work GSH has been represented by a CH_3SH molecule while the Tyr^7 residue has been modeled more completely by a phenol $(C_6H_5)OH$ group. On the basis of this reduced model the proton arrangement has been studied and it was concluded that the strong H-bond formed between the two groups exhibits the $GS\cdots HO$-Tyr configuration as the privileged one. Moreover an interesting result of this work is that when substituting the $(C_6H_5)OH$ molecule with a benzene ring, i.e. modeling the Phe^7 mutant, the experimental difference in pK between the two Tyr^7 and Phe^7 species is not recovered. This discrepancy is speculated to be due to a water molecule stuck between the S atoms and Phe^7 ring, i.e. inside the volume once occupied by the OH group of Tyr^7. The water molecule should act as the acceptor of the thiol hydrogen so that it lowers the pK of glutathione but still exhibiting a deprotonated active site at pH 7.

In this work we address the problem of whether or not a water molecule plays a central role in the modeling of the active site catalitic activity, by high level DFT calculations.

The reduced model of the GSH-GST complex used in ref. [5] seems unsatisfactory from many points of view. At first the limited size of the system does not allow charge transfer due to atoms far enough from the active site once the proton is extracted, as it may occur in the full GSH-GST system. Secondly, the conformation of the complete GSH molecule can play an important role on the energetics of the proton extraction. The size of the considered system does not allow a proper positioning of the groups and subsequent ionic optimization to the energy minimum so to faithfully reproduce the experimental system. Finally computed energy (and consequently pK)

differences do not include the effect of the interaction between the electrostatic environment and the active site atoms, before and after the extraction of the proton.

In order to analyze the importance of these effects, we perform density functional calculations for different representation of the system. On one hand, we consider the same reduced model as in ref. [5]. In this case, we compare the results obtained using a gaussian basis set to those obtained with a planewave basis set. On the other hand, we consider a large scale model including a complete representation of the GSH molecule and five aminoacids of the embedding GST enzyme (class π of the pig lung glutathione S-transferase [6]), studying both the cases of the Tyr^7 and Phe^7 mutants.

For both the reduced and the large models, we analyze both the case of a system in absence of water and with a single water molecule present in the active site region. Indeed, from inspection of the crystallographic data, it is apparent that GSH is partially exposed to water whereas the thiol is the part of the peptide that is most screened to direct contact with solvent, and thus it is conceivable that a water molecule can be stuck in the proximity of the active site.

1.2 Technical Details

1.2.1 Structural Models

Our reduced model (R.M.) systems are the same as in ref. [5]; GSH is represented by a CH_3SH molecule, while the Tyr^7 residue is modeled by a phenol $(C_6H_5)OH$ and the non interacting Phe^7 has not been inserted in our R.M.; we finally considered the adduct formed by CH_3SH with a water molecule.
Thus, in our reduced model representation, 6 different molecules have been considered, labeled as follows:
CH_3SH (**1**), CH_3S- (**2**),
CH_3SH $(C_6H_5)OH$ (**3**), CH_3S- $(C_6H_5)OH$ (**4**),
CH_3SH H_2O (**5**) and CH_3S- H_2O (**6**).
Our large scale model (L.M.) consists of a complete representation of the GSH molecule and includes five blocking aminoacids Tyr^7, Arg^{13}, Trp^{38}, Gln^{49} and Gln^{62} which are modeled by their side chain residues. The groups embedding

the GSH group are free to move except for the C_α atoms of each molecule that have fixed positions in space in order to inhibit uncontrolled drifting in the simulation cell. By imposing this constraint on the C_α we allow each residue to slightly roto-translate in order to accomodate the GSH molecule but still the presence of the enzyme backbone, the less flexible part of a protein [7], is considered. Vice versa the GSH molecule has no constraints on the atomic positions. Its configuration is taken from the X-ray structure of the native complex where the close packing arrangement of atoms prevents the water molecules to approach the active site.

For both models, the whole system is first treated in vacuo conditions. Thus the investigation is about the GSH-GST complex without considering water molecules present around the active site. When treating the Phe^7 mutant we also investigate the possibility of a water molecule interfering directly with the active site.

All the systems that have been investigated in this study have been listed in Table 1.1.

It has to be noticed that the largest system we considered is formed by 131 atoms and for large models of the GST enzyme an average number of 125 atoms has been considered; in this case traditional DFT-gaussian based schemes would have not allowed to tackle these larger system at a sufficient level of accuracy. Indeed, for the largest system, the number of electrons is 468, and adopting a 6-31g* basis set, which may be considered a sort of a standard starting point, a total number of 1988 functions contracted to 1046 basis functions, should be inserted in the calculation. Actually, in order to handle such a large number of basis functions, one should run the correspondent calculations on a supercomputer for several days; this however would result in a global low accuracy in the computation of hydrogen bonding properties, due to the lack of diffues functions in the basis set. When we want to include diffuse functions, i.e. considering a 6-31+g* basis set, a number of 2228 primitive functions, contracted to 1286 basis functions should be considered while using a 6-311+g(d,p) basis set, 2682 primitive functions, contracted to 1754 basis functions should be used, which probably represents an upper limit even for nowadays parallel machines.

In this case the Car-Parrinello method has provided a viable framework to successfully tackle a problem of this dimensions. Indeed the investigated systems represent quite a good case for the CP method, because no heavy elements are present, thus no Vanderbilt pseudopotentials have been used,

and the cavity of the GSH protein shows a compact structure, which allows to reduce the number of plane waves used in the calculations of the electronic structures and geometries of the large model representations of the protein, given that the number of plane waves, at a fixed kinetic energy cutoff, is a function of the volume of the simulation cell.

1.2.2 Theoretical Method

- **Gaussian basis sets calculations**

All the calculations with gaussian basis sets reported in this work have been performed using the Gaussian98 [8] program package. We used two basis sets for our calculations on the R.M., in order to test whether basis set effects were important in the description of hydrogen bond energetics. The two basis sets are both originated from the original 6-311G basis set [9] available in Gaussian98: the first one, hereafter indicated as **A**, is a 6-311+G (d,p), while the second one, hereafter indicated as **B**, is a much larger 6-311++G (3df,2pd). We decided to use such a large amount of diffuse and polarization functions because of the general difficulty encountered in describing anions with gaussian basis sets. Moreover we decided to use two different exchange-correlation functionals in order to test whether differences in the XC functional resulted in substantially different hydrogen bond properties. We used the BPW91 [13] functional, the same one used for the Car-Parrinello calculations, in order to compare results among different (gaussians versus plane waves) basis sets, and the hybrid B3LYP [14] functional including exact exchange in order to check whether substantial differences appear in the energetics. Geometry optimizations were perfomed on **1**, **2**, **3**, **4**, **5** and **6**, without any simmetry constraints using the default Berny [15] algorithm available in Gaussian98, with both functionals using basis sets **A** and **B**. Moreover we optimized the geometry of the phenol molecule, in order to compute interaction energies for the hydrogen bond in the species **3** and **4**. Frequency calculations have been performed on the six considered molecules and on the phenol molecule, using the BPW91 functional in conjunction with basis set **A**, in order to include thermal contributions to the definition of the energetics of the considered deprotonation reactions.
In addition to that, Oniom [16] calculations have been performed on the large model, in order to compare results obtained by a full *ab initio* treatment with

those obtained by combination of Quantum Mechanics (QM) and Molecular Mechanics (MM). Thus the smaller systems 1, 2, 5 and 6 have been embedded in the classical enviroment of the 5 amynoacids, and their geometry have been fully optimized adopting basis set **A**, with both XC functionals, for the quantum part, and the with the UFF [10] force field for the classical part, adopting default cutoffs for the cut beween the quantum and classical parts. Actually embedding was performed in a very natural way, by attaching one of the hydrogens of the methyl group of the quantum systems 1, 2, 5 and 6, in the position the α carbon of the main side chain of Cys, included in the classical part. This particular embedding was adopted because the structure of the quantum part, when the α carbon is substituted by a hydrogen, is exactely the same as in the reduced model,so that direct comparisons between the two cases can be made, and eventual differences may be discussed in terms of environment effects of the classical on the quantum surrounding system.

- **Plane waves basis set calculations**

In our planewave calculations the atomic coordinates are optimized to minimize the total energy using the Car-Parrinello method [11], which provides the electronic structure as well as the forces that act on the ions. Only valence electrons are explicitly considered by means of pseudopotentials to account for the core-valence interactions [12]. We adopte periodic boundary conditions, keeping a minimum of 5Å between repeated images. This condition is sufficient to ensure negligible interaction between the images. The electronic wave functions at Γ-point of the supercell Brillouin zone and the electron density are expanded on a plane-wave basis set with kinetic energy cutoffs of 25 and 96 Ha, respectively. The exchange and correlation energy are evaluated in the generalized-gradient approximation of Perdew and Wang [13].

1.3 Results and Discussion

1.3.1 Reduced Model

The energy differences for the deprotonation reactions, computed with gaussian and PW basis sets for the BPW91 functional and with gaussian basis

only for the B3LYP functional, are collected in Table 1.2. It is interesting to compare these results among them and, for molecules **3** and **4**, with those taken from ref. [5] in which energy differences and geometries were obtained at the MP2 and Hartree-Fock level of theory, respectevely. Moreover crossed comparisons between geometrical parameters and energies computed with gaussians basis sets and plane wave basis will be performed, in order to highline eventual differences. From Table 1.2 it can be noticed that energy differences computed with the two functionals with both gaussian basis sets show negligible (i.e. less than 1 Kcal/mol) variations. Thus we adopted the less computationally expensive BPW91 functional for all the calculations even if we will briefly comment data refering to the B3LYP functional too. We found energy differences computed with basis set **B**, always slightly higher than the correspondent quantities computed with basis set **A**, even if the difference is quite small (less than 2 Kcal/mol). However we found this effect systematically in all the reported calculations and so we tried to rationalize it in terms of an over-stabilization of the protonated molecule with respect to the anion, probably due to the large number of polarization functions. Comparing our results with those taken from ref. [5] we observe that the difference in the value of the computed energy differences for the couples **1**-**2** and **4**-**6**, between our calculations and the MP2/6-31G* value taken from ref. [5], result always lower in our case, with differences of 5.0 and 6.2 Kcal/mol, respectevely. Such differences are reasonable if we consider that our results are obtained with a larger basis set which includes diffuse functions, and indicate that higher stabilization of the anions is achieved in our calculations. The deprotonation energies [17] computed using a plane-wave basis set agree very well with those obtained with a gaussian basis set (see Table 1.2). Indeed, for basis set **A**, we find an average variation of 0.0079 a.u. (1.6 Kcal/mol) and 0.0092 a.u. (1.9 Kcal/mol) for the three considered energy differe nces computed with BPW91 and B3LYP functionals, respectevely (see Table 1.2). Moreover, we notice that the deprotonation energy for the isolated CH_3SH mole cule and with a C_6H_6 differ only by 0.001 a.u. at the Car-Parrinello level of theory, showing that the Phe^7 group does not interact with the sulfur atom of the GSH molecule. However we studied protein environment in this case also, in order to check whet her effects of the enviroment on the energy differences were active in the case of non interacting species too.

In Figures 1, 2 and 3 structures and main geometrical parameters computed

at the BPW91-**A** for molecules **1** and **2**, **3** and **4** and **5** and **6**, respecetvely, have been reported. In Table 1.3, selected geometrical parameters computed with the BPW91 functional both for gaussian and PW basis sets, and for the B3LYP functional with the two gaussian basis sets, for charged molecules **2**, **4** and **6**, are reported. The correspondent geometrical parameters for the neutral molecules **1**, **3** and **5**, are reported in 1.4. It can be noticed that computed geometrical parameters for the anions **2**, **4**, **6** show quite few variations when, varying the gaussian basis set, within the same functional. When analyzing the variations in the geometrical parameters for the neutral species **1**, **3** and **5**, we find, instead, slightly higher variations on going from basis set **A** to **B**. Indeed an average variation of 0.008 $\mathring{A}$ is observed for bond distances in this case, versus an average variation of 0.002 $\mathring{A}$ in the case of the anions, which corresponds to higher basis set dependence for the description of the neutral molecules than for the correspondent anions, reflecting what we have already observed for the energetics of the deprotonation processes. The agreement on geometrical parameters computed with BPW91 using gaussian and PW basis set is excellent, considering that pseudo-potentials are used for PW calculations. Only the carbon-sulfur and the sulfur-hydrogen (not H-bound) distances are always computed to be slightly longer with the P-W basis set than with gaussian basis set **A**, even if the average variations throughout the considered systems are 0.016 and 0.014 $\mathring{A}$, respectevely.
From the computed energy differences and geometrical parameters, we can rationalize the relative strenght of hydrogen bonds in the considered systems. In Table 1.5 computed interaction enthalpies and free energies for **3**, **4**, **5** and **6**, are reported. We will limit our discussion at the BPW91-**A** level of theory, due to the fact that similar arguments hold in the case of PW and **B** basis sets. First of all we can notice that, obviously, the hydrogen bond is much stronger in the deprotonated molecules. Indeed, interaction free energies, are computed to be positive for the neutral species **3** and **5**, which do not result stable, while are computed to be negative for the anions **4** and **6**. Moreover the hydrogen bond in **4** is computed to very strong, at this level of theory, with an interaction free energy more than twice that the one computed for **6**. Geometrical parameters reflect the same scale for hydrogen bonding strength, as indicated by the shorter SH bond distance, the smaller value of the $\angle$O-H-S angle, and the longer OH bond distance, computed for molecules **4** and **6** with respect to **3** and **5**.
In order to understand whether or not localization of the negative charges is

responsible for the observed hydrogen bonding strenght scale, we performed Mulliken charge analysis on the six considere molecules. In paricular it is interesting to compare, for each couple of hydrogen bonding partners, the localization of the atoms and groups, where negative charge if found after deprotonation. In Table 1.6 Mulliken analysis results have been collected for the six considered molecules. Results have been grouped in terms of total charge borne by CH_3, SH or S, OH, Phe or H groups, depending upon which system is considered. As it can be seen, for molecules **1**-**2** deprotonation leads to localization of the negative charge almost on the S and C atoms, with a variation of charge, from the neutral to the charged system, computed to be -0.78 and -0.22 electrons, respectevely. When analysing data refering to the partners **3**-**4**, we see that the variation of charge borne by the S atom is smaller than in the non interacting molecules **1**-**2**, resulting in a diference of -0.47 electrons; the C atom shows a variation of about -0.20 electrons, not significantely different from the correspondent value computed for the couple **1**-**2**, while the value of about -0.20 electrons computed for the phenol ring reflects the charge delocalization effect discussed previously. Indeed, when we compare these data with Mulliken charges computed for partners **5**-**6**, we notice again that charge is amolst borne by the S and C atoms alone. We can therefore rationalize the computed hydrogen bonding strengths, in terms of partial delocalization of the incoming negative charge, originated from hydrogen bond on the oxigen atom, to the aromatic phenol ring, which stabilizes the correspondent anion **4**; this is of course not possible in the case in which a water molecule interacts with the thiol group, resulting in a much lower hydrogen bonding energy.

1.3.2 Large Scale Model

For the large scale model, preliminary calculations on the GSH molecule alone reveal that the ionic positions of the molecule in its local minimum energy configuration is not very different from the starting crystal structure of the embedded GSH. The rms difference in the positions of the heavy atoms of optimized GSH with respect to the crystallographic structure is 0.135 Å, where the largest difference is found at the level of the Glu terminal nitrogen atom (0.338 Å).

When studying the GSH-GST(Tyr) system, relaxation of the ionic positions considering both the OH group of Tyr and the protonated sulfur atom

of GSH we obtain a Kohn-Sham total energy of -518.0475 Ha. The position of H attached to S does not point towards any negatively charged atoms present in the neighborhood of the active site (e.g. the carbonylic oxygens of Gly or Cys of the GSH molecule), indicating that proton extraction is not assisted by secondary groups but it is due mostly to the H-H repulsion and charge polarization due to the incoming H atom. This geometry suggests that the S$\cdots$HO-Tyr configuration is likely to be the stable one, as it is confined once one proton is removed from the active site. When removing the hydrogen atom, the Kohn-Sham energy raises to -517.5309 Ha.

When considering the GSH-GST(Phe) system our starting structure s the X-ray structure of the native Tyr^7 complex, the only available crystallographic structure. Thus we use as the position of the blocked C_α atom of Phe^7 the same position of the native complex. The results obtained with this constraints are compared with those obtained without blockage on the C_α atoms of Phe^7, as described in the following.

When relaxing the ionic positions we obtain an energy of -502.0972 Ha. In this configuration the H atom attached to S looks unperturbed by the presence of the phenol group, i.e. the C-S-H triplet mostly retains its linear configuration. Once the H atom is removed from the sulfur atom, energy raises to -501.5597 Ha.

When allowing the whole Phe^7 group to move (unconstraining the C_α atom of Phe^7) global minimization furnishes an energy of -502.1002 Ha and -501.5650 Ha before and after the removal of the H atom attached to the S atom, respectively. Thus data on energetics enforces the picture that the S atom is protonated in the Phe^7 mutant.

Finally, we investigate the role of water in the case of the Phe^7 mutant, by introducing a water molecule near the active site. We obtain an energy of -519.2413 Ha which raises to -518.7208 Ha when the H atom from sulfur was removed.

Figures 4, 5, 6, 7, 8 and 9 show the optimized geometries of the large scale mod el representations of the GSH-GST interacting systems, obtained from our PW calc ulations.

The deprotonation energies obtained in this large scale model are significantely different from those computed for the reduced model (see Table 1.7), indicating the limits of the reduced model, which does not account for the effects of the protein cavity.
In order to check whether a QM-MM approach represented a good approx-

imation, in the presents case, for the description of the protein cavity, we optimized the geometries of GSH, GS^-, GSH-H_2O and GS^--H_2O adopting the QM-MM Oniom method. We first applied this approach, by including in the quantum part the smaller systems **1**, **2**, **5** and **6**, in order to check whether substantial differeneces appear, both in the energetics and in the geometry, of the considered species, and to compare these results with those obtained from the full quantum treatment. In Table 1.7 the variation in the computed energy differences between the reduced and large models, for the PW basis and the gaussian basis set **A**, with both XC functionals, have been collected. The considered energy differences correspond to the difference in the deprotonation energies, computed with the R.M. and the L.M.; hence a negative value means a lower deprotonation energy in the case of the L.M.. When we analyse data from Table 1.7 we notice that the effect of the environment, for Oniom calculations, is almost negligible; indeed the correspondent variations on the computed energy differences are extremely small, even if a general small amount of stabilization of the anion is achieved. Due to this fact we did not perform Oniom calculations on the larger systems **3** and **4**. Adopting a QM-MM approach we expect our large model to account for electrostatic effects of the cavity on the quantum part only, due to the impossibility of the quantum part to delocalize electron charge density on the classical environment. On the other hand, when we include the environment in the full *ab initio* PW calculation, environmental effects may result in charge transfer processes and polarization of the cavity, which are not taken into account with a simple electrostatic representation of the cavity. Indeed, environmental effects computed with the full *ab initio* PW calculations, are much larger, with an average difference in energy of -21.2 Kcal/mol. This is a clear sign of the role played in this case by the enviromnment, which we may rationalize in terms of the stabilization of the anion due to charge delocalization throughout the protein cavity.

Geometrical rearrangement of the active site should reflect the same kind of effect; we expect geometrical parameters to change much more in the case in which the environment is included in the *ab initio* calculation, than in the case a classical description is adopted. In order to clarify the effect of the cavity in both cases, we analyse the correspondent variations in geometrical parameters for basis set **A**, with both functionals, and for the PW basis. In Table 1.8 the correspondent differences in the main geometrical parameters for **1**, **2**, **5** and **6** are reported, for the gaussian basis set, while the corre-

spondent values for **3** and **4** are those computed with the PW basis set only. Table 1.8 confirms what we discussed for the energies, i.e. the variations on the geometrical parameters are much larger in the case the environment is accounted for in an *ab initio* fashion, than in the classical Oniom picture. It is also interesting to notice that changing the functional does not affect much the observed variations, as confirmed by the fact that the magnitude of the variation is comparable in both cases, and the sign is always the same, for the two considered functionals. From data contained in Table 1.8 it can be seen that the parameters which show larger variations are, both in the Oniom and *ab initio* approach, the r_{SH} distance and the $\angle OHS$ angle. The variations in these parameters indicate that a longer r_{SH} distance and a smaller value of the $\angle OHS$ angle are observed, which both indicate that weakening of the S-H hydrogen bond happens when protein environment effects are taken into account.

1.3.3 Thermal effects

In order to compare our results on the energetics for hydrogen bond formation, with measured pKs, an estimate of the difference in Gibbs free energy associated to the loss of the proton, has to be provided. The measured pKs are the ratio of concentrations of the dissociated species at constant atmospheric pressure and temperature of 300 K, that is one should compute the Gibbs free energies to properly evaluate the absolute pKs. Thus to obtain such a quantity we should perform a finite temperature calculation or to estimate entropies by using a proper statistical analysis. However, a standard way to account for thermal effects, is to compute harmonic vibrational frequencies, from which entropies and enthalpies, can be computed. So we performed frequency calculations on the reduced model, using the BPW91 functional with basis set **A**, and we used the computed corrections in order to evaluate Gibbs free energy for the considered deprotonation processes.
Enthalpies were determined using the following formula:

$$\Delta\ \mathrm{H}_{cal} = \Delta\ \mathrm{E}_e^0 + \Delta\ \mathrm{E}_v^0 + \Delta(\Delta\ \mathrm{E}_v)^{298} + \Delta\ \mathrm{E}_r^{298} + \Delta\ \mathrm{E}_t^{298} + \Delta\ \mathrm{PV}$$

where the first term is the computed hydrogen bonding energy, while the second term is the change in zero-point vibrational energy. The third term is the change in vibrational energy on going from 0 to 298 K. The fourth,

fifth and sixth terms are due to changes in rotation and translation energies, and pressure volume work, respectevely. Entropies were evaluated using a standard statistical mechanics approach. All the terms, except the last one, have been computed, for our R.M., with Gaussian 98, using the BPW91 functional only. The last term in the present case is -RT, -0.59 Kcal/mol at 298 K.
For the PW calculations on both the reduced and large model, we just added the thermal corrections obtained for the reduced model, with gaussian basis set, to the zero temperature energy differences computed with the PW basis. Indeed, when comparing differences in pKs of similar species, such as the Tyr^7 and Phe^7 mutants, one may assume that volume differences and entropic contributions from translations, rotations and harmonic oscillations for the cavity of the two species, are equal in the two systems and thus negligible when considering the differences in pK. In any case the computed free energy contributions to the zero temperature energy differences, have been found to give almost the same contribution for all the considered R.M. deprotonation reactions, with an average difference of about 1 Kcal/mol, each other. Hence, when considering differences of energy differences, thermal contributions from different deprotonation processes, should cancel each other. The Gibbs free energy differences for the R.M. and for the L.M., computed using data from taken from Table reftab:interaction, are reported in Table 1.9.
Following the same procedure as in ref. [5], we compute the variation of the pK_a in the binary complex for the wild-type enzyme and the mutant, considering the computed Gibbs energy differences associated to the loss of hydrogen bond in the neutral complex (Δ G_1), and in the deprotonated complex (Δ G_2). Δ G_1 represents the interaction free energy due to hydrogen bond, for the protonated active site thiol group, given by:

$$\Delta\ G_1 = G\ (GST(Tyr)...GSH(Phe))\ -\ G\ (GST(Tyr))\ -G\ (GSH(Phe))$$

while (Δ G_2) represents the interaction free energy due to hydrogen bond, when the thiol group is deprotonated:

$$\Delta\ G_2 = G\ (GST(Tyr)...GSH(Phe)^-)\ -\ G\ (GST(Tyr))\ -G\ (GS^-(Phe))$$

We then define:

$$\Delta\Delta\ G_{Tyr}\ \Delta\ G_1\ -\ \Delta\ G_2$$

The quantities measured experimentally are the pK_as of the thiol group in the protonated and deprotonated cases, which can be related to the $\Delta\Delta$ G, by considering that:

$$\Delta\Delta \text{ G} = \text{- 2.3 RT } \Delta \text{ pK}_a$$

In the case a water molecule replaces the Tyr group in hydrogen bonding, the interaction free energy for the deprotonated water complex has to be subtracted from Δ G_2. The correspondent expression for $\Delta\Delta$ G then becomes:

$$\Delta\Delta \text{ G}_{H_2O} = \Delta \text{ G}_1 \text{ - } \Delta \text{ G}_2 \text{ - } \Delta \text{G}_i nt \text{ (GS}^- \text{H}_2\text{O)}$$

The intreaction free energies for (GS^-H_2O), in the reduced model representation, have been computed to be 8.3 and 8.0 Kcal/mol, with gaussian and PW basis set, respectevely, which are in excellent agreement with the experimental value of 8 Kcal/mol [18]; extrapolation of the thermal contribution computed with gaussian basis set, has been added, as usual, to the PW energy difference. We can now provide computed values of $\Delta\Delta$ G for the reduced model of the system, with gaussian and PW basis sets, and an estimate for the large model adopting the same corrections as in the reduced model. The experimental value for $\Delta\Delta$ G is 2.2 Kcal/mol, which can be compared with the computed values, for the reduced and large model reported in Table 1.10. When we consider the case of a Tyr residue partecipating to the hydrogen bond in the deprotonated complex we find large discrepancies between theory and experiment, for the reduced model of the system. The computed value of $\Delta\Delta$ G is, for the gaussian basis set, 22.3 Kcal/mol, which is in good agreement with the value of 20.5 Kcal/mol reported in ref. [5]. The correspondent result obtained with a PW basis does not differ much from the value computed with the gaussian basis, even if some lowering in the value of $\Delta\Delta$ G is observed. This is consistent with the intuition that plane waves can better describe hydrogen bonding properties than gaussian basis sets, due to the delocalized nature of plane waves versus the localized nature of gaussian functions centered on atomic sites, which result in stronger hydrogen bonding properties. This point may be also rationalized in terms of a better description of the electronic structure of anions with delocalized basis sets like plane waves, than with localized functions; indeed diffuse functions are usually included in a gaussian basis set in order to account for delocalization of the negative charge on the anion. A better agreement with the experiment

is observed on going to the large model representation of the system, even if the computed value is still too large, resulting in a difference of about 11 Kcal/mol. A possible source of error could be found in terms of inaccuracies in the calculation of thermal contributions, due to errors in the computation of harmonic frequencies, and in the approximated extrapolation procedure, which does not consider changing of entropic contributions to the computed energy differences, on going from the R.M. to the L.M. representation of the system. However these effects alone cannot account for such a large discrepancy.
When considering the contribution of the water molecule, significant improvment in the agreement with the experimental data is observed for the reduced model, even if the computed value for $\Delta\Delta$ G is still too high both with gaussian and PW basis sets. When adopting the large model representation, a computed value for $\Delta\Delta$ G of 5.4 Kcal/mol is found, which is in good agreement with the experiment, taking into account the differences between the restricted computational model and the real entire enzyme, the fact that we are neglecting solvation effects, and the already discussed approximated extrapolation procedure. In this case, properly including environment effects at an *ab initio* level, results in an improvement in the description of the interactions between GSH's thiol group, and possible chemicals interacting through hydrogen bond. However environment effects alone do not account for the measured differences in pK of the protonated and deprotonated thiol group in the case we consider a Tyr residue interacting with the thiol group of GSH, while, when a water molecule is incuded in our large model representation of the enzyme cavity, much better agreement is observed. We can therefore conclude that a water molecule plays a significant role in the activity of the GSH thiol group, substituting the phenolic group of tyrosine in hydrogen bonding when the thiol group is deprotonated.

1.4 Conclusions

The interaction of glutathione (GSH) with the enzyme glutathio-S-transferases (GST) has been studied by means of high level *ab initio* DFT calculations. We considered a reduced model of the interacting species, adopting CH_3SH as a model for GSH, and Ph-OH (C_6H_5OH) as a model for the Tyr[7] residue of GST. Moreover we considered the case of a water molecule interacting with

the thiol group of GSH, in order to clarify whether or not a water molecule plays some role in the catalytic activity of GSH. In order to account for protein environment effects, we have considered a large model representation of the GSH-GST interacting system, including five blocking aminoacids Tyr^{7}, Arg^{13}, Trp^{38}, Gln^{49} and Gln^{62}, modeled by their side chain residues.
Differences in the value computed for the pK of the two species, indicate that the presence of an extra water molecule in the neighborhood of the active site has to be invoked to explain the data on pK, which is in agreement with the observed presence of several cristallographyc water molecules, that could replace the OH group of Tyr^{7} in hydrogen-binding the thiol group of GSH, once it has been deprotonated. Hence, our results show that a proper ab-initio representation of the GSH-GST complex allows to obtain a good agreement between the experimental and computational data on proton extraction energetics.

Bibliography

[1] H. W. Dirr, P. Reinemer, and R. Huber Eur. J. Biochem. **220**, 645 (1994).

[2] S. Liu et al., J. Biol. Chem. **267**, 4296 (1992).

[3] M. M. Kreevoy and T. M. Liang, J. Am. Chem. Soc. **102**, 315 (1980).

[4] W. W. Cleland, Biochemistry **31**, 318 (1992).

[5] Y.-J. Zheng and R. L. Ornstein, J. Am. Chem. Soc. **119**, 1523 (1997).

[6] P. Reinemer et al., EMBO J. **10**, 1997 (1991).

[7] C. L. Brooks III, M. Karplus and B. M. Pettitt, "Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics", Adv. Chem. Phys. **61**, John Wiley and Sons, New York, NY (1988).

[8] Gaussian 98 (Revision A.2), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.

[9] A. D. McLean and G. S. Chandler, J. Chem. Phys. **72**, 5639 (1980).

[10] A. K. Rappè, C.J. Casewit, K.S. Colwell, W. A. Goddard III, W. M. Skiff, J. Am. Chem. Soc. **114**, 10024 (1992).

[11] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[12] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[13] J. P. Perdew and Y. Wang (unpublished), J. P. Perdew, in *Electronic Structure of Solids '91*, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991).

[14] A. D. Becke, J. Chem. Phys. **98**, 5648-5652 (1993).

[15] C. Peng, P. Y. Ayala, H. B. Schlegel and M. J. Frisch, J. Comp. Chem. **17** , 49 (1996).

[16] S. Humbel, S. Sieber and K. Morokuma, J. Chem. Phys. **105**, 1959 (1996).

[17] In our plane-wave calculations, pseudopotentials are used to account for the core-valence interactions, and hence, the absolute value of the total energy is not meaningful. On the contrary, total energy differences (such as the deprotonation energy) are and can thus be compared with calculations using a gaussian basis.

[18] (a) Meot-Ner (Mautner), M. *J. Am. Chem. Soc.* **1988**, *110*, 3854-3858. (b) Gao, J.; Garner, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1986**, *108*, 4784-4790.

#	R.M.	L.M.
I	CH_3SH/CH_3S- + $(C_6H_5)OH$	GSH/GS- + Tyr^7
II	CH_3SH/CH_3S-	GSH/GS- + Phe^7 (blocked)
II'	CH_3SH/CH_3S- + C_6H_6	GSH/GS- + Phe^7
III	CH_3SH/CH_3S- + H_2O	GSH/GS- + Phe^7 + H_2O

Table 1.1: List of all the systems investigated in this study. We consider a reduced model (R.M.), in which the GSH molecule is represented by a CH_3SH molecule, and large scale model (L.M.), in which the GSH is complete.

	BPW91			**B3LYP**	
	A	**B**	**PW**	**A**	**B**
ΔE(# I)	0.54207	0.54469	0.54316	0.54374	0.54656
ΔE(# II)	0.57674	0.57969	0.57236	0.57501	0.57803
ΔE(# III)	0.55655	0.55979	0.55412	0.55593	0.55936

Table 1.2: Computed energy differences for the R.M. using the two basis sets and the two XC functionals. Energy differences in Hartree.

		BPW91			B3LYP	
		A	**B**	**PW**	**A**	**B**
2	r_{CS}	1.848	1.837	1.874	1.847	1.838
4	r_{CS}	1.846	1.835	1.859	1.846	1.836
	r_{OH}	1.077	1.077	1.099	1.034	1.034
	r_{SH}	1.918	1.911	1.888	2.020	2.010
	r_{OS}	2.994	2.987	2.986	3.054	3.044
	∠O-H-S	177.7	177.3	177.0	178.2	178.1
6	r_{CS}	1.847	1.836	1.850	1.848	1.837
	r_{OH}	1.016	1.017	1.039	0.998	0.999
	r_{SH}	2.169	2.149	2.218	2.225	2.207
	r_{OS}	3.173	3.158	3.158	3.206	3.196
	∠O-H-S	168.9	170.7	171.0	167.2	170.2

Table 1.3: Selected geometrical parameters for the R.M. **4** and **6** using basis sets **A**, **B** and the **PW** basis for the BPW91 functional; using basis sets **A**, **B** for the B3LYP functional.

		BPW91			B3LYP	
		A	**B**	**PW**	**A**	**B**
1	r_{CS}	1.838	1.826	1.871	1.835	1.825
	$r_{SH'}$	1.357	1.351	1.371	1.348	1.342
	$\angle$ C-S-H′	96.7	90.0	96.8	96.9	97.3
3	r_{CS}	1.840	1.828	1.870	1.837	1.826
	r_{OH}	0.984	0.983	1.001	0.972	0.971
	r_{SH}	2.392	2.392	2.311	2.426	2.436
	r_{OS}	3.343	3.343	3.300	3.363	3.373
	$r_{SH'}$	1.357	1.351	1.370	1.349	1.342
	$\angle$ O-H-S	162.2	162.6	169.2	161.6	161.9
	$\angle$ C-S-H′	97.1	97.3	97.5	97.3	97.6
5	r_{CS}	1.839	1.827	1.832	1.838	1.826
	r_{OH}	0.979	0.978	0.997	0.970	0.968
	r_{SH}	2.439	2.442	2.424	2.461	2.472
	r_{OS}	3.400	3.407	3.407	3.384	3.397
	$r_{SH'}$	1.357	1.351	1.373	1.348	1.342
	$\angle$O-H-S	167.2	168.8	168.4	158.9	159.6
	$\angle$ C-S-H′	96.9	97.2	97.1	97.2	97.4

Table 1.4: Selected geometrical parameters for the R.M. **3** and **5** using basis sets **A**, **B** and the **PW** basis for the BPW91 functional; using basis sets **A**, **B** for the B3LYP functional. $r_{SH'}$ and r_{SH} refer to the hydrogen bound to the sulfur and to the oxigen atoms respectevely.

	3	**4**	**5**	**6**
Δ E_{INT}^{PW}	-	-	-5.0	-16.4
Δ E_{INT}^{Gau}	-4.0	-25.8	-3.2	-16.0
Δ H_{INT}	-2.7	-26.0	11.6	-14.9
Δ G_{INT}	5.1	-17.2	4.3	-8.3

Table 1.5: Interaction enthalpies and free energies for the reduced models, computed at the BPW91-**A** level of theory.

	1	**2**	**3**	**4**	**5**	**6**
Q_{SH}	-0.05	-0.82	0.04	-0.51	0.00	-0.70
$Q_{Phe,H}$	-	-	-0.07	-0.25	0.24	0.21
Q_{CH_3}	0.05	-0.18	-0.03	-0.22	0.00	-0.20
Q_{OH}	-	-	0.05	-0.02	-0.24	-0.29

Table 1.6: Mulliken charges for molecules **1**, **2**, **3**, **4**, **5** and **6**, computed at the BPW91-**A** level of theory.

	A		**PW**
	BPW91	B3LYP	
Δ_{env} E(# 1)	-	-	-20.4
Δ_{env} E(# 2)	0.2	-0.2	-21.8
Δ_{env} E(# 2′)	-	-	-22.4
Δ_{env} E(# 3)	-0.5	-0.7	-20.1

Table 1.7: Environment effects on the deprotonation energies for basis set **A**, with both functional, and for the PW basis. For the gaussian basis set the large model were simulated by Oniom calculation, while for the PW basis the computed energy differences refer to a full *ab initio* calculation.

		A		**PW**
		BPW91	B3LYP	
1	$\Delta(\mathrm{r}_{CS})$	-0.011	-0.012	-0.005
	$\Delta(\mathrm{r}_{SH'})$	0.000	0.000	-0.004
2	$\Delta(\mathrm{r}_{CS})$	-0.015	-0.017	0.026
3	$\Delta(\mathrm{r}_{CS})$			0.012
	$\Delta(\mathrm{r}_{OH})$			0.002
	$\Delta(\mathrm{r}_{SH})$			-0.112
	$\Delta(\mathrm{r}_{SH'})$			-0.004
	$\Delta(\angle OHS)$			9.7
4	$\Delta(\mathrm{r}_{CS})$			0.001
	$\Delta(\mathrm{r}_{OH})$			0.046
	$\Delta(\mathrm{r}_{SH})$			-0.165
	$\Delta(\angle OHS)$			7.4
5	$\Delta(\mathrm{r}_{CS})$	-0.009	-0.010	-0.045
	$\Delta(\mathrm{r}_{OH})$	-0.002	-0.001	0.005
	$\Delta(\mathrm{r}_{SH})$	-0.029	-0.068	-0.187
	$\Delta(\mathrm{r}_{SH'})$	0.001	0.001	-0.002
	$\Delta(\angle OHS)$	19.1	7.5	23.9
6	$\Delta(\mathrm{r}_{CS})$	-0.013	-0.012	-0.004
	$\Delta(\mathrm{r}_{OH})$	0.014	.014	0.020
	$\Delta(\mathrm{r}_{SH})$	0.049	0.038	-0.130
	$\Delta(\angle OHS)$	14.2	8.3	1.3

Table 1.8: Environment effects on the geometries for basis set **A** , with both functionals, and for the PW basis. The values reported correspond to the difference between geometrical parameters in the reduced and large model, such that a negative value means an increment of the considered parameter, when passing from the R.M. to the L.M.. For the gaussian basis set the large model were simulated by Oniom calculation, while for the PW basis the computed energy differences refer to a full *ab initio* calculation. $\mathrm{r}_{SH'}$ and r_{SH} refer to the hydrogen bound to the sulfur and to the oxigen atom, respectevely.

	R.M.		L.M.
	A	**PW**	**PW**
ΔE(# I)	0.53174	0.53283	0.50627
ΔE(# II)	0.56735	0.56297	0.52813
ΔE(# III)	0.54747	0.54504	0.51296

Table 1.9: Energy differences for the R.M. and the L.M. representations of GSH-GST, including thermal effects. For the PW calculations thermal contributions have been extrapolated from data in Table reftab:interaction. Energy differences in Hartree.

	R.M.		L.M.
	A	**PW**	**PW**
$\Delta\Delta$ G $_{Tyr}$	22.3	18.9	13.7
$\Delta\Delta$ G $_{H_2O}$	14.0	10.6	5.4

Table 1.10: $\Delta\Delta$ G $_{Tyr}$ and $\Delta\Delta$ G $_{H_2O}$, computed including thermal effects extrapolated from the gaussian basis set calculation of Table1.5 to the plane-wave basis set results. The energies are expressed in kcal/mol.

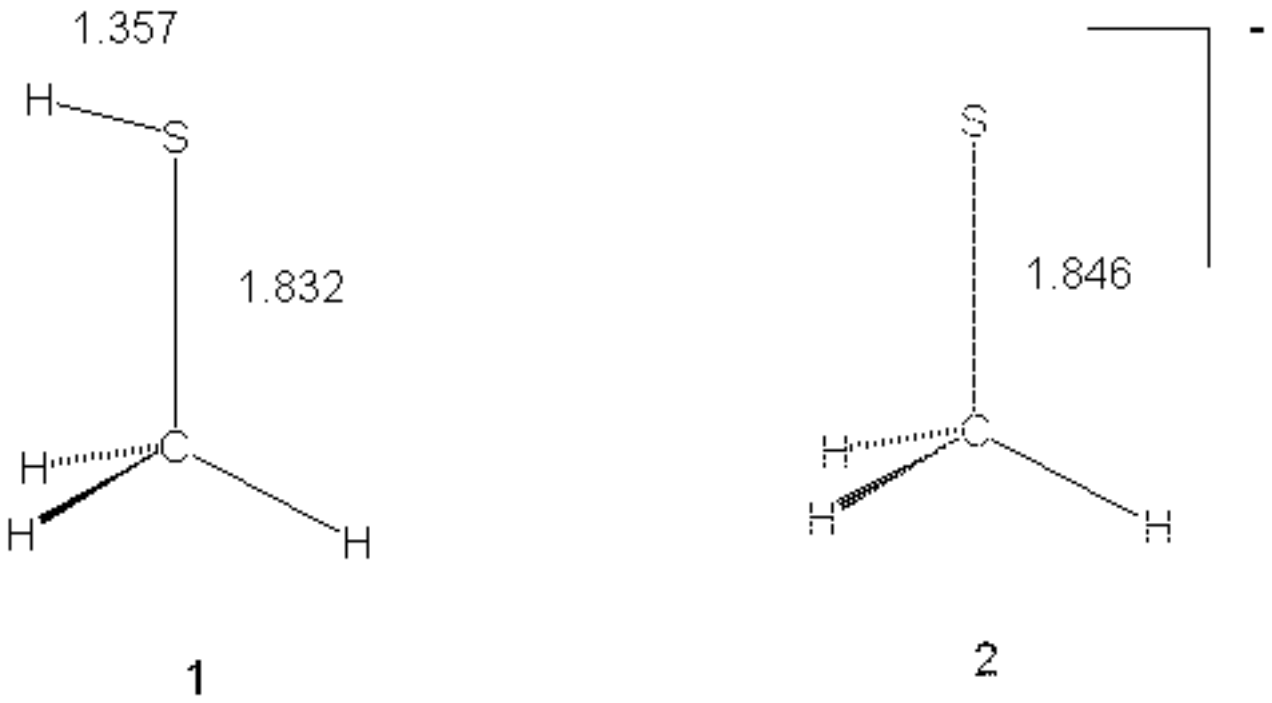

Figure 1.1: Reduced models: main geometrical parameters for molecules **1** and **2**, computed at the BPW91-**A** level of theory.

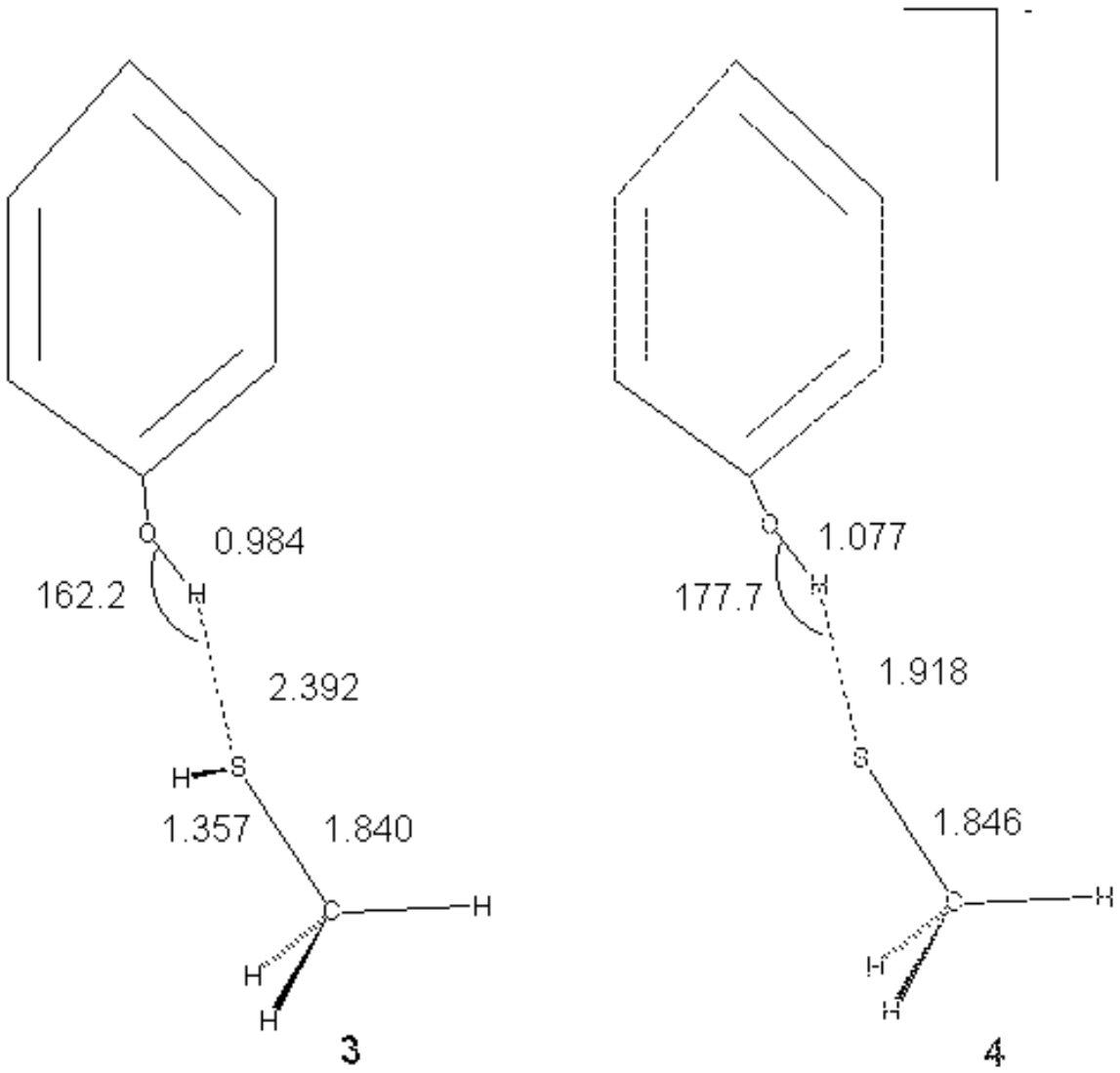

Figure 1.2: Reduced models: main geometrical parameters for molecules **3** and **4**, computed at the BPW91-**A** level of theory.

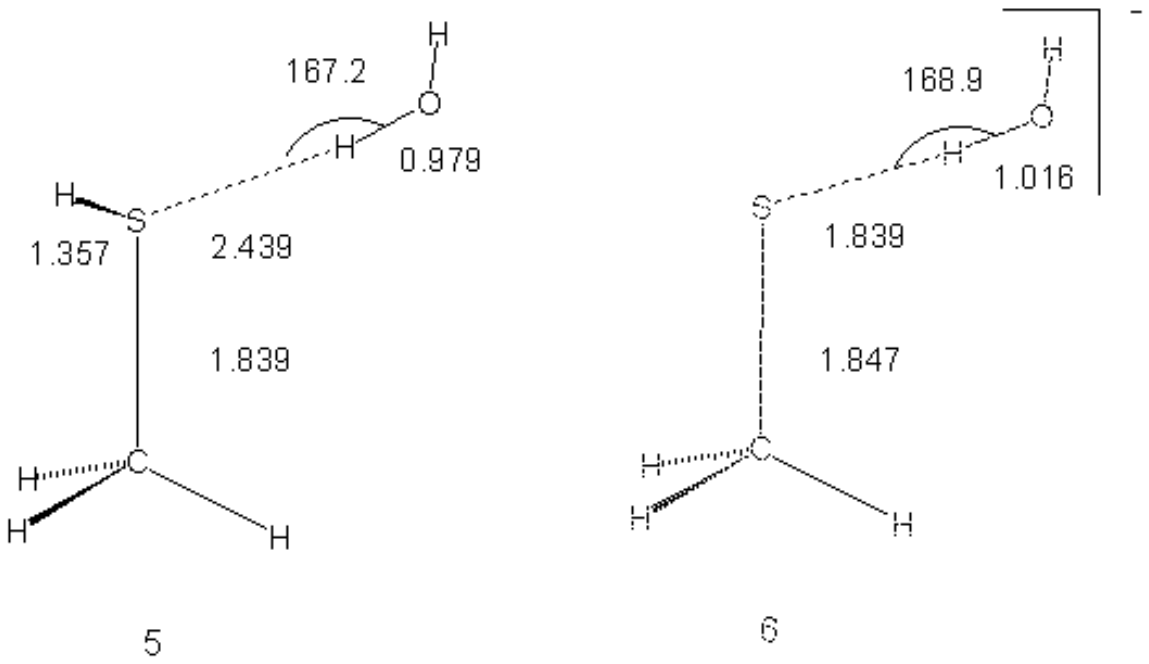

Figure 1.3: Reduced models: main geometrical parameters for molecules **5** and **6**, computed at the BPW91-**A** level of theory.