THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

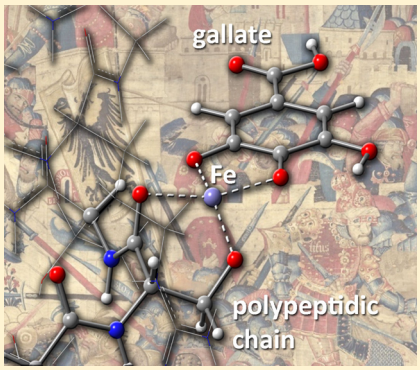
pubs.acs.org/JPCB

A Strategy for the Study of the Interactions between Metal−Dyes and Proteins with QM/MM Approaches: the Case of Iron−Gall Dye

Sandro Jurinovich,* Ilaria Degano,* and Benedetta Mennucci*

Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Risorgimento 35, Pisa 56126, Italy

ABSTRACT: Historical textiles dyed with tannins usually show more extended degradation than fabrics dyed with other coloring materials. In order to shed light on this phenomenon we investigated the molecular interactions between tannin dyes and protein-based textiles using quantum-mechanical tools. In particular, we focused on the iron−gall complex with a fragment of α-helix wool keratin. We developed a step by step protocol which moves from the simplest ternary complexes with free amino acids (all treated quantum mechanically) to the more realistic system of the polypeptide fragment (treated at QM/MM level), passing through an intermediate model of interacting sites to evaluate the local environmental effects. The analysis of the interactions between the iron−gall complexes and free amino acids allowed us to identify possible coordination modes as well as determining their relative geometries. However, we also showed that only with the addition of the proteic environment a detailed picture of the interaction sites and binding modes can be achieved. An important role is in fact played by the microenvironment which can favor specific coordinations with respect to others due to both structural and electronic changes in the possible interaction sites.



1. INTRODUCTION

State-of-the-art analytic procedures and newly developed spectroscopic techniques have become of fundamental importance in the cultural heritage field[1,2] as they allow the identification of dyeing materials and their degradation products. However, these approaches cannot give information on the molecular-level phenomena which determine the properties of the investigated materials. Computational chemistry can provide a valid complementary approach to the experimental investigations.[3−5] In particular, it can show important structure−property relations of the fundamental artwork components and disclose the role played by the various components (e.g., the dye, the metal ion and the environment) in determining their behavior. Within this line, here, we present a computational study on the interactions between a metal−dye and an α-helix polypeptide fragment representative of wool keratin proteins. In particular, we focused on tannin-based dyes because historical textiles dyed with tannins showed more extended degradation phenomena than textiles dyed with other coloring materials, even in the same condition of preservation.[6] The peculiar role of tannins and metal ion mordants in the degradation processes has not been clarified yet but an important role seems to be played by their tendency to promote redox reactions. In most cases, mordants are metal salts that can form a metal complex with natural colorants, which exhibit increased affinity to the substrate. Depending on the metal character, complex formation does not only strengthen dye fixation on the substrate but it also changes the color of the dyeing.[7]

In the past, the most commonly used mordant for dyeing with tannins was *copperas* (iron(II) sulfate), which both enhances dye binding process and allows the rise of the dark color.[8] Iron metal ions form complexes with components derived from tannins, in particular with gallotannins, essentially constituted by polyphenols, which are the hydrolyzable part of tannin natural extracts. Gallic acid and ellagic acid are the main components of hydrolyzed gallotannin mixture in tannin-based dye baths.[9] The dark color rises from the large UV−Vis absorption due to formation of a charge-transfer complex between polyphenols and iron. The resulting iron−gall complexes are known to bind with the protein substrate but neither the coordination sites nor the complexes' geometries are known. In this work, we developed an analytic strategy aimed at clarifying these aspects by explicitly simulating the interactions among the three components: (1) the organic dye (3,4,5-trihydroxybenzoic acid, gallic acid, which is the main component of tannin dyes); (2) the metal, iron(II), used as mordant; and (3) the α-helix polypeptide chain fragment. The strategy was based on a full QM description of the iron−gall complexes to obtain an accurate estimation of both the structural and the electronic characteristics of the system, whereas the protein has been modeled using a mixed QM and classical description. The analysis of the interactions between the iron−gall complexes and free amino acids allowed us to identify possible coordination modes as well as determining their relative geometries. However, we also showed that only with the addition of the proteic environment a detailed picture of the interaction sites and binding modes can be achieved. An important role is in fact played by the microenvironment which can favor specific coordinations with respect to others due to

Received: August 21, 2012
Revised: October 16, 2012
Published: October 25, 2012

dx.doi.org/10.1021/jp3083002 | J. Phys. Chem. B 2012, 116, 13344−13352

both structural and electronic changes in the possible interaction sites.

2. METHODS AND SYSTEMS

Many spectroscopic and electrochemical experimental studies confirmed that both iron(II) and iron(III) ions form complexes with gallic acid.[10−15] Iron(III) complexes are slightly soluble in water and they precipitate. They are characterized by a strong absorption peak which gives a dark-blue shading. Iron(II) complexes are soluble in water and their formation is generally a slow reaction. The addition of ferrous salts as mordants, reported in many recipes for dye bath preparation, does not exclude the presence of iron(III) in solution. The redox chemistry with both free and complexed iron (II/III) is complicated. In this work, we investigated iron(II)−gall complex, GAFe(II), as a molecular model for the chromophore interacting with wool fibers. This complex is also the main component in analogous iron−gall ink mixture, commonly used for writing in black.[16] As concerns its stoichiometry in aqueous solution, we followed what is reported in the literature, namely a 1:1 complex.[15]

All complexes of GAFe(II) and free amino acids were described at density functional theory (DFT) level. GAFe(II)−amino acid ternary complex geometry optimizations were performed with M06 hybrid exchange functional, which proved suitable to describe main-group chemistry as well as transition metals.[17,18] A reasonably large 6-311G(d,p) basis set was adopted to correctly describe the valence shell electrons of iron. All calculations were performed in vacuo using Gaussian09 suite of programs[19] with the default criteria for the SCF density-based convergence, the optimization thresholds and integration grid. Iron(II) ion exists in two stable electronic spin states: a singlet closed-shell low-spin (LS) state and a quintet high-spin (HS) state. Both high-spin and low-spin states were taken into account for iron(II) complexes calculations.

The study of the interaction of the iron−gall complex with a polypeptide chain was performed on a molecular model system consisting of a small B0LKP1 α-helix protein fragment,[20] one of the typical keratinaceous proteins in wool keratin fiber, very rich in cysteine residue. The molecular structure of B0LKP1 protein was obtained by the homology modeling method using the integrated Web-based system developed by Swiss Model,[21,22] set to default options to generate the target structure. Afterward, a small fragment consisting of 16 amino acidic residues and containing one cysteine residue was selected in order to describe the peculiar interactions involving thiol group. The two ends of the polypeptide chain were capped with methyl groups and the acidic residues were considered in the protonated form except Glu391, which can interact with the basic protonated residue of Arg398.

The structure of the selected polypeptide fragment was preliminarily optimized at MM level of theory, using AMBER force field.[23] Terminal methyl groups were kept frozen during the optimization process to retain the α-helix domain, which was well described by the starting structure obtained by homology modeling. Figure 1 shows the MM fragment optimized structure used as input to build ternary complexes with GAFe(II) and highlights the investigated binding sites.

The geometry of the GAFe(II)−polypeptide ternary complex was optimized using an ONIOM(QM/MM) approach combining the M06/6-311G(d,p) level for the QM part with the AMBER force field for the classical part.[24] According to the ONIOM scheme, the full system was partitioned into a "model system" formed by the so-called "interaction site" and the hydrogen link atoms used to saturate the dangling bonds. In particular, the interaction site is made of (1) complex atoms; (2) the atoms of the amino acidic residues in the polypeptide chain that directly interact with the metal; and (3) the chain fragment required to cut a C−C bond instead of C−N or C—(C═O), as indicated in Figure 2. In order to include the electrostatic effects of the low layer atoms on the model system, we used the electronic embedding formalism implemented in Gaussian09.

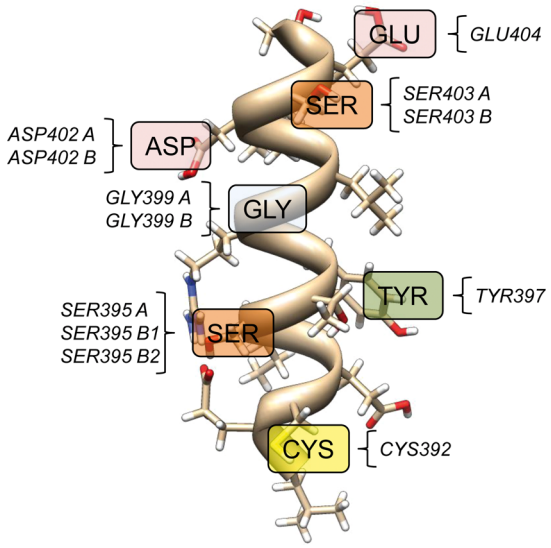


Figure 1. AMBER-optimized structure of 390−405 BOLKP1 protein fragment. Different interaction sites are identified on the α-helix.

Finally, to compare energies of different isomers, we performed a single point energy calculation at full M06/6-311g(d,p) level of theory on the optimized ONIOM structures. Each model system was also optimized out of the protein

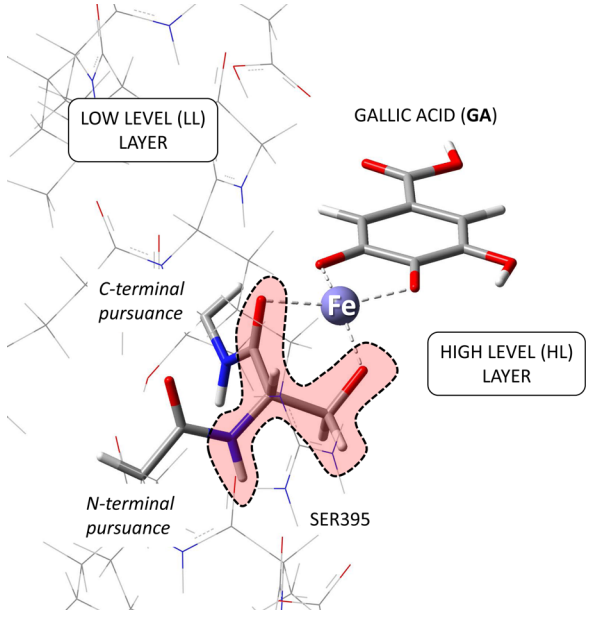


Figure 2. Components of the ONIOM scheme using [GAFe(II)-Ser395]$^-$ complex as an example. The interaction site, which constitutes the high-level layer, is represented in tube. The red highlight identifies the Ser395 residue atoms.

dx.doi.org/10.1021/jp3083002 | J. Phys. Chem. B 2012, 116, 13344−13352

environment at M06/6-311g(d,p) level of theory, capping the terminal carbons with three hydrogen atoms and keeping the terminal methyl group frozen during the optimization process. This strategy allowed us to compare the QM optimized model system geometry on the "real" polypeptide one, in order to investigate the protein environment effects on each specific interaction site.

3. RESULTS AND DISCUSSION

The analysis is divided in two sections: an initial study on iron–gall complexes with free amino acids, which represent a first-order approximation of a polypeptide, followed by the main investigation involving different binding sites on a real polypeptidic system. A final, third, section summarizes and discusses the relative stability of protein-bound GAFe(II) complexes with respect to the electronic and structural effects due to the environment.

3.1. Ternary Complexes with Free Amino Acids. It is well-known that amino acids can chelate metal ions and the literature reports several experimental[25−27] and theoretical[28−34] works describing binary complexes between free amino acids or small peptides and metal ions. Nonetheless, not much data can be found about ternary complexes, where another ligand is present together with the amino acid,[14] even if this information can be important in the modeling of metal ion–protein or –peptide complexes.[35] For this reason, in the first part of the study we investigated ternary complexes between GAFe(II) and selected free amino acids in order to compare energies and structures of different coordination modes.

Depending on the structure, we identified three potential metal binding sites in which the amino acid acts as mono-, bi-, or tridentate ligand. Metal binding sites are (1) the amino nitrogen; (2) the carboxylic oxygen; and (3) the donor functional group in the side chain, if present. Figure 3 shows the selected coordination modes for a generic amino acid.

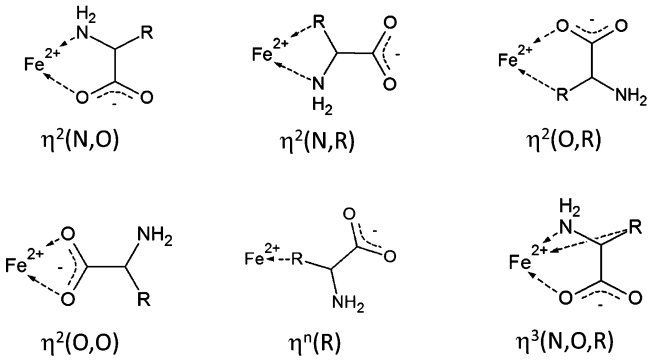


Figure 3. Schematic structures of different coordination modes involving a generic amino acidic with only one side functional group. Interaction η^2(N,O) is possible for all amino acids. Amino acids are here represented in a deprotonated form.

DFT geometry optimizations were performed on [GAFe(II)AA]$^{n-}$ complexes where AA stands for glycine (Gly), serine (Ser), cysteine (Cys), or aspartic acid (Asp) and n indicates the total charge. The selected amino acids are those present in higher percentage in the wool keratin protein composition as indicated by different authors.[36−39] In our approach, all structures were considered as negatively charged (n = 2) and the carboxylic, hydroxyl, and thiol groups were assumed as deprotonated while the amino group was $-NH_2$. Gly, which has no functional groups on the side chain, was considered mono-deprotonated (n = 1). These ionization forms correspond to very alkaline conditions and cannot be considered as representative of the normal complexation conditions in the dye bath. However, this model allowed us to compare different ligand donors in the same conditions and also to directly compare relative energies.

Our results showed that HS complexes are always about 50 kcal/mol more stable than LS ones and also metal to ligand bond distances are about 0.2 Å longer in HS cases than in LS ones. Iron cation is always maintained on the same plane of gallate aromatic ring. In the following, we will discuss only one isomer for each coordination mode even if all possible isomers were investigated.

3.1.1. [GAFe(II)Gly]$^-$ Complexes. Gly is the smallest and simplest amino acid. It is largely used as model system for computational and theoretical investigations on the interactions between metal cations and biological systems in many life processes.[32] We investigated η^2(N,O) and η^2(O,O) coordination modes starting from different initial guess structures in order to identify all possible conformers. Optimized structures are reported in Figure 4.

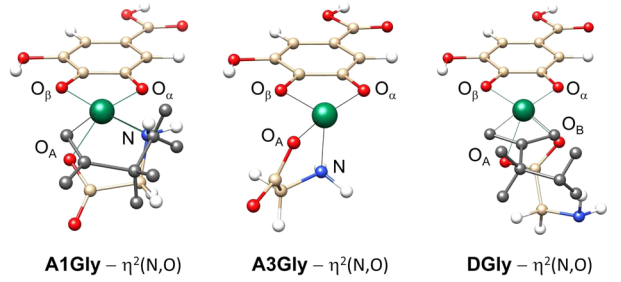


Figure 4. [GAFe(II)Gly]$^-$ optimized structures at M06/6-311G(d,p) level of theory. Colored structures refer to the low-spin state. Structures of high-spin complexes are reported in gray; in this case GAFe(II) atoms are overlaid over the low-spin ones.

In the A1Gly and DGly HS structures, the amino acid donor atoms are placed on the same plane of GAFe(II) whereas in LS structures planar geometry is slightly distorted and O_A atoms are about 20° out of plane. The enhanced planarity in the HS structures is confirmed by the absence of A3Gly "see-saw" structure for the HS electronic configuration. η^2(N,O) coordination mode gives rise to a stable five-membered chelate ring and it results in a 13.7 kcal/mol more stable structure than the η^2(O,O) coordination mode involving the carboxylate group.

[GAFe(II)Gly]$^-$ hydrated structures were also optimized in order to investigate the effect of the presence of two water molecules saturating the octahedral coordination sphere of iron. The relative energy differences between HS/LS isomers were almost maintained, but in the hydrated complexes, ligand atoms are always placed on the iron–gall plane, also in LS structures.

3.1.2. GAFe(II) Complexes with Functionalized Amino Acids. The same computational strategy used for Gly complexes was applied to the ternary complexes with serine (Ser), cysteine (Cys), and aspartic acid (Asp). These amino acids can act as bi- and tridentate chelates. The results for Ser and Cys were quite similar because of their analogous structures. These complexes showed similar metal coordination geometries with different Fe–S and Fe–O distances because of the larger radius of the sulfur atom. Asp is characterized by a

dx.doi.org/10.1021/jp3083002 | J. Phys. Chem. B 2012, 116, 13344−13352

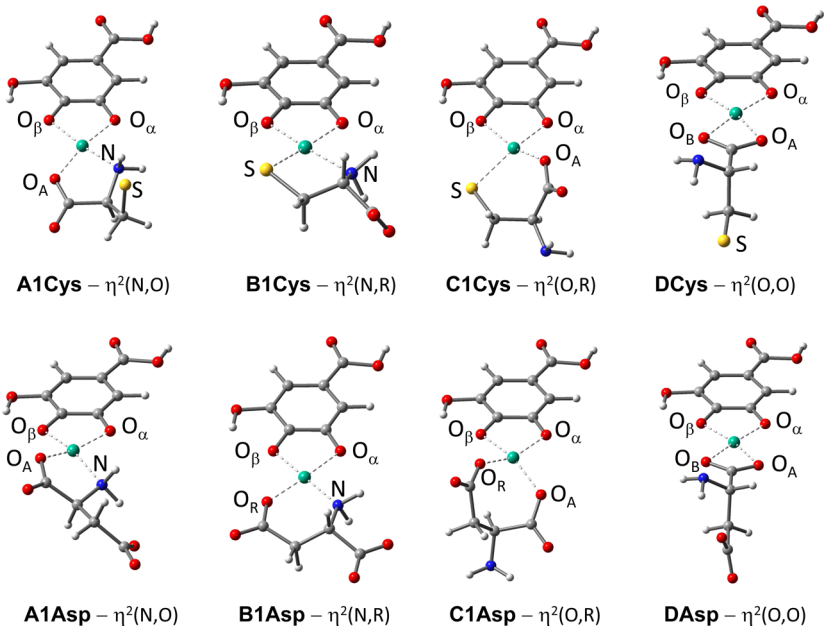


Figure 5. Selected bidentate $[GAFe(II)Cys]^{2-}$ (top) and $[GAFe(II)Asp]^{2-}$ (bottom) high-spin optimized structures at M06/6-311G(d,p) level of theory.

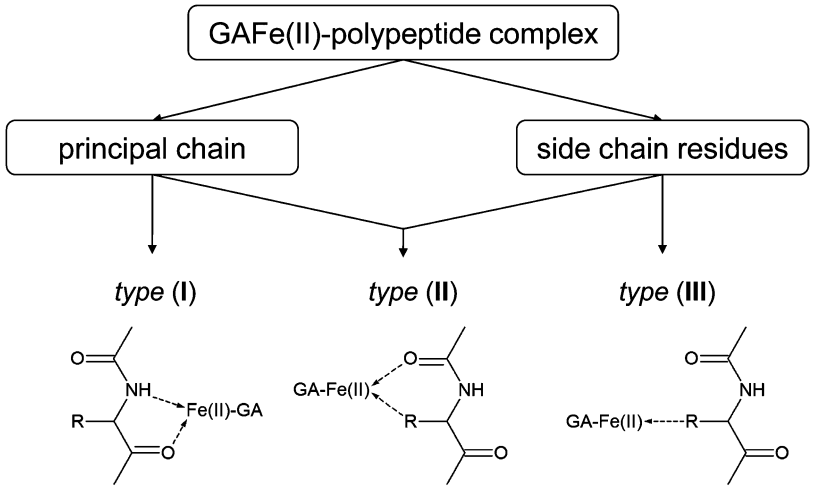


Figure 6. Possible interactions between GAFe(II) and the α-helix polypeptide fragment. Type (I) interactions involving only N,O atoms; type (III) involving only the side chain functional groups; type (II) involving both atom types.

longer side chain. Also in this case, HS structures resulted to be the most stable and their geometries were similar to the LS ones; HS-optimized structures for bidentate $[GAFe(II)Cys]^{2-}$ and $[GAFe(II)Asp]^{2-}$ complexes are reported in Figure 5.

The obtained η^2(N,O) and η^2(O,O) geometries were very similar for all amino acids. Side functional groups were involved in η^2(O,R), η^2(O,O) and η^2(N,O,R) coordination modes and ring sizes vary from Ser and Cys to Asp. Ligand donor atoms tended to be placed on the same GAFe(II) plane, consistent with the ring size. Five-membered rings showed a typical "envelope" conformation whereas six-membered rings showed a "pseudo boat" conformation.

$[GAFe(II)Asp]^{2-}$ differs from the other complexes because of the larger ring size. In particular, we obtained a quite tensioned seven-membered ring for C1Asp isomer. For $[GAFe(II)Asp]^{2-}$, another η^2(O,O) coordination mode involving the COO^- side-chain group instead of the amino acidic COO^- is also available. By comparing the energies, we saw that η^2(N,R) and η^2(O,R) are more stable than the "basic" η^2(O,O). In all cases, η^2(O,O) coordination mode is energetically unfavored whereas the η^3(N,O,R) tridentate complex shows the same stability of the η^2(N,R) and η^2(O,R) coordination modes. The increasing number of donor group does not provide a sensible complex stabilization as the ring tension rises in such tridentate complexes.

3.2. GAFe(II) Complexes with an α-Helix Polypeptide Chain. Further, we investigated the interactions between iron–gall complexes and a model α-helix polypeptide chain with an amino acidic composition that can be considered representative of wool keratin fiber proteins. The polypeptide fragment allowed us to investigate the interaction sites including all the amino acids treated above and also other peculiar residues like glutamic acid (Glu) and tyrosine (Tyr). Seven binding sites, highlighted in Figure 1, were investigated at QM/MM level of theory with respect to both the low-spin and the high-spin electronic configurations.

dx.doi.org/10.1021/jp3083002 | J. Phys. Chem. B 2012, 116, 13344–13352

The electronic structure of the peptide is different from that of a free amino acid because the carboxyl and amino groups are converted in the amide group by the peptide condensation reaction. This leads to a modification of the donating properties: amino lone pair is partially engaged in a C−N double bond formation, according to the amide resonance formula. On the other hand, carbonyl lone pairs can be a good metal ion ligand as suggested by the experimental review of Siegel and Martin[25] and also by theoretical studies.[30]

In the amide structure it is impossible to reproduce a η^2(N,O) coordination mode on the same amino acidic residue. However, this coordination mode can be obtained involving O and N amide atoms belonging to different residues in the chain. Furthermore, the polypeptide geometry is characterized by the secondary α-helix structure and O and N amide atoms are involved in strong intramolecular H-bonds. For these reasons it is fundamental to take into account the role of the side-chain functional groups in the metal−polypeptide coordination. Amino acidic side chains are accessible and suitable for metal coordination. On the selected polypeptide fragment, we therefore hypothesized three different coordination modes, which are reported in Figure 6. In the following discussion, iron−gall polypeptide complexes are indicated with [AA]$^{n-}$ label, where AA stands for the specific binding site and n indicates the total charge.

3.2.1. Interaction Type I. The type I interaction involves only O and N main-chain atoms. A careful examination of the fragment geometry suggests that interaction type I can be obtained from binding sites near the Gly residue. The absence of side groups on Gly residue allows the GAFe(II) complex to approach the α-helix main chain without any steric hindrance. Focusing on Gly399 residue, two binding sites were identified, as shown by the optimized QM/MM structures reported in Figure 7.

In [Gly399A] structure there are no side groups that can directly interact with iron, which is only bound to threonine (Thr) 396 O_A and Gly399 N_B atoms. The gallate hydroxyl group also interacts with both the Thr396 hydroxyl group and the arginine (Arg) 398 NH_2 via H-bonds. Ligand atoms are all on the same plane whereas the gallate aromatic ring is quite bent down, probably because of its intermolecular interactions.

[Gly399B] structure differs from [Gly399A] because of Ser403−OH group located near the complex. We investigated two structures, which differ by the $N_B-C_1-C_2-O_H$ dihedral angle. The optimization process gave two different minima on the potential energy surface. The most stable was [Gly399B2] where the −OH can interact with iron along the axial vacant binding site on the octahedral iron coordination sphere. Energy difference between the two [Gly399B] structures was about 20 kcal/mol, suggesting the presence of a strong Fe−OH interaction. In this case, HS structures differed from LS ones only for the Fe-ligand bond length and they were also about 50 kcal/mol more stable than the corresponding LS, as shown in Table 1.

Binding interactions between the iron and the ligands were characterized by evaluating C═O and C−N bond length variation (Δr) due to polypeptide complexation with the iron−gall molecule. We noted that C_A═O_A were elongated while C_A-N_A were shortened as expected from the amide resonance formula. On the contrary, C_B-N_B were elongated while C_B-O_A were shortened due to the fact that the Fe−N interaction involved the amino lone pair subtracting electron density from C−N bond.

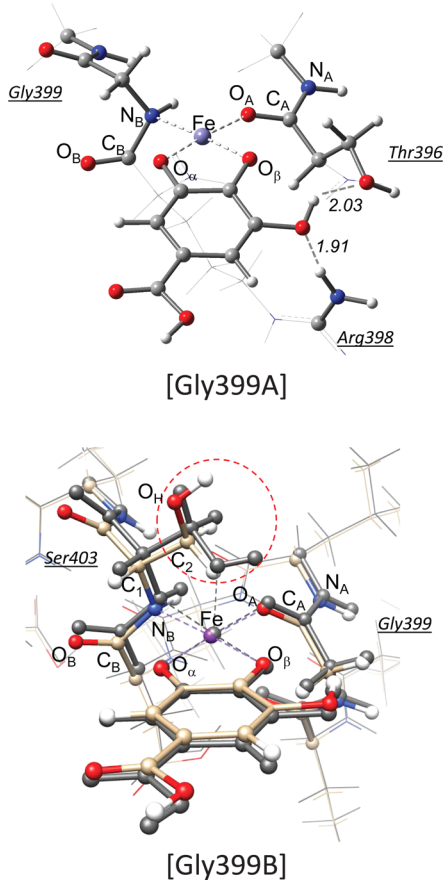


Figure 7. [Gly399] LS-optimized structures at QM/MM level of theory. Gly399B overlaid structures mainly differ in the Ser403−CH_2OH conformation (atoms in dashed red circle); the colored layer refers to [Gly399B1] and the gray layer refers to [Gly399B2] conformers. HS structures are very similar to the LS ones and thus they were omitted.

Table 1. [GAFe(II)Gly399] Structural and Energetic Data of in Vacuo QM/MM-Optimized Structures[a]

	B1		B2		A	
ΔE	20.3	*14.1*	0	*0*	29.1	*26.3*
ΔE_{HL}	52.4		46.1		49.0	
$r(Fe-O_1)$	1.865	*1.913*	1.867	*1.916*	1.785	*1.876*
$r(Fe-O_2)$	1.894	*1.958*	1.910	*1.970*	1.904	*2.001*
$r(Fe-O_A)$	1.988	*2.087*	2.030	*2.229*	1.929	*2.011*
$r(Fe-N_B)$	2.047	*2.172*	2.061	*2.218*	2.044	*2.229*
$\Delta r(C_A-O_A)$	0.011	*0.016*	0.013	*0.018*	0.023	*0.023*
$\Delta r(C_A-N_A)$	−0.007	*−0.013*	−0.007	*−0.012*	−0.012	*−0.013*
$\Delta r(C_B-O_B)$	−0.019	*−0.019*	−0.017	*−0.016*	−0.019	*−0.017*
$\Delta r(C_B-N_B)$	0.077	*0.080*	0.077	*0.075*	0.081	*0.059*

[a] ΔE (kcal/mol) refers to M06/6-311G(d,p) energy difference with respect to the most stable structure. ΔE_{HL} (kcal/mol) represents the energy difference between high spin and low spin isomers and it is defined as $\Delta E_{HL} = E_{HS} - E_{LS}$. Bond lengths ($r$) are expressed in Å and Δr refers to bond length difference between free and complexed polypeptide. Normal text values refer to LS complex and italic values refer to HS isomers.

3.2.2. Interaction Type II. The type II interaction refers to the η^2(O,R) coordination mode. Ser405, Ser395, Asp402, and Cys396 binding sites were investigated as examples. Depending on the local side-chain distribution, we modeled two different

dx.doi.org/10.1021/jp3083002 | J. Phys. Chem. B 2012, 116, 13344−13352

binding possibilities: (A) involving C═O and R groups of the same interacting amino acidic residue; and (B) involving a C═O belonging to a different amino acid residue.

An example is offered by Ser binding site. Figure 10 reports the LS optimized $[Ser395]^-$ and $[Ser403]^-$ structures for both A and B isomers. No relevant structural differences were found between HS/LS isomers.

In $[Ser395A]^-$ there is no steric hindrance due to closer side chains. On the contrary, in $[Ser395B]^-$, iron−gall complex directly interacts with Arg398 (B1) and Cys392 (B2). For B structures only C═O groups belonging to the fourth below residue interacted with iron. In interaction type (II) complexes, Fe−O_A bond lengths varied in a range of 0.15 Å depending on the specific site. The average HS Fe−O_A distance was 2.15 Å. This value is in agreement with 2.05 Å found in $[GAFe(II)AA]^{2-}$ complexes. Also the Fe-R bond lengths are similar. Fe−O_A bonds can be further characterized in terms of the C_A═O_A and C_A−C_N bond length variation (Δr) with respect to the "free" and iron−gall complexed polypeptide: Δr values, reported in Table 2, showed the same trend observed for [Gly399] complexes.

Table 2. Bond Length Variations in Interaction Type II Structures[a]

	$\Delta r(C_A{=}O_A)$		$\Delta r(C_A{-}N_A)$	
Ser395A	0.025	*0.004*	−0.018	*−0.016*
Ser395B1	0.018	*0.011*	−0.019	*−0.013*
Ser395B2	0.022	*0.026*	−0.017	*−0.023*
Ser403A	0.027	*0.015*	−0.003	*−0.005*
Ser403B1	0.080	*0.035*	−0.003	*−0.027*
Ser403B2	0.016	*0.025*	−0.010	*−0.016*
Asp402A1	0.024	*0.015*	−0.021	*−0.018*
Asp402A2	0.037	*0.034*	−0.011	*−0.019*
Cys392	0.017	*0.018*	−0.021	*−0.024*

[a] Δr (Å) refers to bond length difference between free and complexed polypeptide. Positive values indicate a bond elongation whereas negative values indicate a bond shortening. Normal text values refer to LS complex and italic values refer to HS structures.

3.2.3. Interaction Type III. We investigated the interaction type III for the glutamic acid (Glu) 404, Asp402, and tyrosine (Tyr) 397 residues: QM/MM-optimized structures are reported in Figure 8. Asp402 and Glu404 iron−gall complexes refer to the bidentate η^2(O,O) coordination mode. In $[Glu404]^-$, the iron−gall complex COO^- group is placed on the same plane of iron−gall and an intermolecular H-bond between the O carbonyl gallate and −OH tyrosine phenolic group is present. COO^- coordination geometry found for Asp404 was different from the expected planar metal coordination observed in other η^2(O,O) interactions even if Fe−O bond lengths are similar. Gallate carboxylic group was also about 10° out of plane because of the strong interaction with the $-NH_2$ arginine group below.

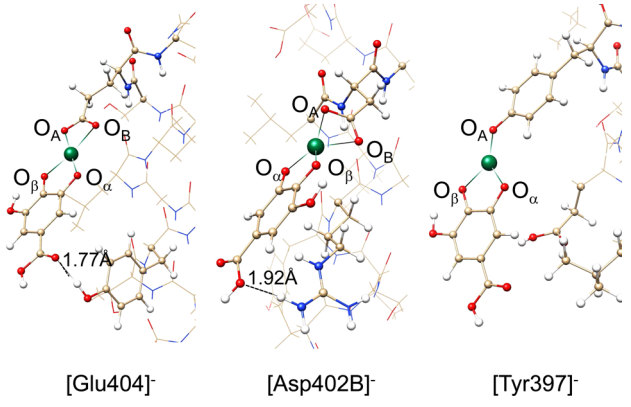


Figure 8. HS interaction type III optimized structures at the QM/MM level of theory. LS structures are not reported because there are no relevant geometry differences between LS/HS isomers.

Tyr397 iron−gall complex acts as a monodentate η^1(R) ligand. Only the phenolic group in the tyrosine is bound to the iron. The Fe−O_R bond length is similar to that found in $[GAFe(II)Ser]^{2-}$ complexes. Tyrosine and gallate aromatic rings are placed on the same plane, and gallate is affected by the proximity of Glu and isoleucine (Ile) side-chain groups.

3.3. Discussion. The relative stability of the various structures can be estimated on the basis of a full QM energy comparison. The graph in Figure 9 reports the energy differences (ΔE) calculated with respect to the most stable structure, for both the type II and III interactions and the low-spin and high-spin structures. [Gly399] iron−gall complexes were omitted because of the different polypeptide chain protonation.

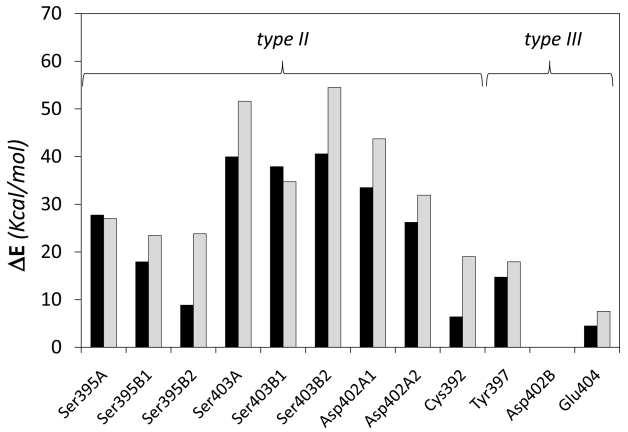


Figure 9. Relative energy difference calculated at M06/6-311g(d,p) level of theory on optimized $[AA]^-$ QM/MM structures with respect to the most stable $[Asp402B]^-$, set at ΔE = 0. Black histograms refer to LS structures, and gray histograms refer to HS.

High-spin isomers were on average about 58 kcal/mol more stable than low-spin ones. This ΔE_{HL} is similar to the difference found in GAFe(II) and $[GAFe(II)AA]^{n-}$ complexes. The maximum deviation with respect to the ΔE_{HL} average difference was about 10 kcal/mol for $[Ser403B1]^-$. For this kind of systems, we found that ΔE_{HL} energy gaps are mainly dependent on an intrinsic stabilization of the high-spin state. Geometry changes due to spin-state configuration do not seem to be relevant.

A general discussion about the stability order may be formulated on the LS only (or HS only) results because the two sets of data show an analogous trend. Even if there is not a very clear trend that shows a preference for one type of interaction with respect to the others, type III interactions seem to be more stable with respect to type II interactions, with η_2(O,O) ordination modes in $[Asp402B]^-$ and $[Glu404]^-$ being the most stable ones. $[Tyr397]^-$ ΔE is slightly higher because of its monodentate coordination unlike the bidentate coordination of

dx.doi.org/10.1021/jp3083002 | J. Phys. Chem. B 2012, 116, 13344−13352

the others. The type II interaction shows a wide range of ΔE depending both on the specific amino acid residue and the local protein environment of each site. $[Cys392]^-$ and $[Ser395B2]^-$ are the most stable in this group, but in the latter system a strong intermolecular interaction with Arg398 contributes to the energy stabilization.

The energy differences shown in Figure 9 result from the combined effects determined by three major aspects: (1) the type of coordination η^2(O,O) or η^2(O,R); (2) the type of amino acidic residue; and (3) the local proteic microenvironment that surrounds each interaction site. Unfortunately, it is very difficult to separate these contributions because they are all coupled; however, the analysis reported in the previous sections on the ternary complexes with the free amino acids allow us to appreciate the main effects due to points (1) and (2).

In order to point out the role of the microenvironment, we introduced a more detailed analysis based on the QM investigation of electronic and structural changes induced by the protein. This analysis was applied to the two serine residues placed in different areas of the polypeptide chain, as shown in Figure 1. Structures of A, B1, and B2 isomers were similar for both $[Ser395]^-$ and $[Ser403]^-$, as reported in Figure 10. It was

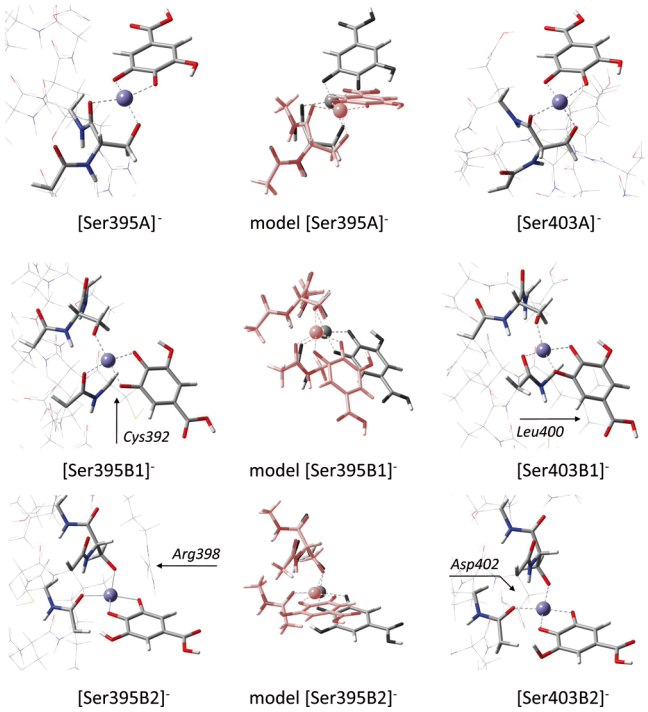


Figure 10. $[Ser395]^-$ (left column) and $[Ser403]^-$ (right column) optimized at the QM/MM level of theory (LS). Structures in the center column represent the isolated model systems $[Ser395]^-$ optimized at the QM level of theory (red structures) starting from the QM/MM-optimized geometries, represented in gray.

also found that $[Ser395]^-$ structures were always more stable than the corresponding $[Ser403]^-$, which showed the same metal coordination geometry, but inserted in a different protein environment.

To better understand these differences, we extracted the interaction sites' structures (from now indicated as *model systems*) from the "full" QM/MM calculations. Afterward, geometry optimizations of these model systems were performed freezing the terminal methyl groups (which represent the link atoms to the polypeptide chain). The resulting structures, which correspond to the interaction sites' relaxed geometries, did not significantly differ from the original ones obtained in the QM/MM optimization of the "full" polypeptide. In all cases, Fe-ligands' coordination geometries were maintained. Differences in Fe-ligands' bond lengths were generally less than 0.04 Å whereas other internal coordinates, such as dihedral angles, varied in consequence of (i) the absence of N−H and C═O main-chain groups giving H-bonds which support the α-helix skeleton, and (ii) the absence of steric hindrance due to closer residues which may give rise to intermolecular repulsions with the ligand.

Significant differences appeared instead by comparing the energies of the QM model systems when relaxed or extracted from the "full" calculations. Such energies, reported as histograms in Figure 11, are defined in terms of eq 1

$$\Delta E_{TOT} = \langle \psi_p | H_p^0 | \psi_p \rangle - \langle \psi_r^0 | H_r^0 | \psi_r^0 \rangle \quad (1)$$

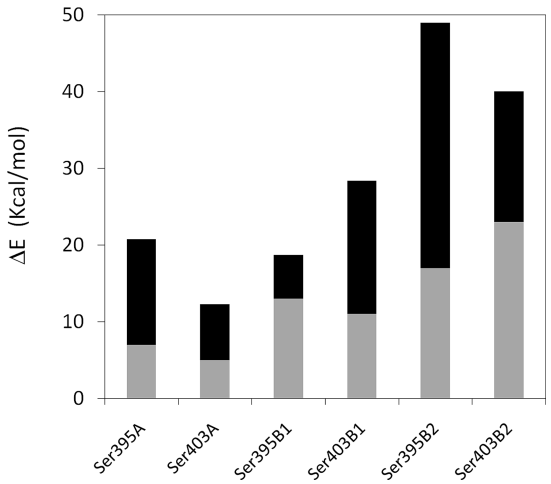


Figure 11. Energy differences (kcal/mol) between $[Ser]^-$ isomers in isolated relaxed QM model system and in QM model systems resulting from QM/MM calculations taking into account the wave function polarization due to the MM protein environment. Each column refers to ΔE_{TOT} quantity (eq 1), the black portion is the ΔE_{el} contribution (eq 2) and the gray portion is the ΔE_r contribution (eq 3).

where the p subscript refers to the model system geometry as extracted from the "full" calculation whereas r refers to the model system relaxed geometry. ψ^0 is the wave function associated with the specific geometry, whereas ψ is the polarized wave function due to the effects of the classical electrostatic field used in the MM description within the "full" calculations.

$[SerB2]^-$ isomers showed the largest ΔE_{TOT} values: this is due to the strong intermolecular interactions between the GAFe(II) and the local surrounding proteic environment, such as (i) the electrostatic interactions with positive charged Arg398 and in $[Ser395B2]^-$, and (ii) the H-bond between the GA hydroxyl group and Asp402 COOH group in $[Ser395B2]^-$. For the other isomers, which were not affected by so strong electrostatic interactions, energy differences were smaller.

In order to separate the electronic and the geometrical contributions, ΔE_{TOT} energy differences were partitioned in ΔE_{el} and ΔE_r, defined by the following equations:

$$\Delta E_{el} = \langle \psi_p | H_p^0 | \psi_p \rangle - \langle \psi_p^0 | H_p^0 | \psi_p^0 \rangle \quad (2)$$

dx.doi.org/10.1021/jp3083002 | J. Phys. Chem. B 2012, 116, 13344−13352

$$\Delta E_r = \langle \psi_p^0 | H_p^0 | \psi_p^0 \rangle - \langle \psi_r^0 | H_r^0 | \psi_r^0 \rangle \quad (3)$$

In $[Ser395B2]^-$ complex the strong electrostatic interactions with positive charged Arg392 residue determined a ΔE_{el} energy gap of 32 kcal/mol, whereas in the analogous interaction site $[Ser395B1]^-$, which was far from the charged residue, ΔE_{el} were lowered at 6 kcal/mol.

The component ΔE_r mainly takes into account the structural effects. High values of this quantity indicate that the protein environment strongly affects the GAFe(II) interaction site geometries. ΔE_r contribution was smaller for $[SerA]^-$ isomers than for the other systems and this behavior was justified by the absence of closer amino acidic residues, which may create a steric hindrance for the polypeptide complexation in $[SerB]^-$ isomers.

4. CONCLUSIONS

We developed an analytic strategy to investigate molecular interactions in the gallate−iron−polypeptide system. The strategy addresses the problem using a step by step protocol.

We started from the simple ternary systems, where the interaction of the iron−gall complex with different free amino acids were investigated. In this first step, the theoretical approach was based on a full QM description in order to obtain an accurate estimation of both the structural and the electronic characteristics of the systems. Interactions between GAFe(II) and free amino acids contributed to predict coordination Fe-ligand geometries and bond lengths. We found that amino acidic coordination mode involving the amino or carboxylic group and the donor side-chain group and forming a 5- or 6-membered chelate rings were the most favored.

We then moved to the real gallate−iron−polypeptide system where we have mainly considered coordination modes allowed by the secondary structure offered by the α-helix fragment. Here, besides the type of coordination η^2(O,O) or η^2(O,R) and the type of amino acidic residue, an important role is played by the local proteic microenvironment surrounding each interaction site. To account for the protein effects, we included in the computational model a α-helix fragment, thus improving the description of the real "interaction site". To this aim, an ONIOM QM/MM formulation was used. Seven different binding sites and three interaction types were investigated with respect to the coordination geometry, the bond lengths, and the relative isomer stability. In particular, the type III interaction involving only the side-chain functional groups and η^2(O,O) coordination modes resulted the most stable. Also, interactions with C═O and donor side groups are possible, but the molecular stability depends on the amino acidic residue type and also on its position in the polypeptide chain. A detailed analysis on both the electronic and structural changes induced by the local environment due to the residues around the different coordination sites was also performed on selected complexes.

The developed protocol is a very general one and it can be extended to other organic dyes in analogous α-helix peptide chains. Also, other iron oxidation states and different metal ions, used as mordant in natural dye bath preparations, can be investigated with exactly the same protocol. An important aspect has been, however, neglected, namely the solvation effect. This indeed represents a further perturbation which can significantly affect the nature and the relative strengths of the interactions between the metal−dye complex and the different protein residues. Some preliminary tests performed using a continuum approach through a PCM external shell[40] mimicking a solvated protein matrix have shown that all the different coordination sites are equally stabilized: this seems to indicate that no important differential stabilizations should be expected. Clearly, this is what comes out from a "mean field" approach and a different picture could be obtained by investigating specific solvent interactions, such as the presence of water molecules in the iron coordination sphere or other intermolecular H-bonds between free −OH and COOH gallic acid groups with water or other aminoacidic residues. Further developments in the direction of including these effects are therefore necessary to get a final and more complete picture of how dyes interact with protein-base textiles.

AUTHOR INFORMATION

Corresponding Author

*E-mail: s.jurinovich@studenti.unipi.it (S.J.); ilariad@dcci.unipi.it (I.D.); bene@dcci.unipi.it (B.M.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge VAT project "The Short Life of Tannins", founded by PAR-FSE Regione Toscana 2011-2013, for financial support.

REFERENCES

(1) Janssens, K.; van Grieken, R. *Non-destructive microanalysis of cultural heritage materials*; Elsevier: Amsterdam, 2005; Vol. *42*.

(2) Degano, I.; Ribechini, E.; Modugno, F.; Colombini, M. P. *Appl. Spectrosc. Rev.* **2009**, *44*, 363−410.

(3) Fantacci, S.; Amat, A.; Sgamellotti, A. *Acc. Chem. Res.* **2010**, *43*, 802−813.

(4) Jacquemin, D.; Preat, J.; Wathelet, V.; Perpète, E. *J. Chem. Phys.* **2006**, *124*, 074104.

(5) Tilocca, A.; Fois, E. *J. Phys. Chem. C* **2009**, *113*, 8683−8687.

(6) Hofenk de Graaf, J. H. *The colourful past: origins, chemistry and identification of natural dyestuffs*; Archetype Publications: London, 2004.

(7) Mussak, R.; Bechtold, T. In *Handbook of Natural Colorants*; Mussak, R., Bechtold, T., Eds.; John Wiley and Sons: New York, 2009; Chapter 18.

(8) Onal, A. *J. Sci. Ind. Res.* **2005**, *64*, 491−495.

(9) Julkunen-Tiitto, R.; Häggman, H. In *Handbook of Natural Colorants*; Mussak, R., Bechtold, T., Eds.; John Wiley and Sons: New York, 2009; Chapter 5.

(10) Kipton, H.; Powell, J.; Taylor, M. *Aust. J. Chem.* **1982**, *35*, 739−756.

(11) Jančovičova, V.; Čeppan, M.; Havlinova, B.; Rehakova, M.; Jakubikova, Z. *Chem. Pap.* **2007**, *61*, 391−397.

(12) Hynes, J. M.; Coinceanainn, M. *J. Inorg. Biochem.* **2001**, *85*, 131−142.

(13) Andjelković, M.; Camp, J. V.; Meulenaer, B. D.; Depaemelaere, G.; Socaciu, C.; Verloo, M.; Verhe, R. *Food Chem.* **2006**, *98*, 23−31.

(14) Fazary, A. E.; Taha, M.; Ju, Y. *J. Chem. Eng. Data* **2009**, *54*, 35−42.

(15) Lu, L.; Li, Y.; Lu, X. *Spectrochim. Acta Part A* **2009**, *74*, 829−834.

(16) Kolar, J.; Strlič, M. *Iron Gall Inks: on manufacture characterisation degradation and stabilisation*; National and University Library: Ljubljana, Slovenia, 2006.

(17) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157−167.

(18) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci,

B.; Petersson, G. A.; et al. *Gaussian09, Revision A.1*; Gaussian Inc.: Wallingford, CT, 2009.
(20) Crewther, W.; Inglis, A. S.; McKern, N. M. *J. Biochem.* **1978**, *173*, 365−371.
(21) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. *Bioinformatics* **2006**, *22*, 195−201.
(22) Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. *Nat. Protoc.* **2009**, *4*, 1−13.
(23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(24) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct. THEOCHEM* **1999**, *461*−*462*, 1−21.
(25) Sigel, H.; Marin, B. *Chem. Rev.* **1982**, *82*, 385−426.
(26) Kiss, T.; Sóvágó, I.; Tóth, I.; Lakatos, A.; Bertani, R.; Tapparo, A.; Bombi, G.; Martin, R. B. *Dalton Trans.* **1997**, 1967−1972.
(27) Téllez, C. A.; de Moraes Silva, A.; Felcman, J. *J. Raman Spectrosc.* **2004**, *35*, 19−27.
(28) Shankar, R.; Kolandaivel, P.; Senthilkumar. *J. Phys. Org. Chem.* **2011**, *24*, 553−567.
(29) Pesonen, H.; Aksela, R.; Laasonen, K. *J. Phys. Chem. A* **2010**, *114*, 466−473.
(30) Xu, J. *J. Mol. Struct. THEOCHEM* **2005**, *757*, 171−174.
(31) Hongqi, A.; Yuxiang, B.; Ping, L.; Zhiqiang, L.; Xiangquan, H.; Chen, Z. *J. Phys. Org. Chem.* **2005**, *18*, 26−34.
(32) Marino, T.; Toscano, M.; Russo, N.; Grand, A. *J. Phys. Chem. B* **2006**, *110*, 24666−24673.
(33) Ai, H.; Bu, Y.; Li, P.; Li, Z.; Hu, X.; Chen, Z. *J. Phys. Org. Chem.* **2005**, *18*, 26−34.
(34) Fleming, G. J.; McGill, P. R.; Idriss, H. *J. Phys. Org. Chem.* **2007**, *20*, 1032−1042.
(35) Hoyau, S.; Ohanessian, G. *J. Am. Chem. Soc.* **1997**, *119*, 2016−2024.
(36) Maclaren, J. A.; Milligan, B. In *The Chemical Reactivity of the Wool Fibre*; Press, S., Ed.; Science Press: Marricksville, Australia, 1981; pp 1−16.
(37) Jones, L. N.; Rivett, D. E.; Tucker, D. J. In *Wool And Related Mammalian Fibres*; Lewin, M., Pearse, E. M., Eds.; CRC Press: New York, 2006; Chapter 5.
(38) Leeder, J. D.; Marshall, R. C. *J. Textile Inst.* **1941**, *32*, T83−T108.
(39) Gillespie, J. M.; Broad, A.; Reis, P. J. a. *Biochem. J.* **1969**, *112*, 41−19.
(40) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3093.