Understanding the Structural and Binding Properties of Collagen: A Theoretical Perspective

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JULY 2004

Impact Factor: 3.3 · DOI: 10.1021/jp049172z

3 AUTHORS, INCLUDING:

Simona Bronco
National Research Council (CNR, Pisa, Italy)
74 PUBLICATIONS **864** CITATIONS
SEE PROFILE

Chiara Cappelli
Scuola Normale Superiore di Pisa
117 PUBLICATIONS **1,866** CITATIONS
SEE PROFILE

Understanding the Structural and Binding Properties of Collagen: A Theoretical Perspective

Simona Bronco, Chiara Cappelli,* and Susanna Monti*

POLYLAB-INFM Pisa, c/o Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Risorgimento 35, I-56126 Pisa, Italy, and Istituto per i Processi Chimico-Fisici (IPCF-CNR), Area della Ricerca, via G. Moruzzi 1, I-56124 Pisa, Italy

Received: February 24, 2004; In Final Form: April 23, 2004

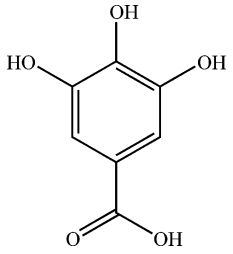
The change in the stability and conformational dynamics of a collagen-like microfibril segment (CMS), 23 amino acid residues long (around $5[3(GLY-X-Y)_8]$), as a result of the interaction with water, formaldehyde, and gallic acid is studied by performing a series of nanosecond molecular dynamics simulations. Major changes in the conformation of CMS occur when interacting with water while in the other solutions: (1) pure formaldehyde, (2) 8.0% (v/v) formaldehyde/water, and (3) 1.4% (v/v) gallic acid/water; the variation is less evident. Possible CMS/modifying agents' binding sites are evidenced by the analysis of radial distribution functions, coordination number, and H-bonding network. In particular, in the case of formaldehyde a high preference for ARG, LYS, and GLN residues is found. Gallic acid molecules preferably bind to PRO and HPR residues. The examination of average interaction energies per residue and their van der Waals and electrostatic components confirms the results obtained with the structural analysis. Furthermore, both van der Waals and electrostatic terms are important for the stabilization of the CMS. This is especially noted when CMS interacts with gallic acid. A comparison between calculated and available experimental findings is proposed. The agreement is very satisfactory, thus validating the computational approach to these systems.

1. Introduction

In this paper the structural and binding properties of collagen as a result of the interaction with different polar chemicals (from now on called modifying agents) are studied. Among the various classes of molecules able to interact with collagen, we focus on aldehydes and vegetal tannins, and in particular on formaldehyde and gallic acid (see Scheme 1), which can be considered the simplest prototypes of the two classes of compounds. The ability of formaldehyde to interact with collagen has been known for a long time[1]. In fact, it has been used as a preserving agent for biological specimens or in embalming, as well as in the tanning process for the production of leather. Even more interesting is the binding between collagen and gallic acid, or more generally polyphenolic compounds. Not only can these molecules act as modifying agents[1], but their antioxidant role in carcinogenesis has also been evidenced (see, for example, Tang et al.[2] and Gomes et al.[3]). Thus, the understanding of their interaction with biopolymers can be used as a starting point to elucidate their biological role. In addition, recently the collagen-gallic acid binding properties have been exploited to develop an optical sensor based on FTIR spectroscopy.[4]

Collagens are a large and growing family of proteins sharing some common traits but also exhibiting wide differences and achieving different functional roles concerned principally with the maintenance of tissue and cellular shape, strength, and structural integrity. Collagens are one of the most crucial components of the extracellular matrix and the fibril-forming biochemical types I, II, III, V, and XI are the most abundant and extensively studied. In addition to biological function they are the target of a variety of applications in industry (medical, pharmaceutical, and the food and leather sectors).

SCHEME 1: Gallic Acid



Only a small number of papers on the experimental study of the interaction between collagen and modifying agents is present in the literature. With regard to the binding properties of collagen, it has been shown that water is one of the major contributors to the enthalpy of denaturation of collagen.[5] Experimental findings about the interaction between collagen and aldehydes suggest that the enhanced collagen chemical and hydrothermal stability is a result of the fiber cross-linking. In particular, aldehydic compounds bond covalently with the $-NH_2-$ and $-NH-$ groups of collagen amino acidic side chains.[1,6]

Moving to the binding to polyphenolic compounds, there is experimental evidence, obtained by differential scanning calorimetry (DSC), that the interaction with polyphenols increases the hydrothermal stability of collagen[2] and that the interaction between polyphenols and collagen fibers is a synergistic effect of hydrophobic association and hydrogen bonding.[6] However, the reaction is a reversible process and polyphenols bind in a relatively weak way to each individual binding site.[6] In more detail, studies about the polyphenol/protein binding suggest that polyphenols interact preferentially with PRO[7] and HPR residues.

* To whom correspondence may be addressed: E-mail: chiara@dcci.unipi.it (C.C.); s.monti@ipcf.cnr.it (S.M.). Tel.: +39-050-2219293 (C.C); +39-050-3152520 (S.M.). Fax: +39-050-2219260 (C.C.); +39-050-3152442 (S.M.).

10.1021/jp049172z CCC: $27.50
Published on Web 06/12/2004

In addition, Baxter et al.[8] have shown that PROs perform dual functions: they act as binding sites and at the same time keep the peptide extended, thus maximizing the available binding surface. NMR spectroscopy experiments have provided further evidence that the dominant mode of interactions between the PRO-rich peptides and polyphenol molecules is a hydrophobic association, although hydrogen bonding may play a role. On the other hand, PRO and HPR are not the only possible binding sites and, as demonstrated by Nuclear Overhauser Effect (NOE) measurements, strong interactions also occur with ARG side chains.[9] Moving to the effect of binding on the collagen structure, it has been shown that the structure of collagen is almost unaffected by the interaction with modifying agents. In fact, even a chemical modification of the amino groups has little influence on collagen at the level of triple helical stability.[10]

This short summary on the available experimental studies on the subject of this work shows the difficulties in treating this matter from the experimental point of view. This is mainly due to the intrinsic complexity of the collagen system, which indeed increases if the interaction with the "environment" is taken into account. For this reason, in this case especially the computational investigation can be of help in supplementing and rationalizing experimental findings. However, the treatment of a complex system needs the use of complex models and perhaps this is the reason the literature on computational approaches to this matter is very limited. In fact, at least to the best of our knowledge, no computational studies on the interaction between collagen and formaldehyde have been performed so far. Moving to gallic acid, a series of recent computational studies has been performed on this topic.[11−14] However, in the cited references the analysis is limited to a small portion of a collagen-like peptide (a triple helical portion, at most) treated by using molecular mechanics, molecular dynamics, PM3 Hamiltonian, or density functional theory, depending on the complexity of the system. In particular, in Madhan et al.[13] molecular dynamics has been used to investigate the possible interaction sites between gallic acid and a small collagen-like peptide. In this paper a far more complex atomistic picture of the collagen structure is exploited. In particular, the spatial assembly of five triple helix segments to build up a microfibrillar structure interacting with water, formaldehyde, or gallic acid is considered.

The determination of the collagen structure has been a matter of study for many years. It is widely accepted that a collagen molecule consists of three polypeptide chains, called α-chains, arranged in a tight triple-helical entity.[15−18] The helical conformation of each chain is dependent on the presence of Glycine (GLY) residues every two residues and on the high content of Proline (PRO) and Hydroxyproline (HPR) residues. The HPR residues increase significantly the conformational stability of a collagen triple helix. The stabilizing effect is stereoselective and position dependent; only HPR residues with R stereochemistry and a suitable position along the sequence (Y position) are stabilizing.[5,19−26] In all fibrillar collagens the α-chains include an uninterrupted sequence of about 300 GLY-X−Y triplets flanked by much shorter terminal domains of different structure. Depending on the collagen type and on the tissue, triple helices can either be homo- or heterotrimers. Collagen type I, the most abundant, is usually a heterotrimer formed by two α_1 helices plus an α_2 helix. The triple helix organization leads to the formation of a long, rodlike structure, stiff but flexible, about 1.5 nm wide and over 300 nm long with globular domains at both ends. A careful analysis of the primary structure of collagen[16] reveals the presence of some patterns and motifs, most of them consisting of a periodic distribution of certain sequences or features. Polar and hydrophobic residues are periodically clustered along the sequence of collagen I every 234 residues. The precise nature of the packing of collagen helices has been widely investigated and it is generally agreed that groups of four to six triple helices aggregate to form microfibrillar structures[27−29] where molecular translations are cyclic. In reality, each model is a compromise; each has its strength and weakness and receives its share of confirmations and refutations from the various experimental techniques available.[16]

From what has been discussed so far it should be clear that, due to the complexity of the collagen structure, to obtain a reliable description (and hopefully prediction) of the collagen structural and binding properties, an extended model is needed. The definition of such a model is the subject of the following section. The system has been studied by means of molecular mechanics (MM) and molecular dynamics (MD) simulations with the AMBER force field,[31] which has already been successfully exploited for the study of collagen triple helices.[19] The details on the calculations are given in Section 3, whereas the results are reported in Section 4. A short summary and some conclusions end the paper.

2. Reduced Collagen Model System

A three-dimensional computer model of the bovine type I collagen microfibril based on the Smith description, consistent with both X-ray diffraction and electron microscopy data, was kindly made available to us by Professor Eleanor M. Brown.[31] In this model a microfibril is defined as a bundle of five triple-helical molecules pentagonally grouped, resulting in 15 polypeptide chains having 102 residues each, with a center-center distance of around 10 Å between adjacent triple helical molecules. The different residues defining the helices are not distributed uniformly throughout the length of the molecule, but form instead regions of relatively high and low hydrophobicity. Hydrophilic segments are wider than hydrophobic segments and have a lower amount of helical character.

In light of the computational resources at our disposal (see the following section), molecular simulations of such a huge matrix (even bigger if the modifying agent and solvent molecules are considered), are feasible only by restricting the number of repeating units of the collagen sequence. For this reason, a short segment 23 amino acid (AA) residues long (around $5[3(GLY-X-Y)_8]$), rich in hydrophilic residues, thus particularly apt to the study of the interactions with water, formaldehyde, and gallic acid, was extracted from the microfibril model. A picture of this segment, from now on called a collagen microfibril segment (CMS) is reported in Figure 1. The amino acid content of CMS in comparison with the whole collagen model is reported in Figure 2 in the form of histograms. All the amino acidic residues present in the large microfibril model can also be found in CMS. The ratio between the total amount of hydrophilic and hydrophobic residues is also preserved in CMS.

To build up the reduced model, the hydrogens were added by taking into account the pH of the environment; hydrogen coordinates were optimized with AMBER7[32] and its 1999 force field while keeping fixed all the other atoms. In this work, depending on the nature of the interacting agent, two different CMS structures have been considered. One corresponds to a neutral solution, in which the residues GLU and ASP are not protonated, the residues LYS and ARG are protonated, and the residues HIS are neutral (in this case a charge of +4e results in

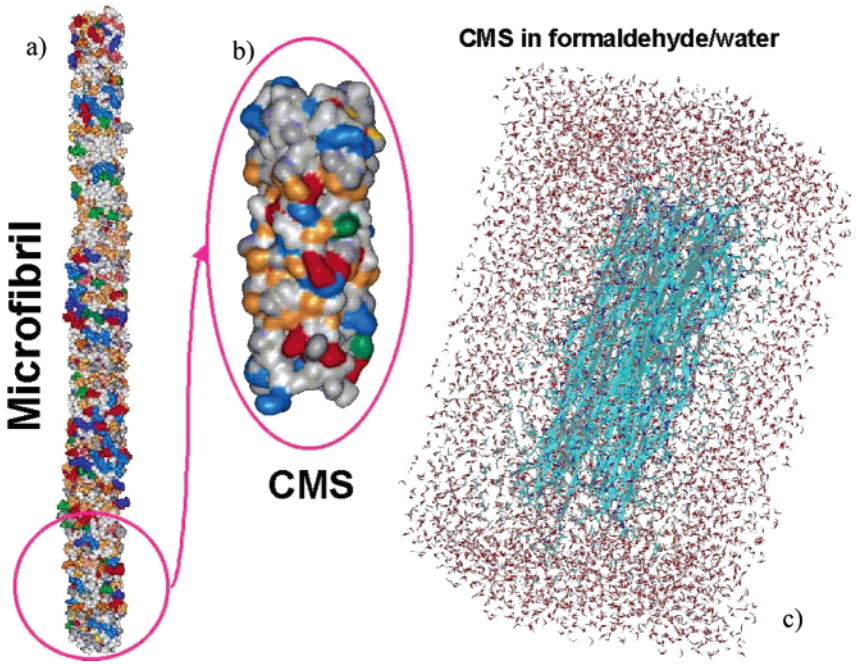


Figure 1. (a) Space-filling representation of a three-dimensional computer model of the bovine type I collagen microfibril based on the Smith description[32] and (b) the extracted segment (CMS) with side chains colored as follows: HPR in gold, ASN and GLN in green, positively charged side chains (ARG, LYS) in red, negatively charged side chains (ASP, GLU) in blue, and all the other side chains in gray-white. (c) CMS in the simulation box filled with formaldehyde and water molecules.

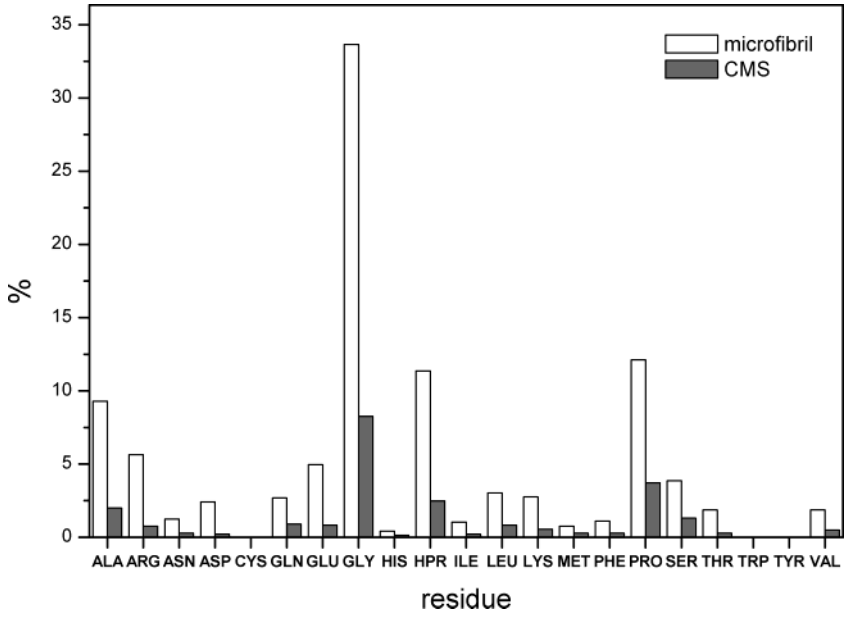


Figure 2. Percentage of amino acid residues present in the microfibril model and in the CMS.

the total system). The second structure corresponds to an acidic solution in which all the residues GLU, ASP, LYS, ARG, and HIS are protonated, thus leading to a total charge of +21e in the system. Due to the missing chains at both ends of the CMS, to maintain the structural integrity of the model without dramatically changing its dynamics harmonic position restraints were applied to the backbone atoms of the residues placed at both ends of the 15 α chains. The side chains of such residues are instead allowed to move freely. The forces applied to the border regions were such that the atoms were not fixed but could reasonably move.

3. Computational Details

MD simulations were carried out to obtain a detailed microscopic description of four systems: (1) CMS in pure formaldehyde (neutral solution), (2) CMS in around 8.0%(v/v) formaldehyde/water (neutral solution), (3) CMS in 1.4%(v/v) gallic acid/water (acidic solution), and (4) CMS in water (neutral solution). The use of the acidic pH in case 3 is due to the choice of simulating the interaction between CMS and the neutral form of gallic acid, which exists only at a low pH value (below 4.26, Martel and Smith[33]). The simulations at neutral pH level were performed without the addition of counterions. This approximation should be reasonable, because the total charge of the model is small compared with the system size and the electrostatic interactions are damped by the presence of many shells of explicit solvent molecules.[34]

Water was described using the TIP3P model.[35] The starting configurations in cases 2 and 3 were constructed by first randomly positioning the modifying agent molecules around the

collagen segment, then surrounding the system with water molecules, which were also incorporated in the "modifying-agent-rich" phase where space was adequate for an insertion.

The +21e charged system (acidic solution) was neutralized by adding 21 Cl^- charge-balancing counterions. The position of the counterions was determined by using a method that explores the electrostatic energy surface near charged residues looking for favorable counterion-residue interactions. In summary, in case 3 the starting configuration consisted of CMS surrounded by gallic acid and water molecules with charge-balancing counterions at the protein surface.

All MD simulations were performed using the AMBER7 package and the Cornell et al.[30] and General AMBER (GAFF)[32] force fields, whose atom types are more general and cover most of the organic chemistry space.

The charges of formaldehyde and gallic acid were determined, consistently with the AMBER ones, with the RESP methodology,[36] by fitting them to the B3LYP/6-31G* electrostatic potential in the space surrounding the molecule with hyperbolic restraints on the non-hydrogen atoms. The electrostatic potential was calculated by using the Merz−Kollman fitting method as implemented in Gaussian03.[37] MD simulations were carried out in the NPT ensemble in a rectangular parallelepiped box by using periodic boundary conditions, a time step of 1 fs, and Ewald summations to handle long-range electrostatic interactions. The pressure was kept at 1 bar with isotropic-based scaling;[38] a coupling constant of 1 ps and a dielectric constant, $\epsilon = 1$, were employed. A 12 Å cutoff was applied to the nonbonded interactions and the nonbonded list was updated every 20 steps. After 5000 steps of energy minimization to remove bad steric contacts, the system was heated under constant volume conditions to 600 K over 10 ps of dynamics, freezing the solute coordinates in order to randomize the positions of the solvent molecules. The system was allowed to equilibrate at this temperature for 10 ps, then the temperature was lowered over 10 ps to 305 K, the constraints were removed, and the system was equilibrated for 10 ps with 100 kcal/mol $Å^2$ position restraints on the backbone atoms of the residues at both ends of the α -chains to avoid border effects. The equilibration continued at constant pressure for 25 ps to bring the artificial box to a reasonable density. The results of MD simulations should, in principle, be independent of the starting structure but the purpose of the equilibration phase is to prepare the system in a form such that nonphysical forces do not cause displacements that are beyond the radius of convergence of the numerical integrators.

To determine if 25 ps was suitable to allow the system to relax, the pressure, volume, and density were monitored. Although it is not shown here, these indicators were easily equilibrated within the time chosen for equilibration. Starting from the last equilibrium configuration, a simulation of around 1 ns was performed in the NPT ensemble for data collection. During the equilibration and production phases Berendsen's temperature bath coupling[39] was used. The SHAKE[40] algorithm was exploited to constrain all hydrogen atom-heavy atom bond lengths to their equilibrium values. The coordinates were saved every 1 ps to analyze the trajectory. Analysis of the various trajectories was carried out by using the carnal, anal, and ptraj modules of AMBER7.

The simulations were run on a Compaq DS20E workstation and on a Digital Alpha XP1000. The analysis of the trajectories was done on SGI O^2 and SGI Indigo2 workstations.

4. Results and Discussion

4.1. Bundle Stability and Structural Features. To compare the relative stability of the models and the dynamic behavior of the helix bundles, the root-mean-square deviation (rmsd) of the C_α atoms from the initial conformation and the root-mean-square fluctuation (rmsf) of the C_α atoms from their average values over the duration of the simulations were calculated. All the reported quantities were computed over the last 800 ps of the simulations leaving the first 200 ps as a further equilibration phase. A picture of the behavior of rmsd as a function of time is reported in Figure 3.

In all simulations rmsd increases over the equilibration phase (data not shown); this can be due to a sizable conformational change. Then, rmsd continues to gradually increase during the production run. The final deviations are in the range 2.0−2.8 Å in all cases except the simulation of CMS in pure water (case 4), whose final deviation is around 3.7 Å. Evidently, major changes in the conformation of collagen occur when the CMS is in water. In fact, the high water activity makes different regions of the conformational space accessible. This is in agreement with the findings of Privalov,[5] who has determined that water molecules are one of the major contributors to the enthalpy of denaturation of collagen. Contrarily, minor changes in the collagen structure occur in the other cases, at least on the time scale sampled here. These findings are again in agreement with experiments on the direct determination of the influence of chemical modification on collagen molecular stability.[10] In fact, it has been shown that even the alkylation of the collagen amino groups in the presence of $NaBH_4$ (at least performed with mild conditions) has little or no influence on collagen at the level of triple helical stability.[10]

Comparing the two simulations of collagen with formaldehyde (cases 1 and 2 − Figure 3), a larger deviation from the starting conformation is noted when CMS is surrounded by an aqueous solution of formaldehyde (case 2). This is not surprising if the behavior in water (case 4) is considered.

The rmsd in case 3 (aqueous solution of gallic acid) is less than the final value shown in case 1 (pure formaldehyde). This slight modification of the CMS structure as a consequence of the interaction with gallic acid is in accordance with experimental findings by Tang et al.[2] obtained by differential scanning calorimetry (DSC), showing that the interaction with polyphenols increases the hydrothermal stability of collagen.

Another indicator of the structural flexibility of the protein is the rmsf. In Figure 4 the trend of the rmsf of the C_α of the five triple helical segments for CMS in the various environments here studied is shown. The minimum values correspond to the restrained residues placed at both ends of the CMS. The introduction of position restraints on the backbone atoms of these residues seems reasonable if we consider that CMS is just a small portion of a longer microfibril segment embedded in a pure or mixed "solvent". Due to the missing parts, the helix termini interact with a greater number of solvent molecules. Thus, they can exhibit greater fluctuations which might give rise to artifacts and unrealistic effects.

Examination of rmsf (Figure 4) suggests that in all cases the greatest fluctuations are at the residues on the helical side exposed to the solvent. This is not surprising, because such residues have a large number of possible interactions with modifying agent and water molecules. The fluctuations seem to be more marked in case 1 (pure formaldehyde). On the other hand, rmsf values for the core residues are predominantly less than 1 Å, thus indicating a general stability of CMS in all cases.

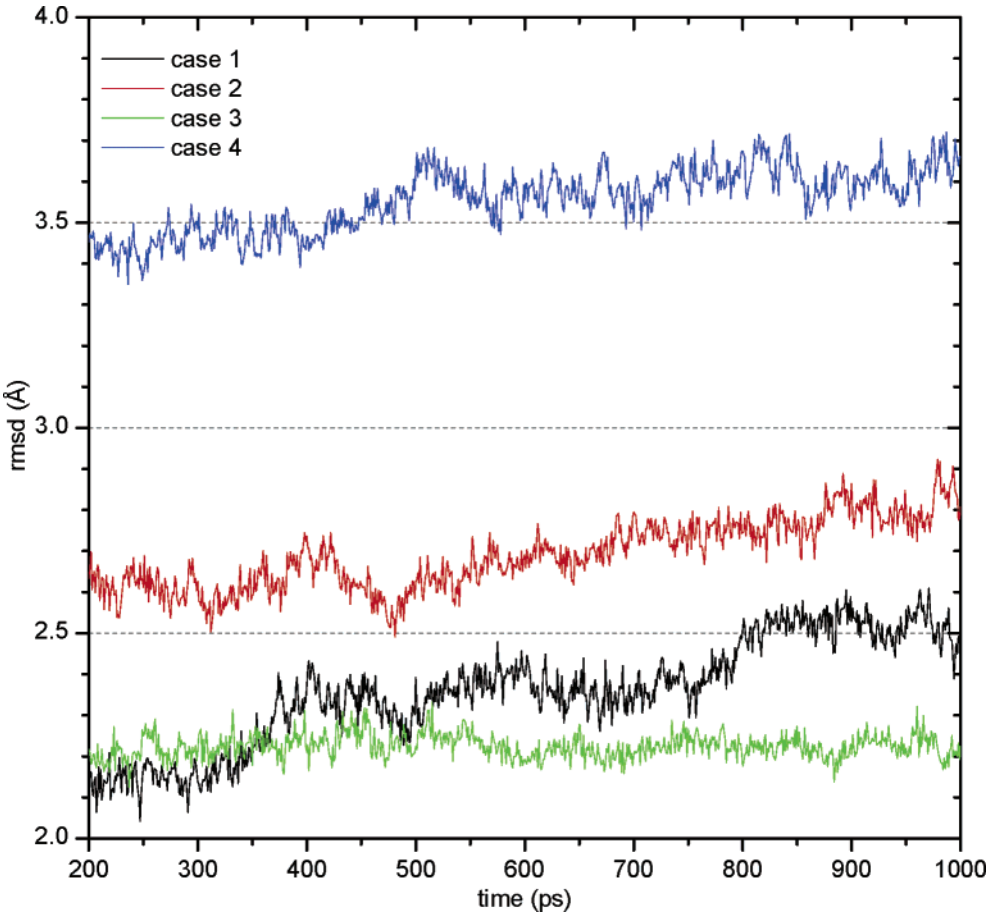


Figure 3. Root-mean-square deviations (rmsds) of the C_α-carbons from the starting structure in each simulation as a function of time.

These findings, if combined with the behavior of the C_α rmsds, show a different stability of the bundle in case 1 with respect to the other cases. In fact, the bundle in formaldehyde appears to "expand" somewhat as time progresses while the structure in case 3 (aqueous solution of gallic acid) is the most stable over the time of the simulation.

4.2. Bundle-Solvent Interactions. This section is focused on the evaluation of the distribution of the "solvent" (water + modifying agent) species around the CMS. The radial distribution functions (rdfs) here reported are average values calculated over the trajectory and the rdf plots for the entire CMS or combined groups of CMS atoms are calculated for each of the following solvent atom species: water oxygen, formaldehyde oxygen, and gallic acid oxygen united and individual atoms. The rdfs are plotted on a probabilistic scale calculated by normalizing the densities relative to the bulk density. The CMS-solvent hydrogen bonding interactions that involve mainly the molecules of the first solvation shell are explored. To keep track of the potential hydrogen bond interactions a cutoff distance of 3.3 Å between donor and acceptor atoms and a H-donor−acceptor cutoff angle of 30 degrees have been specified.

To give the overall picture of the hydrogen bonding network between collagen residues and modifying agent molecules, the fraction of time each residue spends forming an H-bond with modifying agent molecules is also shown. This fraction f_H[41] is calculated as the ratio between the total number of time steps a given residue forms an H-bond (n_H), the product between the total number (n_T) of time steps in the simulation, and the number (q) of observed types of H-bond formed by the residue under examination during the simulation; thus

$$f_H = \frac{n_H}{n_T q} \tag{1}$$

After discussing each simulation individually, a comparison among the various models is presented.

4.2.1. Case 1. CMS in Pure Formaldehyde. The CMS-formaldehyde oxygen radial distribution function (Figure 5-1) obtained from the simulation shows a peak for the first solvation shell at a distance of about 2.0 Å and a second peak centered around 2.5 Å. Integration of the first peak up to the first minimum at approximately 2.32 Å gives an average coordination number of 146 formaldehyde molecules while the coordination number corresponding to a distance of 4.0 Å is about 567 (Table 1). In Figure 6-1, the partial rdfs (prdfs) of the formaldehyde oxygen with various atoms or groups of atoms belonging to the side chains of the following AA residues are shown: ARG[N(NH),N(NH_2)], ASP[O(COO^-)], ASN[N(NH_2)], GLU[O(COO^-)], GLN[N(NH_2)], HIS[N(NH),N], HPR[O(OH)], LYS[N(NH_3)$^+$], SER[O(OH)], THR[O(OH)]. The prdfs of ARG, ASN, GLN, HPR, LYS, SER, and THR show a distinct first solvation shell peak at a distance in the range 2.8−3.0 Å and coordination numbers of 66.4, 0.7, 16.5, 14.3, 16.7, 3.0, and 1.9 respectively (see Table 2 for a summary). The prdfs of ASN, HPR, SER, and THR show a second peak at a distance in the range 3.9−4.7 Å while the prdfs of ASP, GLU, and HIS exhibit only one peak at about 4.2 Å.

The pattern of H-bond formation between the CMS residues and formaldehyde molecules (Figure 7-1) has been analyzed in detail by considering f_H, percentage of occupancy (percentage of structures exhibiting the particular type of H-bond), and the

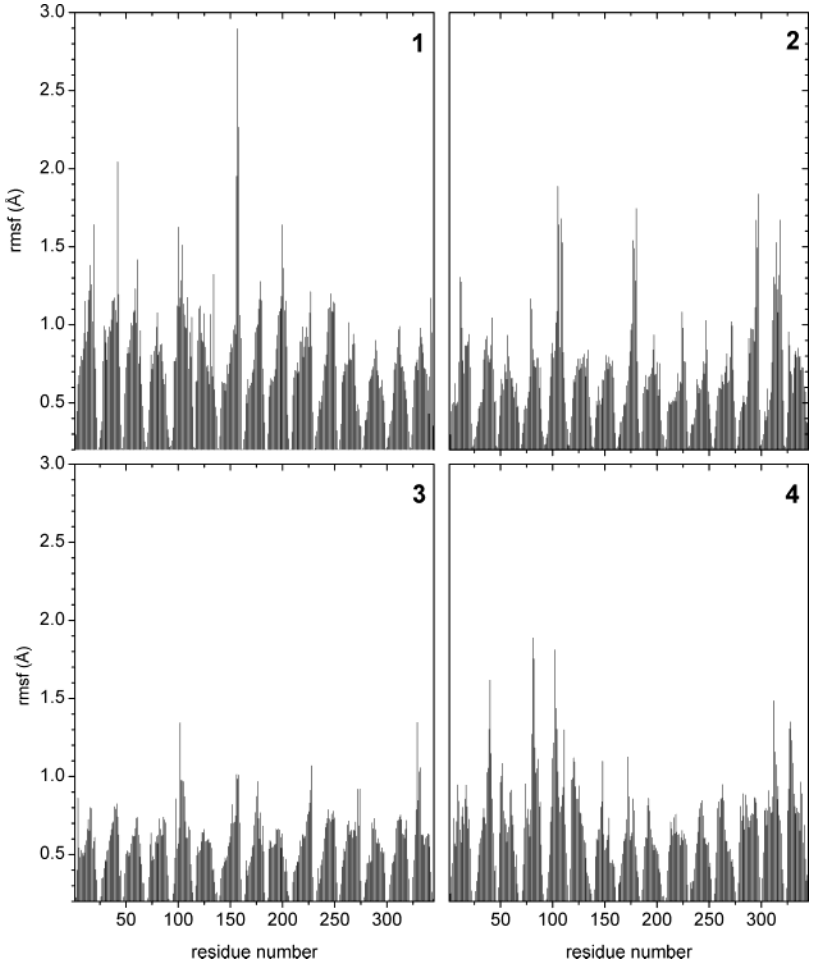


Figure 4. Residue by residue C_α-rms fluctuations about their average coordinates for: **case (1)** CMS in formaldehyde; **case (2)** CMS in formaldehyde/water; **case (3)** CMS in gallic acid/water; and **case (4)** CMS in water.

TABLE 1: Total Coordination Number for Water and Modifying Agents in the Various Cases under Examination. See Text for the Integration Ranges

		case 2		case 3		
shell	case 1	H_2CO	water	gallic acid	water	case 4
first	146	10	169	10	175	209
second	567	89	1021	213	908	1310

TABLE 2: Coordination Number for the Various Residues in the Different Cases Under Examination. See Text for the Integration Ranges

residue	case 1	case 2	case 3 (OHp)	case 3 (OHc)	case 3 (O)
ARG	66.4	8.2	4.3	2.9	3.8
ASN	0.7	0.4	0.8[a]	0.3	1.1
ASP	19.5[a]	1.2[a]	0.2	0.1[a]	0.0
GLN	16.5	0.7	10.0	0.9	0.8
GLU	54.0[a]	1.3[a]	4.9	4.5[a]	2.8
HIS	10.6[a]	1.0[a]	1.0[a]	0.0	0.0[a]
HPR	14.3	26.0[a]	15.4	0.9	1.4
LYS	16.7	1.8	1.4	1.4[a]	0.6
SER	3.0	2.8	6.4	1.2	1.3
THR	1.9	2.0[a]	0.6	0.1[a]	0.1[a]

[a] Second coordination shell ($r \geq 4$).

average distance and angle of the atoms involved. A consistent pattern can be seen for the entire sequence: 59 AA have f_H <0.1, 25 AA have f_H in the range 0.1−0.2, 18 AA have f_H in the range 0.2−0.4, and 14 AA have f_H in the range 0.4−0.9. Among the residues with a greater f_H value, 8 AA are placed near the helix termini, their H-bonds occur predominantly from backbone atoms and their percentage of occupancy (the percentage of structures where the H-bond is present with respect to the total sampled conformations) is greater than 40%.

All the H-bonds have an average distance lower than 3.03 Å with a maximum standard deviation of 0.18 Å, and an average H-donor−acceptor angle lower than 22.7 degrees with a maximum standard deviation of 7.6 degrees. The total number of AA involved in hydrogen bonding interaction is 100, which is about 29% of the whole AA sequence; 33 AA form H-bonds to formaldehyde with their side chains, 63 AA form H-bonds with their NH donor backbone atoms whereas 22 interact with formaldehyde via both side chain and backbone atoms. As the total number of accessible AA that might be involved in H-bonding interactions via both backbone and side chain atoms is 75, in this case about 29% of them (22 over 75) are "active". All ARG and LYS AA interact via both backbone and side-chain atoms.

4.2.2. Case 2. CMS in 8% (v/v) Formaldehyde/Water. The CMS-formaldehyde oxygen radial distribution function (Figure 5-2) shows a first peak at 2.1 Å for the first solvation shell and a second peak at 2.6 Å with coordination numbers of 10 and 89 respectively (Table 1). Due to the presence of water molecules the CMS-water oxygen radial distribution function is also plotted (Figure 8). It shows a distinct first solvation shell peak at 1.94 Å and a second peak at 2.7 Å with coordination numbers of about 169 and 1021, respectively (Table 1). The prdfs of the formaldehyde oxygen with the atoms or groups of atoms belonging to the side chains of the residues considered in case 1 are reported in Figure 6-2. The prdfs of ARG, ASN, GLN, and LYS show a peak corresponding to the first solvation

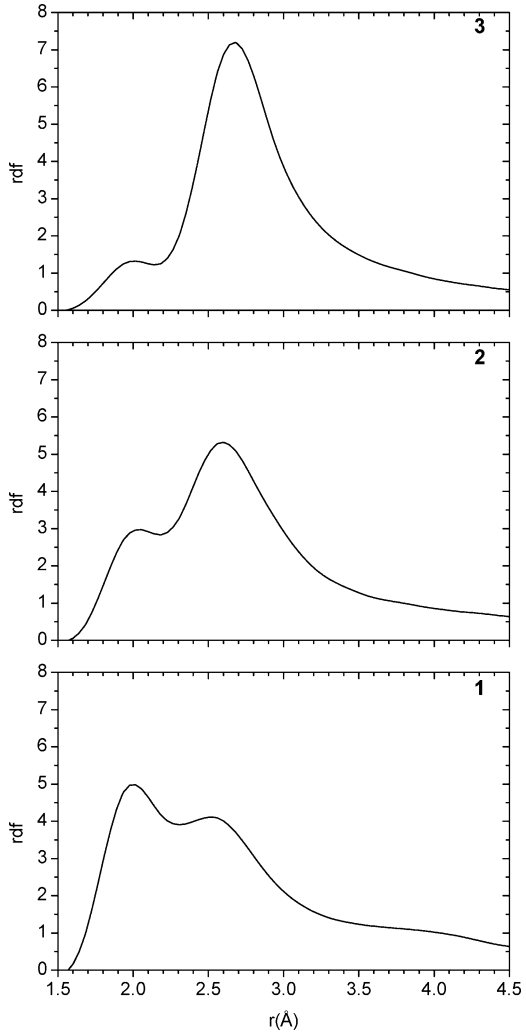


Figure 5. Rdfs for the entire solute (CMS) with formaldehyde oxygen atoms (bottom: case 1, medium: case 2) and with the entire group of gallic acid oxygen atoms (top: case 3).

shell at a distance in the range 2.8−3.1 Å and coordination numbers of 8.2, 0.4, 0.7, and 2.0 respectively (Table 2). Among these prdfs only ASN and GLN have a second large and a lower peak at about 4.4 Å and 3.6 Å, respectively. The prdfs of ASP, GLU, HIS, SER, and THR exhibit large and not well-defined peaks at a distance in the range 4.1−5.2 Å.

A reduced H-bonding pattern with respect to case 1 can be seen for the entire sequence (Figure 7-2). There are 78 AA with f_H lower than 0.1, 6 AA interact with formaldehyde molecules with f_H in the range 0.1−0.2, and 4 AA have $f_H > 0.2$. The 10 AA whose f_H value is greater than 0.1 interact predominantly via their backbone atoms (NH donors) and their percentage of occupancy is greater than 12. The total number of hydrogen bonded residues is 87, which means 25% of the whole AA sequence: 30 AA are H-bonded to formaldehyde oxygen with their side chains, 42 AA interact via their NH donor backbone atoms, whereas 15 AA interact with both side chain and backbone atoms. Considering that 75 is the total number of accessible AA that might be involved in both backbone and side chain H-bonding interactions, in this case 20% of them form H-bonds.

4.2.3. Case 3. CMS in 1.4% (v/v) Gallic Acid/Water. The CMS−gallic acid rdf (Figure 5-3) shows a small peak at 2.0 Å and another peak at 2.7 Å with coordination numbers of about 10 and 213, respectively. The CMS−water oxygen rdf (Figure 8) shows the first peak at 1.94 Å for the first hydration shell and the second peak at 2.7 Å, with coordination numbers of 175 and 908, respectively (Table 1). The rdfs of the hydroxylic oxygen atoms (OHp), carbonylic oxygen atoms (O), and carboxylic oxygen atoms (OHc) with various atoms and groups of atoms of the side chains of the AA previously mentioned, but considering their actual protonation state, are shown in Figure 9. The prdfs of OHp with ARG, ASP, GLU, GLN, HPR, LYS, SER, and THR exhibit a first peak at a distance in the range 2.9−3.3 Å and coordination numbers of 4.3, 0.2, 4.9, 10.0, 15.4, 1.4, 6.4, and 0.6 respectively (Table 2). The prdfs of OHp with ASN shows a first peak at about 4.2 Å with a coordination number of 0.8 and a second peak for a distance greater than 5.5 Å.

The OHp−HIS prdf shows only one peak at 4.3 Å with a coordination number of 1. The prdfs of OHp with ARG, ASP, and SER display a second peak at a distance in the range 4.6−4.8 Å. OHc−ARG, OHc−ASN, OHc−GLN, OHc−HPR, and OHc−SER prdfs exhibit a first peak at a distance in the range 2.8−3.7 Å with coordination numbers of 2.9, 0.3, 0.9, 0.9, and 1.2 respectively (Table 2). OHc−ASP, OHc−GLU, and OHc−THR prdfs do not have well-defined peaks; the most pronounced is at about 4.8 Å. The OHc−LYS prdf has only one peak at 4.6 Å with a coordination number of 1.4. The prdfs of OHc with ARG, ASN, and GLN display a second peak at a distance in the range 4.8−5.0 Å. O−ARG, O−ASN, O−GLN, and O−LYS prdfs show a first sharp peak at a distance in the range 2.90−2.97 Å with coordination numbers of 3.8, 1.1, 0.8, and 0.6 respectively (Table 2).

O−GLU and O−HPR prdfs exhibit a large and low peak at about 3.1−3.2 Å with coordination numbers of 2.8 and 1.4, respectively. The O−SER prdf shows a peak at 3.1 Å with a coordination number of 1.3.

The fraction of H-bonding f_H per AA is reported in Figure 7-3. A considerable pattern of H-bonding to gallic acid molecules with high f_H values is shown for the entire sequence. There are 37 AA with $f_H < 0.1$, 13 AA with f_H in the range 0.1−0.2, and 23 AA have f_H values in the range 0.4−1.0. There are 33% of the AA with a greater f_H value that consist of GLY AA that possess the highest average percentage of occupancy (about 79%) and form hydrogen bonds with gallic acid OHps. The remaining residues are HPR, SER, and GLN; their average percentage of occupancy is lower and their side chains and backbone atoms are both involved in H-bonding interactions with gallic acid OHps. The total number of H-bonded AA to gallic acid molecules is 105; this means the 30% of the whole AA sequence. There are 57 AA that interact with gallic acid oxygen atoms with their backbone atoms; 20 AA are H-bonded with their side chains atoms whereas 28 AA interact via both backbone and side-chains atoms. This last number is about 37% of the accessible residues whose side chains and backbone atoms might participate in H-bonds.

4.3. Discussion. The distributions of the water molecules in the mixed solution (cases 2 and 3) and that in the pure water simulation (case 4) are very similar (Figure 8). The main difference is due to the presence of a larger number of water molecules in the first hydration shell when only water is present. The vast majority of water molecules are concentrated in close proximity to the relatively polar surface of CMS and a very low water density is found for distances greater than 4 Å.

The formaldehyde and gallic acid distributions for cases 2 and 3 indicate that both molecules are present in the primary solvation layer of the CMS and participate in direct interactions with its residues. The first peaks of the formaldehyde rdfs are

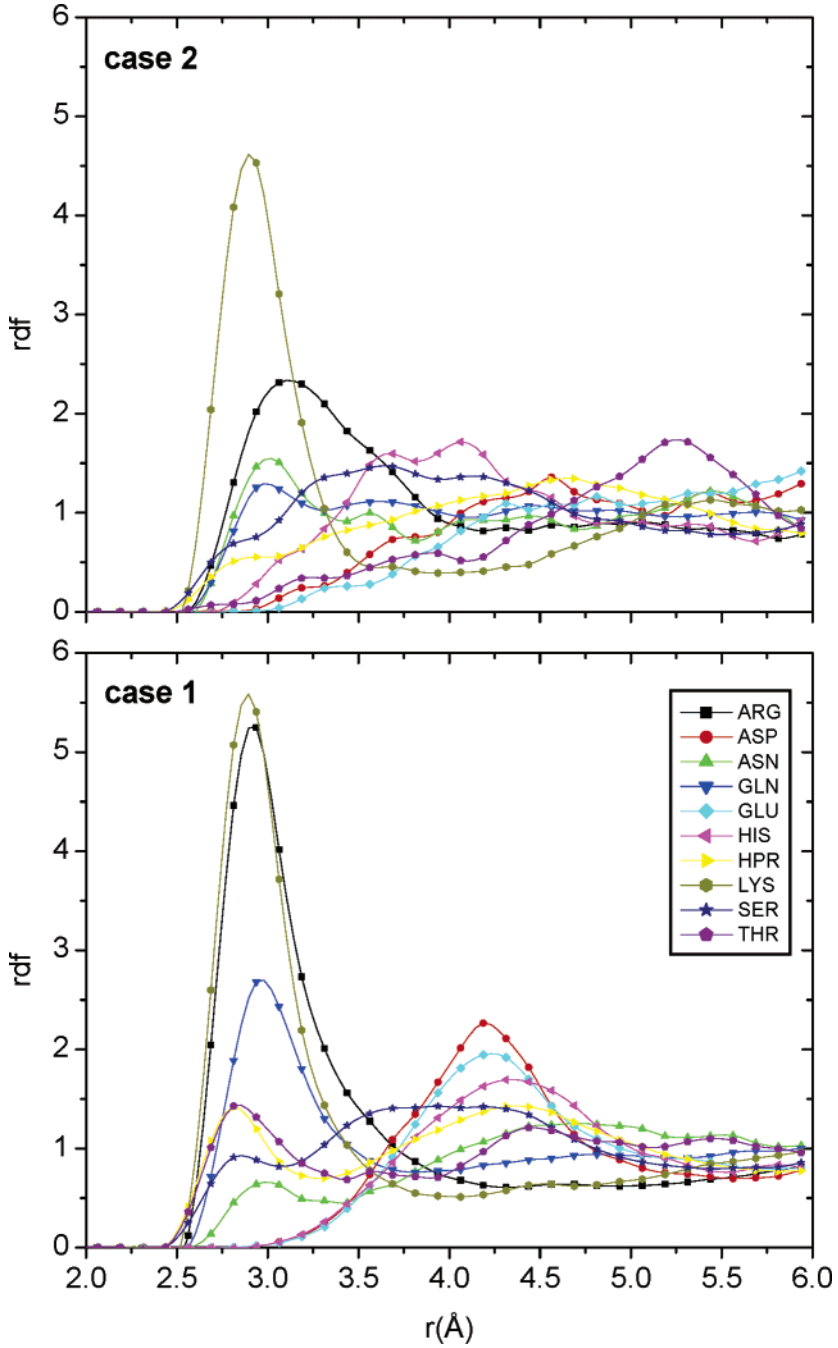


Figure 6. Prdfs of formaldehyde oxygen atoms with various atoms or groups of atoms belonging to the side chains of the reported residues (see text).

at 2.0 and 2.1 Å for cases 1 (pure H_2CO) and 2 (aqueous solution of H_2CO), respectively, which is somewhat larger than the distance found for the water peak. The second peak is at a distance of 2.5 and 2.6 Å for case 1 and 2, respectively, which is shorter than the distance of the water second peak. Another difference can be seen in the peak heights. In case 1, the first peak height is roughly 5 times its bulk value whereas the second peak is about 4 times its bulk value. On the contrary, in case 2 the first peak is about 3 times its bulk value and the second peak is greater than 5 times bulk. This different behavior is due to the presence of water molecules in the first solvation shell that interact with formaldehyde molecules and with CMS residues.

In case 3 (aqueous solution of gallic acid) the height of the first peak of gallic acid is about 1.3 times its bulk value and the water peak is about 4 times the bulk. On the contrary, the gallic acid second peak is more than 7 times its bulk value whereas the water second peak is less than 5 times the bulk. This suggests that a larger number of gallic acid molecules can be found at a larger distance from the CMS surface.

The prdfs and coordination numbers together with the tight H-bond analysis can be useful to illustrate the hydrogen bonding behavior of the formaldehyde carbonylic oxygen with AA side chain functional groups and to explain changes in the structure of the solution when the concentration of formaldehyde is reduced from 100% to 8% (v/v). Our simulations indicate that ARG, LYS, ASN, and GLN AA (i.e. residues having amino functional groups in their side chains) interact more favorably with formaldehyde than other AA with hydroxylic or carboxylic functional groups in their side chains such as HPR, SER, THR, ASP, and GLU. This behavior is seen in both simulations 1 and 2, however in the mixed solution (case 2) both the

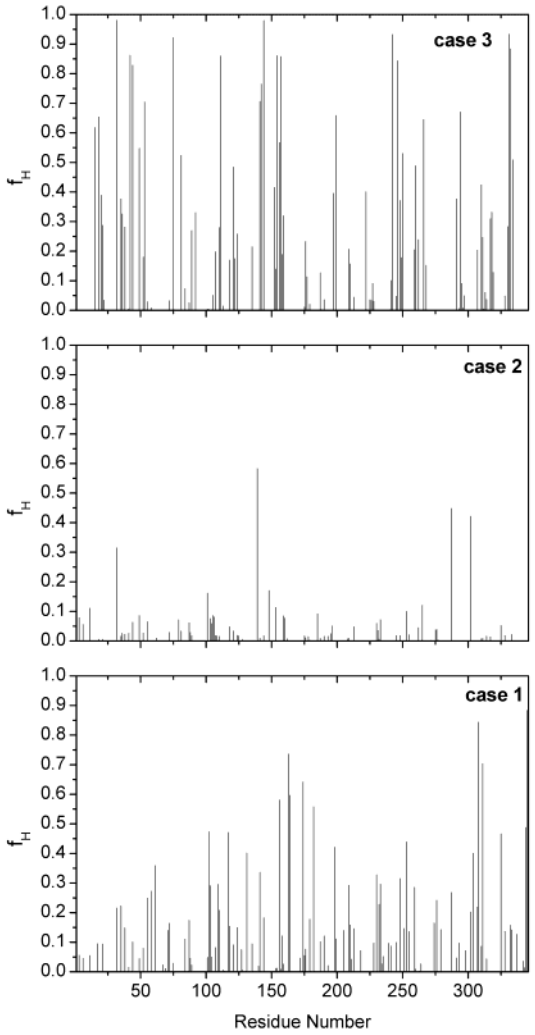


Figure 7. Fraction of H-bonding, f_H, per residue for: **case (1)** CMS in formaldehyde; **case (2)** CMS in formaldehyde/water; **case (3)** CMS in gallic acid/water.

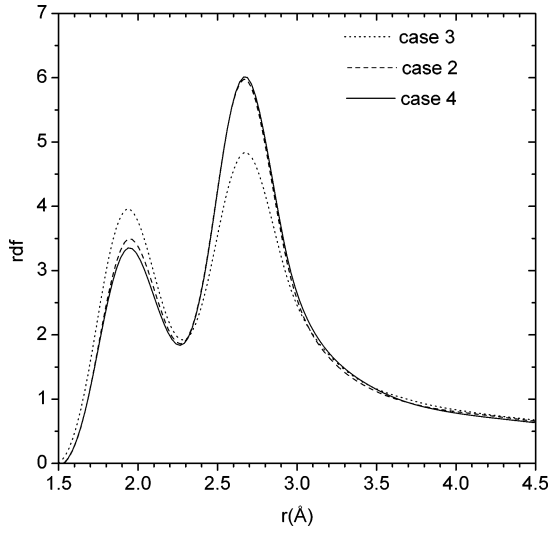


Figure 8. Rdfs for the entire solute (CMS) with water oxygen atoms for: CMS in formaldehyde/water (case2: dashed line), CMS in gallic acid/water (case 3: dotted line), and CMS in water (case 4: solid line).

interaction and the number of coordinated molecules is reduced due to strong competition with water molecules.

These results explain the fact that aldehydic compounds are used for enhancing collagen chemical and hydrothermal stability.

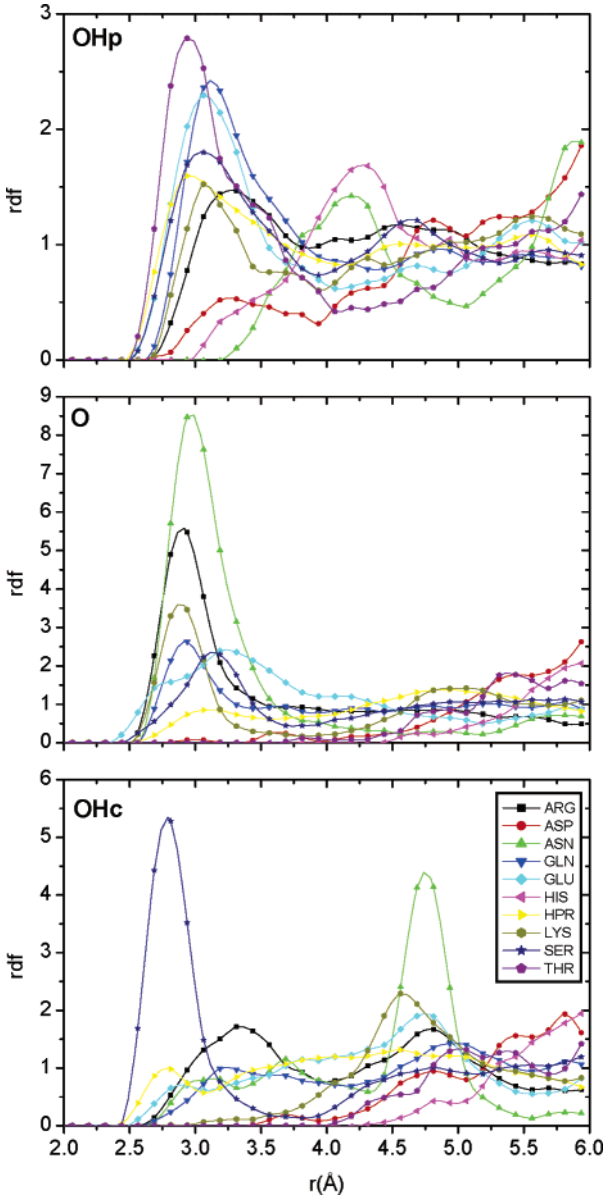


Figure 9. Prdfs of gallic acid phenolic oxygen atoms (OHp, top), gallic acid carbonylic oxygen atoms (O, medium), and gallic acid carboxylic oxygen atoms (OHc, bottom) with various atoms or groups of atoms belonging to the side chains of the reported residues (see text).

In fact, aldehydes probably cause cross-linking of the collagen fibers as a result of covalent bonding with the $-NH_2-$ and $-NH-$ groups of collagen AA side chains.[1,6] Unfortunately, our approach is not able to simulate covalent binding mechanisms and the only possible conclusions are the ones just drawn.

In contrast to formaldehyde molecules, which can only act as hydrogen bond acceptors, gallic acid molecules are rich in oxygen atoms and exhibit both proton-donor and proton-acceptor properties. All the oxygen atoms are involved in hydrogen bonding interactions with a greater number of CMS AAs (with respect to case 2) and a high percentage of these H-bonds may last for a long time in the simulation. The competition with water molecules is strongly reduced. Gallic acid molecules interact, in fact, with greater surface areas of the CMS and render them inaccessible to water molecules. It is difficult to establish a clear correlation between the H-bonding pattern and the stability of the helix bundle, but it seems that the H-bond network may be important for reducing its flexibility. However, it has been proven that the interaction of polyphenols and

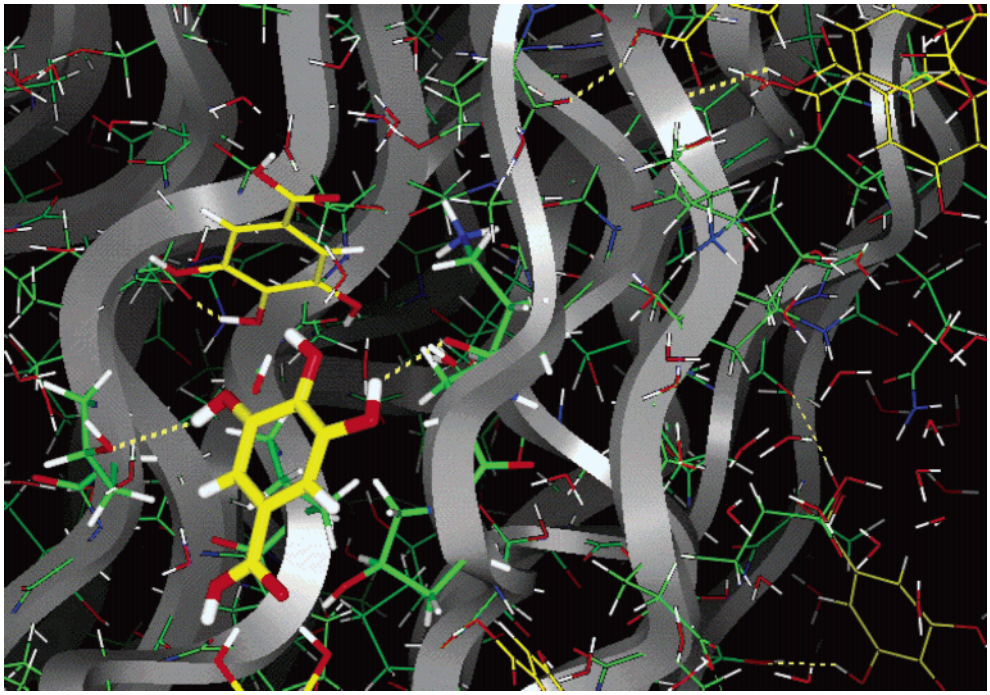


Figure 10. Simulation snapshot showing gallic acid molecules (carbon in yellow, oxygen in red and hydrogen in white sticks) H-bonded to AA side chains and backbone atoms (gray solid ribbon) of the CMS (carbon in green, oxygen in red, nitrogen in blue, hydrogen in white).

collagen fibers is a synergistic effect of hydrophobic association and hydrogen bonding. But, the reaction is a reversible process and polyphenols bind in a relatively weak way to each individual binding site.[6] The data reported in Table 2, suggest that the majority of gallic acid molecules point their phenolic groups toward CMS residues, and a visual inspection of the sampled structures confirms this tendency (Figure 10). The same data are also in agreement with previous computational studies on the interaction between gallic acid and a small collagen-like peptide,[11−14] which evidenced the formation of hydrogen bonds between the phenolic OH groups of gallic acid and the hydroxyl groups of the model collagen-like peptide.

To end this section, it is worth noticing that, at least on the time scale here sampled, there is no evidence of penetration of water, formaldehyde, or gallic acid molecules inside the protein bundle.

4.4. Bundle-Modifying Agent Interaction Energies. The role played by modifying molecules in stabilizing the model system structure is examined in more detail in this section by analyzing the interaction energy components between CMS and formaldehyde/gallic acid, to identify which residues have the largest relative contributions.

Energetic analysis of the sampled structures has been done by decomposing the total energy among the different groups of atoms, AA, and modifying agent molecules, to find the interaction energies between the various parts of the system and to identify the relative contributions of the CMS polar residues.

Calculation of the CMS-modifying agent interaction energies was carried out by using a two-term potential function (van der Waals and electrostatic) with geometric and energy parameters derived from the force field of Cornell et al.[31] The choice of the force field is crucial to obtain a good and reliable picture of the interaction. AMBER force field performance has been validated in DNA base pair interactions;[42] previous studies of ours,[43,44] as well as other authors,[42] showed a satisfactory behavior of this force field in the evaluation of the stability of either H-bonded or van der Waals complexes.

The total average interaction energy and its van der Waals and electrostatic components are shown in Figure 11 and in Figure 12 as a function of the residue number.

A distance dependent dielectric function $\epsilon = 4r$ was used to simulate the polarization effect in attractive interactions and to provide solvent screening effects. Thus, closer interactions have a higher weight and long-range Coulomb interactions are damped more than short-range ones. The restrained residues were excluded from the calculations. The interaction energies were evaluated for the "solvent" (water and modifying agent) molecules within a radius of 6 Å with respect to CMS, and in order to compare the two simulations with formaldehyde (cases 1 and 2) a normalization factor was applied.

Case 1: Pure Formaldehyde. CMS residues having a high interaction energy ($E^{TOT}_{int} > 4$ kcal/mol), are 3.5% of the total residues: 60% are ARG, 30% ASP, 10% LYS, and the main component of the interaction is the electrostatic term. CMS AA having medium-high interaction energies ($2 < E^{TOT}_{int} < 4$ kcal/mol) are 16.8% of the total: 19% of them are HPR, 17% PRO, 15% GLN, 10% GLU, 7% LYS, and the main contribution to the interaction energy is the van der Waals term.

Case 2: 8% (v/v) Formaldehyde/Water. CMS residues having $E^{TOT}_{int} > 4$ kcal/ mol represent 4.9% of the total number of residues considered: 36% of them are ARG, 14% PRO, 14% GLN, and the van der Waals contribution is greater than the electrostatic term. CMS AA with $2 < E^{TOT}_{int} < 4$ kcal/mol are 20.0% of the total (26% PRO, 26% HPR, 9% ALA, 5% LYS, and 5% GLY). In this case the interaction energy mainly consists of van der Waals contributions.

Case 3: 1.4% (v/v) Gallic Acid/Water. CMS residues having $E^{TOT}_{int} > 4$ kcal/mol represent 16.5% of the total (26% PRO, 19% HPR, 11% ARG, 9% GLN, 7% LYS, 4% SER, and 4% GLU). CMS AA with $2 < E^{TOT}_{int} < 4$ kcal/mol are 18.9% of the total (22% HPR, 33% GLY, 13% PRO, 9% SER, 6% GLN, 6% GLU, and 6% ALA). The main stabilizing component of the interaction energy is the van der Waals term, whereas the electrostatic contribution is the largest term for SER residues.

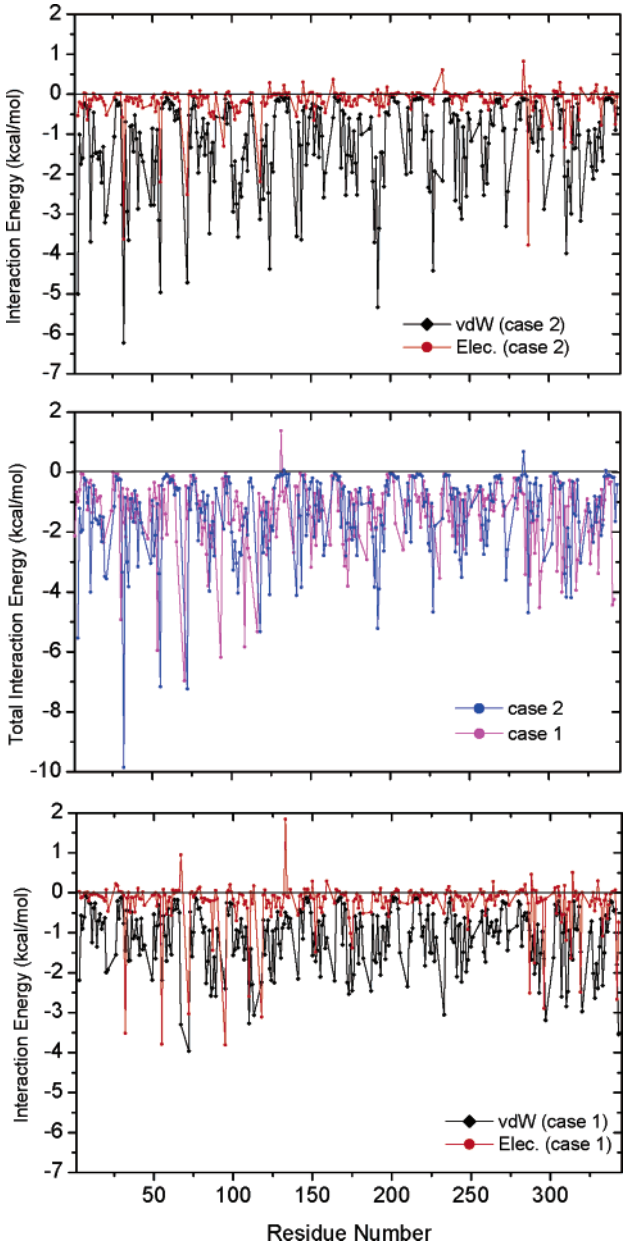


Figure 11. Total average interaction energies and their van der Waals and electrostatic components between formaldehyde molecules contained within a radius of 6 Å with respect to CMS and each CMS residue (helix termini excluded).

Discussion. The results indicate that the presence of ARG, LYS, and GLN residues is crucially important for a strong interaction between formaldehyde and CMS. This interaction has an essentially electrostatic nature even if the Coulomb term is reduced in case 2 due to the presence of water molecules; the penetration of organic compounds into polar contacts solvated by water must, in fact, involve competition with water molecules. A weaker but favorable interaction with HPR and PRO residues is also evident.

With regard to gallic acid, our results are in agreement with early studies of polyphenol/protein binding suggesting that polyphenols bind preferentially to PRO[7] and HPR residues. Baxter et al.[8] have shown that PROs perform dual functions: they act as binding sites and they keep the peptide extended, thus maximizing the available binding surface. NMR spectroscopy experiments have provided further evidence that the dominant mode of interactions between the PRO-rich peptides and polyphenol molecules is a hydrophobic association although

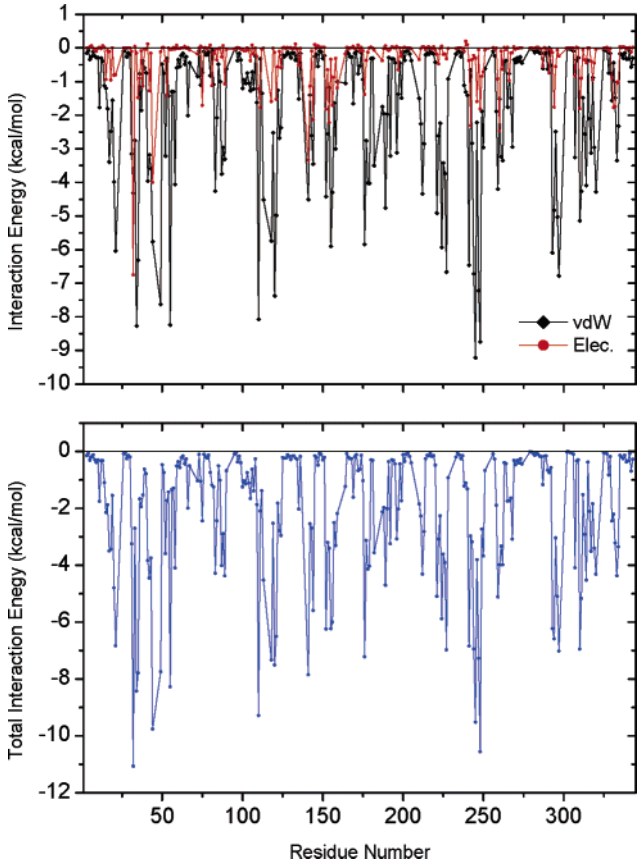


Figure 12. Total average interaction energies and their van der Waals and electrostatic components between gallic acid molecules contained within a radius of 6 Å with respect to CMS and each CMS residue (helix termini excluded).

hydrogen bonding may play a secondary role. However, PRO and HPR are certainly not the only possible binding sites. Strong interactions also occur with ARG AA side chains and these findings are supported by Nuclear Overhauser Effect (NOE) measurements.[9]

5. Summary and Conclusions

The simulations presented in this paper are the first detailed MD studies of structural and binding properties of a collagen microfibril segment (CMS) surrounded by pure and mixed solvents. Our findings show that major changes in the conformation of the CMS occur in water whereas in pure formaldehyde and in the mixed solutions the deviation from the starting structure is smaller. The greatest fluctuations are at the residues located on the helical side exposed to the solvent; analysis of the core residues' fluctuations indicates a general stability for all the models. Both formaldehyde and gallic acid molecules are present in the primary solvation layer of the CMS and participate in direct interactions with its residues.

ARG, LYS, ASN, and GLN AA interact more favorably with formaldehyde molecules than other residues and the interaction is essentially of an electrostatic nature. All the oxygen atoms of gallic acid molecules are involved in hydrogen bonding interactions with CMS residues and point their phenolic groups toward CMS residues. Preferred binding to PRO and HPR residues is evidenced. Both van der Waals and electrostatic contributions are relevant to the interaction. There is no evidence, at least on the time scale here sampled, of penetration of water, formaldehyde, or gallic acid molecules inside the protein bundle. Our computational results are in satisfactory

agreement with experimental data found in the literature and demonstrate that MD simulations can provide a realistic description of the CMS in different environments.

It is also worth noting that the detailed information derived from the present theoretical approach can provide a new significant help to the experimental study and application of collagen-related materials. Indeed, it certainly provides a better understanding of the effect of chemicals on a very complex natural macromolecular system and can be of help in designing theoretically driven experiments. These last are certainly needed for the present and future use (and for the treatment) of collagen. In particular, the leather industry could obtain, in a rapid and economic way, indications for a more efficient and environmentally suitable procedure for hide tanning. Also, collagen is a fundamental component of the human body and its proper and safe chemical "modification" can provide improved ways to develop biomaterials for medical applications.

Acknowledgment. The authors thank Prof. E. M. Brown for allowing the use of her collagen structure model. S.M. is grateful to James W. Caldwell (UCSF) for granting her the use of the AMBER 7 package. The authors also thank Prof. Francesco Ciardelli for useful discussion and revising of the manuscript.

References and Notes

(1) Covington, A. D. *Chem. Soc. Rev.* **1997**, *26*, 111−126.

(2) (a) Tang, H. R.; Covington, A. D.; Hancock, R. A. *Biopolymers* **2003**, *70*, 403−413. (b) Tang, H. R.; Covington, A. D.; Hancock, R. A. *J. Agric. Food Chem.* **2003**, *51*, 6652−6656.

(3) Gomes, C. A.; Girão da Cruz, T.; Andrade, J. L.; Milhazes, N.; Borges, F.; Marques, M. P. M. *J. Med. Chem.* **2003**, *46*, 5395−5401.

(4) Edelmann, A.; Lendl, B. *J. Am. Chem. Soc.* **2002**, *124*, 14741−14747.

(5) Privalov, P. L. *Adv. Protein Chem.* **1982**, *35*, 1−104.

(6) pin Liao, X.; bing Lu, Z.; Shi, B. *Ind. Eng. Chem. Res.* **2003**, *42*, 3397−3402.

(7) Hagerman, A. E.; Butler, L. G. *J. Biol. Chem.* **1981**, *256*, 4494−4497.

(8) Baxter, N. J.; Lilley, T. H.; Haslam, E.; Davies, A. P.; Williamson, M. P. *Biochemistry* **1997**, *36*, 5566−5577.

(9) Charlton, A. J.; Baxter, N. J.; Khan, M. L.; Moir, A. J. G.; Haslam, E.; Davies, A. P.; Williamson, M. P. *J. Agric. Food Chem.* **2002**, *50*, 1593−1601.

(10) Giudici, C.; Viola, M.; Tira, M. E.; Forlino, A.; Tenni, R. *FEBS Lett.* **2003**, *547*, 170−176.

(11) Madhan, B.; Dhathathreyan, A.; Subramanian, V.; Ramasami, T. *Proc. Indian Acad. Sci. (Chem. Sci.)* **2003**, *115*, 751−766.

(12) (a) Madhan, B.; Fathima, N. N.; Raghava Rao, J.; Subramanian, V.; Nair, B. U.; Ramasami, T. *J. Am. Leather Chem. Assoc.* **2003**, *98*, 263−272. (b) Madhan, B.; Subramanian, V.; Raghava Rao, J.; Nair, B. U.; Ramasami, T. *J. Am. Leather Chem. Assoc.* **2003**, *98*, 273−278.

(13) Madhan, B.; Thanikaivelan, P.; Subramanian, V.; Raghava Rao, J.; Nair, B. U.; Ramasami, T. *Chem. Phys. Lett.* **2001**, *346*, 334−340.

(14) Madhan, B.; Parthasarathi, R.; Subramanian, V.; Raghava Rao, J.; Nair, B. U.; Ramasami, T. *Chem. Phys. Lett.* **2003**, *369*, 131−138.

(15) Madhan, B.; Muralidharan, C.; Jayakumar, R. *Biomaterials* **2002**,

(16) Arqués, D. G.; Fallot, J.-P.; Michel, C. J. *Int. J. Biol. Macromol.* **1996**, *19*, 131−138.

(17) (a)Wess, T. J.; Hammersley, A. P.; Wess, L.; Miller, A. *J. Struct. Biol.* **1998**, *122*, 92−100. (b)Wess, T. J.; Hammersley, A. P.; Wess, L.; Miller, A. *J. Mol. Biol.* **1998**, *275*, 255−267.

(18) (a) Ottani, V.; Martini, D.; Franchi, M.; Ruggeri, A.; Raspanti, M. *Micron* **2002**, *33*, 587−596; (b)Ottani, V.; Raspanti, M.; Ruggeri, A. *Micron* **2001**, *32*, 251−260.

(19) Orgel, J. P. R. O.; Miller, A.; Irving, T. C.; Fischetti, R. F.; Hammersley, A. P.; Wess, T. J. *Structure* **2001**, *9*, 1061−1069.

(20) Klein. T. E.; Huang, C. C. *Biopolymers* **1999**, *49*, 167−183.

(21) Jenkins, C. L.; Bretscher, L. E.; Guzei, I. A.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125*, 6422−6427.

(22) (a) Chopra, R. K.; Ananthanarayanan, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 7180−7184. (b) Berg, R. A.; Prockop, D. J. *Biochem. Biophys. Res. Commun.* **1973**, *52*, 115−120.

(23) Benzi, C.; Improta, R.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2002**, *23*, 341−350.

(24) Mooney, S. D.; Kollman, P. A.; Klein. T. E. *Biopolymers* **2002**, *64*, 63−71.

(25) (a) Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. *Chem. Biol.* **1999**, *6*, 63−70. (b) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, L. T. *J. Am. Chem. Soc.* **2001**, *123*, 777−778. (c) Vitagliano, L.; Berisio, M.; Mazzarella, L.; Zagari, A. *Biopolymers* **2001**, *58*, 459−464.

(26) DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 2497−2505.

(27) Buevich, A. V.; Baum, J. *J. Am. Chem. Soc.* **2002**, *124*, 7156−7162.

(28) Smith, J. W. *Nature* **1968**, *219*, 157−158.

(29) Fraser, R. D. B.; MacRae, T. P.; Miller, A.; Suzuki, E. *J. Mol. Biol.* **1983**, *167*, 497−521.

(30) Piez, K. A.; Trus, B. L. *Biosci. Rep.* **1981**, *1*, 801−810.

(31) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(32) (a) King, G.; Brown, E. M.; Chen, J. M. *Protein Eng.* **1996**, *9*, 43−49. (b) Brown, E. M.; Dudley, R. L.; Elsetinow, A. R. *J. Am. Leather Chem. Assoc.***1997**, *62*, 225−233. (b) Qi, P. X.; Brown, E. M. *J. Am. Leather Chem. Assoc.* **2002**, *97*, 235−342.

(33) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crouley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 7*; University of California: San Francisco, 2002.

(34) Martell, A. E.; Smith, R. M. *Critical Stability Constants*; Plenum: New York, 1989; Vol. 3.

(35) Walser, R.; Hünenberger, P. H.; van Gunsteren W. F. *Proteins: Struct., Funct., and Genet.* **2001**, *44*, 509−519.

(36) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335−350.

(37) (a) Cieplak, P.; Cornell, W. D.; Bayly, C. I.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357−1377. (b) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision A.1; Gaussian, Inc.: Pittsburgh, PA, 2003.

(39) Berendsen, H. J. C.; Postma, J. P. M.; vanGunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Comput. Phys.* **1984**, *81*, 3684−3690.

(40) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327−341.

(41) Forrest, L. R.; Tieleman, D. P.; Sansom, M. S. P. *Biophys. J.* **1999**, *76*, 1886−1896.

(42) Hobza, P.; Kabelác, M.; Sponer, J.; Mejzlík, P.; Vondrásek, R. *J. Comput. Chem.* **1997**, *18*, 1136−1150.

(43) Alagona, G.; Ghio, C.; Giolitti, A.; Monti, S. *Theor. Chem. Acc.* **1999**, *101*, 143−150.

(44) Alagona, G.; Ghio, C.; Monti, S. *Int. J. Quantum Chem.* **2002**, *88*, 133−146.