Toward the Supramolecular Structure of Collagen: A Molecular Dynamics Approach

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JULY 2005

Impact Factor: 3.3 · DOI: 10.1021/jp0440941 · Source: PubMed

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

Toward the Supramolecular Structure of Collagen: A Molecular Dynamics Approach

Susanna Monti*

Istituto per i Processi Chimico-Fisici del Consiglio Nazionale delle Ricerche (IPCF-CNR), Area della Ricerca, via G. Moruzzi 1, I-56124 Pisa, Italy

Simona Bronco and Chiara Cappelli*

PolyLab−INFM, c/o Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Risorgimento 35, I-56126 Pisa, Italy

Received: December 28, 2004; In Final Form: April 14, 2005

The structure, stability, and conformational dynamics of an assembly of two pentameric bundles made of collagen-like triple helical segments are explored using 1.2 ns molecular dynamics simulations in three environments: 8.0% (v/v) formaldehyde/water solution, 1.4% (v/v) gallic acid/water solution, and pure water. Stable supramolecular arrangements, where the two collagen units are very close to each other at interacting distances, are identified via docking and energy minimization procedures. Analysis of the interaction with formaldehyde and gallic acid suggests that they perturb the protein in a similar way depending on hydrogen-bonding capability, hydrophobic association properties, and the size and concentration of the compound.

1. Introduction

Though connective and supporting tissues appear in disparate forms in different parts of the body, there is a fundamental similarity in their features, and any particular modification of form represents an adaptation to function.

Three major components are demonstrable and can be recognized in almost all connective tissues: the cells, the extracellular fibers, and the extracellular nonfibrillar ground substance.

The fibrillar and ground substance material, present in great abundance, gives the connective tissues their main characteristics. The ropelike tensile strength of a tendon, for example, is related to the dense bundles of collagen fibers of which a tendon is composed.

Collagen is a protein with a complex supramolecular organization; its molecules interact with each other at different hierarchical levels forming a characteristic higher-order periodic structure with distinctive features and specific functions. The precise nature of collagen packing and self-assembly in tissues has been widely investigated, but the absolute three-dimensional (3D) arrangement of collagen molecules, their lateral supermolecular organization, is not well understood. In 1968, Smith[1] proposed the so-called five-stranded pentagonal microfibril model that consisted of five molecules coiled into a hollow cylinder, a description that was consistent with both the X-ray diffraction and the transmission electron microscopy (TEM) data. In 1979, Hulmes and Miller[2] suggested that type I collagen molecules adopted a tilted quasi-hexagonal packing arrangement. Their model was also in good agreement with the X-ray diffraction data. From the investigation of the fibrillar substructure of a wide variety of tissues, Ruggeri et al.[3−5] found that collagen molecules could arrange as straight microfibrils with rodlike fibrils in cartilage, microfibrils could wind along a very small angle to the fibril axis in tendons and ligaments, and their winding angle could be of about 18° in blood vessel walls, interstitial tissues, corneas, and nerve sheaths.[6−8] Moreover, this last arrangement was compatible with a quasi-crystalline hexagonal packing.[9]

However, many other models attempting to combine X-ray diffraction data, electron microscopic results, and biochemical data have been suggested over the years.[10−22]

Each was a compromise and received its share of confirmations and confutations from the various experimental techniques available. Unfortunately, no consensus structure has emerged.

In a previous work of ours,[23] we studied the change in the stability and conformational dynamics of a collagen-like microfibril segment (CMS) 23 amino acid (AA) residues long (around 5[3(Gly-X-Y)$_8$]) that was a small portion of a longer model whose 3D structure, based on the Smith description, was obtained by Brown and co-workers[24] through computer modeling. In particular, the analysis of the dynamics of the CMS surrounded by pure and mixed solvents (water, formaldehyde, and gallic acid (3,4,5,trihydroxybenzoic acid)) showed that major changes in the conformation of the CMS occur in water whereas in pure formaldehyde and in the mixed, water/H_2CO and water/gallic acid, solutions the deviation from the starting structure is smaller.[23] A more detailed analysis of the molecular dynamics (MD) runs also showed that the Arg, Lys, Asn, and Gln residues of the CMS interact more favorably with formaldehyde molecules, whereas preferred binding to Pro and Hpr (hydroxyproline) residues was shown for gallic acid. The computational results of ref 23 are in satisfactory agreement with experiments, thus demonstrating that MD simulations (performed using the AMBER7 package with the Cornell et al. force field[25] and the general AMBER force field (GAFF)) can provide a realistic description of the CMS in different environments. Starting from these premises and with the aim of extending the study toward the understanding of the collagen supramolecular structure and how it is modified when collagen is surrounded by chemicals, here for the first time we shall examine the mechanism of both intra- and inter-microfibril cross-linking of a dimeric CMS

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

10.1021/jp0440941 CCC: $30.25
Published on Web 05/13/2005

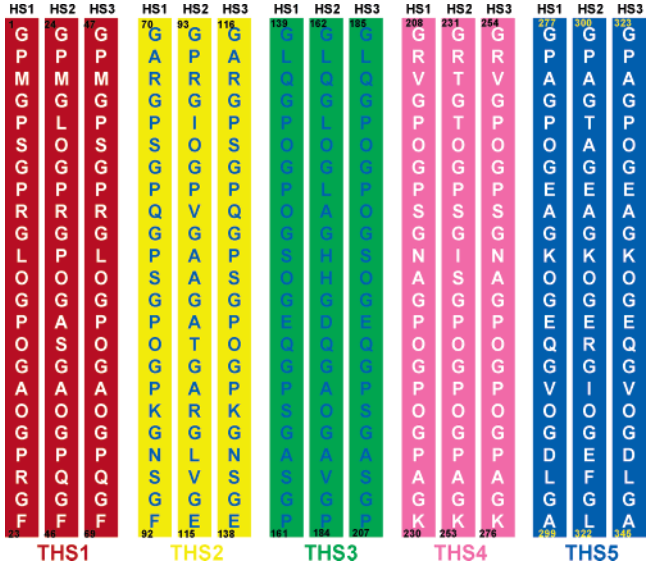

Figure 1. Schematic representation of the CMS amino acid sequence. The one-letter code is used to label the AA residues. O stands for Hpr.

complex. Similar to what was done in our previous paper,[23] to permit a direct comparison between the single-fibril and the two-fibril models we shall limit the analysis to the same chemicals already considered, i.e., formaldehyde and gallic acid. For comparison, the results of a simulation in pure water are also shown.

2. Methods

2.1. Collagen Model. The collagen-like microfibril segment (CMS) used in these studies was generated as described in a previous work of ours.[23] Briefly, a short segment 23 AA residues long (around 5[3(Gly-X-Y)$_8$]) thus made of five triple helical segments (THSs), rich in hydrophilic residues, was extracted from a 3D computer model of the bovine type I collagen microfibril based on the Smith description[1] made available to us by Professor Eleanor M. Brown.[24] The AA content of CMS is reported in Figure 1. Before simulation, side-chain ionization states were adjusted on the basis of the pH of the environment. All of the C-terminal carboxylates and N-terminal amino groups were neutral. The net charge of the CMS bundle used in the simulation at neutral pH was +4 whereas at acidic pH the total charge of the system was +21.

2.2. Generation of the Two CMS Bundle Models. Exactly 24 models (12 for each pH value) of dimeric CMS helices were generated using a medium-resolution docking approach followed by in vacuo energy minimizations. Molecular surface complementarity was found to be useful to efficiently scan and identify potential interaction sites between the two molecular assemblies. A geometric recognition algorithm that takes advantage of techniques applied in the field of pattern recognition, based essentially on the maximization of the intermolecular surface contact, avoiding overlap, was used.[26−29] The 3D atomic structures of the two CMSs (I and II) were projected on a 3D grid. The size of the grid step, which determines the resolution or accuracy of molecular representation, was set to 4.1 Å and a 90° angle limit between the main axis of the two bundles was chosen. No detail smaller than the grid step is reflected in the molecular 3D projections; thus our choice corresponded to an averaging of atom−atom interactions at the level of individual residues. All possible orientations of the second CMS with respect to the first one were tried systematically with three different predefined angular steps: 10°, 15°, and 20°. Both parallel and antiparallel orientations of the CMS pair were explored. At each orientation, the position of the CMSs with the highest score, proportional to the area of contact between the two systems, was retrieved. The procedure yielded an ensemble of 6000 possible conformations for each pH value: 3000 with parallel and 3000 with antiparallel arrangements from which the 6 highest structures were selected as starting points of in vacuo energy minimizations. Among the 24 models obtained, the most stable parallel and antiparallel conformations at neutral and at acidic pH values were chosen for molecular dynamics (MD) simulations in mixed solutions. The analysis of the energies of the minimized structures shows that the chosen antiparallel structure represents 50% of the Boltzmann population, whereas the parallel one accounts for 80%. To end this section, it should be noted that our system, which models the packing of two identical microfibril segments, despite being a reasonable starting point toward the understanding of the collagen supramolecular structure, does not take into account the fact that in reality the interacting microfibril segments could be offset along the sequence.

2.3. Setup of the Simulation Systems. The four models of the CMS dimer, parallel arrangement at acidic pH, antiparallel arrangement at acidic pH, parallel arrangement at neutral pH, and antiparallel arrangement at neutral pH, were solvated with cross-linking agent, water molecules, and chloride counterions (added to maintain electroneutrality) and then energy minimized. After brief MD equilibration stages (for details see ref 23), ∼1.2 ns production simulations were performed, restraining the backbone atoms of the residues placed at both ends of the 30 α-chains to avoid border effects. However, the restraining forces were such that the atoms were not fixed but could reasonably move. Three systems were considered: (1) CMS dimer in about 1.4% (v/v) gallic acid/water (acidic solution), (2) CMS dimer in about 8.0% (v/v) formaldehyde/water (neutral solution), and (3) CMS dimer in pure water (neutral solution). All of the simulations were performed with the addition of counterions, and the gallic acid molecules were in the neutral form, which exist only at a pH value below 4.26.[30]

2.4. Molecular Dynamics Simulation Protocol. MD simulations were carried out as described in our earlier paper[23] using periodic boundary conditions in the NPT ensemble. A constant pressure of 1 bar was applied with a coupling constant $\tau_p = 1$ ps,[31] and the system was maintained at 305 K using a coupling constant $\tau_T = 0.1$ ps.[31] Long-range electrostatic interactions were dealt with using Ewald summations. A 12 Å cutoff was applied to the van der Waals interactions. The time step was 1 fs, and the SHAKE algorithm[32] was used to constrain all hydrogen atom−heavy atom bond lengths to their equilibrium values. Water was described using the TIP3P model.[33] All MD simulations were performed using the AMBER7 package,[34] the Cornell et al. force field,[25] and the general AMBER force field (GAFF) running on the IBM Linux Cluster 1350 (CLX) at CINECA. The coordinates were saved every 1 ps to analyze the trajectory. The analysis of the trajectories was done on Compaq DS20E and Digital Alpha XP1000 workstations and on SGI O^2 and Indigo^2 workstations at the IPCF.

3. Results and Discussion

3.1. Packing of the Two CMS Bundles: Structural Features and Dynamic Behavior. The studied structures will be abbreviated in the text as follows. PA and PN are the parallel arrangements at acidic and neutral pH values, respectively; APA and APN are the antiparallel arrangements at acidic and neutral pH values, respectively. The starting antiparallel associated

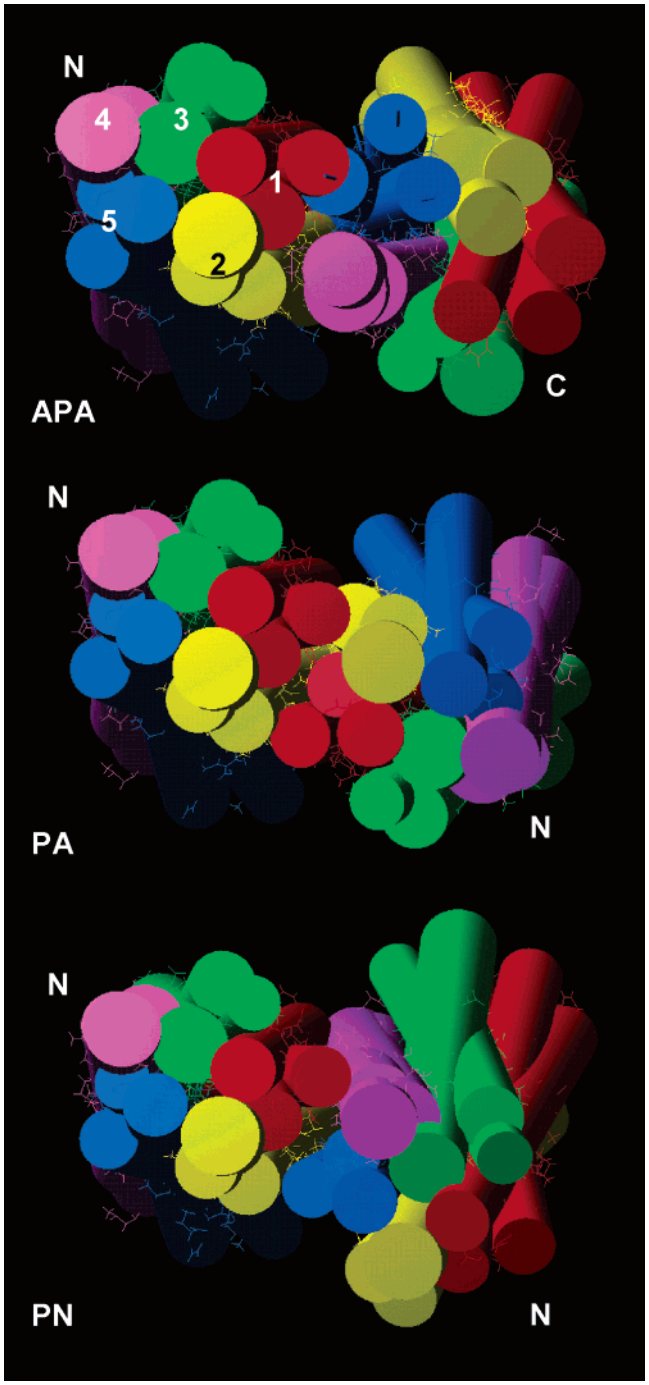

Figure 2. Three-dimensional structure of the starting APA, PA, and PN models. The 30 helices are drawn as cylinders. Three cylinders form a THS and five THSs form a CMS. Each THS can be identified by its color code: THS1 (red cylinders), THS2 (yellow cylinders), THS3 (green cylinders), THS4 (magenta cylinders), and THS5 (cyan cylinders). The CMS N- and C-terminal regions appearing in the foreground are indicated by the letters N and C, respectively.

complexes are more stable than the parallel ones. The dipolar interaction among the helical dipoles is more favorable when the CMS helices are antiparallel arranged. However, in all complexes van der Waals interactions promote the detailed close packing of the CMS helices.

The initial conformations of the models, generated using restrained in vacuo energy minimizations, are shown in Figure 2. In the antiparallel structures, APA and APN, the triple helical segments 1 and 2 of CMS_I ($THS1_I$, $THS2_I$; see Figure 1) are opposite the triple helical segments 5 and 4 of CMS_{II} ($THS5_{II}$, $THS4_{II}$), respectively, and are located at the interbundle interface. $THS1_I$ C_α atom−$THS5_{II}$ C_α atom and $THS2_I$ C_α atom−$THS4_{II}$ C_α atom distances are longer toward the C-terminal and N-terminal regions, whereas shorter distances are found in the middle of the helices with minimal values in the range 4.7−5.7 Å, involving mainly alanine residues.

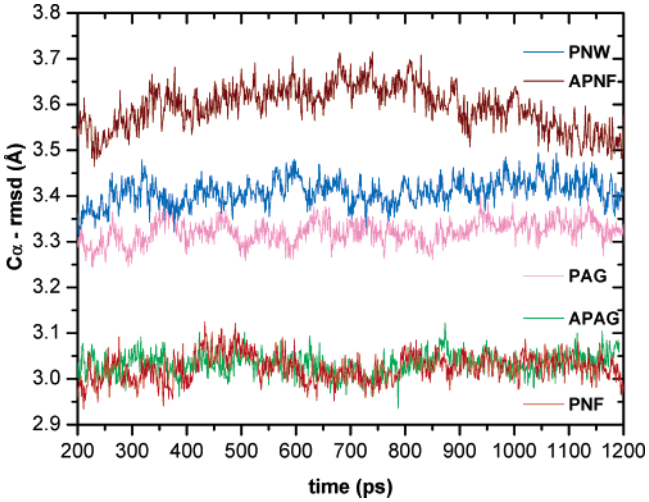

Figure 3. Root-mean-square deviations (rmsd's) versus time for the C_α atoms of each simulation.

The two bundle axes are not parallel but form an angle of about 160°, and the angles between $THS1_I$ and $THS5_{II}$ and between $THS2_I$ and $THS4_{II}$ axes are around 170°. The total exposed surface area is about 19 560 Å^2. APA and APN models show a great similarity in their 3D arrangements; indeed the root-mean-square deviation obtained superimposing the C_α traces is only 0.12 Å.

In the parallel structure PA, $THS1_I$ and $THS2_I$ are opposite $THS2_{II}$ and $THS1_{II}$, respectively, and are located at the interbundle interface. Tight contacts between $THS1_I$ and $THS2_{II}$ and between $THS2_I$ and $THS1_{II}$ are found in the central region of the helices with minimal $C_{\alpha I}$−$C_{\alpha II}$ distances in the range 4.8−6.0 Å. Also in this case among the closest contacts the alanine C_α prevails. The two bundle axes form an angle of about 19°, and the angles between $THS1_I$ and $THS2_{II}$ axes and between $THS2_I$ and $THS1_{II}$ axes are about 10°. The total exposed surface area is about 19 750 Å^2.

In the parallel structure PN, $THS1_I$ and $THS2_I$ are opposite $THS4_{II}$ and $THS5_{II}$, respectively, and are located at the interface between the two bundles. Closer $C_{\alpha I}$−$C_{\alpha II}$ distances are found in the central region of the helices with minimal values in the range 4.0−6.0 Å. The two bundle axes form an angle of about 23°, and the angles between $THS1_I$ and $THS4_{II}$ axes and between $THS2_I$ and $THS5_{II}$ axes are about 8°. The total exposed surface area is about 19 850 Å^2.

The data reported above show that the four arrangements (APA, APN, PA, and PN), despite their structural differences and protonation states, share common features in that the angles between the two CMS bundles and the solvent-exposed surfaces are very similar. In addition, in all cases the two CMSs are very close to each other, and the lateral chains are located at interacting distances.

Similar to what was reported in ref 23, to examine some of the conformational changes that occur in the starting bundle assemblies as they relax in solution, we have considered the C_α root-mean-square deviation (rmsd) versus time, the C_α root-mean-square fluctuation (rmsf) as a function of the residue number, the bundle crossing angles, the C_α distance map, and the hydrogen-bonding pattern. All the reported quantities have

been computed over the last 1 ns of the simulations leaving, the first 200 ps as a further equilibration phase.

3.1.1. C_α rmsd and C_α rmsf. An overall measure of the drift from the initial structure is provided by the behavior of the C_α rmsd relative to the initial conformation as a function of time. In Figure 3, the C_α rmsd values versus time of each simulation are shown. Most of the observed overall deviation from initial structure occurs during the equilibration phase, indicating that the complexes are stable and they do not change substantially in the last 1 ns of the production runs. The difference in the solvent does not have a marked effect on the progress of the simulations. The C_α rmsd for both APA gallic acid/water (APAG) and PN formaldehyde/water (PNF) models rises to about 3.03 Å during the equilibration period then fluctuates between 2.94 and 3.12 Å with a standard deviation of about 0.03 Å.

In PA gallic acid/water (PAG), PN water (PNW), and APN formaldehyde/water (APNF) models, the starting C_α rmsd values, 3.32, 3.41, and 3.59 Å, respectively, are higher and different from one another, but their fluctuations are similar to those shown by APAG and PNF complexes with a standard deviation of about 0.04 Å. The fluctuation ranges are 3.24−3.41 Å for PAG, 3.31−3.49 Å for PNW, and 3.45−3.71 Å for APNF.

The fluctuations of the individual residues have been measured as C_α rmsf values averaged over time and across all 30 helices of the two bundles. As expected, the fluctuations are extremely reduced at the helix termini due to the positional restraints applied during the whole simulation time, whereas at the center of the helices greater fluctuations are observed especially in the PNW case. The C_α rmsf values for all of the residues are predominantly less than 1 Å, indicating a general stability of all five models. The greatest fluctuations are at the residues on the helical side exposed to the solvent. However, about 7% of the PNW residues and about 6% of both the APNF and the PNF residues show C_α rmsf values greater than 1 Å, while in the other cases, APAG and PAG, this percentage is <3%.

Such a stability of the bundle and the greatest fluctuations observed in a pure water environment confirm the findings shown in ref 23. In particular, once again the behavior in water is in agreement with experimental studies reporting water as one of the major contributors to the denaturation enthalpy of collagen.[23,35]

3.1.2. Bundle Crossing Angles. The relative orientation of the bundles has been measured in terms of bundle crossing angles, i.e., the angle between the long axes of the two CMSs. Crossing angles were calculated for each structure sampled during the simulation, and their values for each pair of opposed THSs at the interbundle interface were also examined.

The average CMS crossing angle in APA models is about 160°, whereas PA models have CMS crossing angles of about 21°. Examination of the various crossing angle trends as a function of time shows that its range of variation is quite narrow from 156° to 163° for APA complexes and from 17° to 26° for PA complexes. In addition, the relative orientation of the helices at the interbundle interface has small fluctuations. All of these findings suggest that there is no rearrangement of the helix packing within the various models during the course of the simulation, thus confirming the general stability of all five models.

3.1.3. C_α Distance Maps. The C_α distance map of each complex is a 69 × 69 matrix. (We recall that in our model each helix is composed of 23 AA residues, so that the triple helical segment consists of 69 AA residues.) The matrix elements *ij* are equal to the mean distance between the $C_{\alpha i}$ and the $C_{\alpha j}$ atoms of the CMS triple helical segments at the interbundle interface. The C_α distance maps of the starting structures were also calculated for comparison. Two residues were considered as contacting each other when the distance between their C_α atoms is smaller than 6 Å. The number of contacts for each simulation, the contact with the minimum mean distance, and statistical parameters defining its variation during MD simulations are given in Table 1 (data refer to the last 1 ns of the simulation). The antiparallel complexes have a minor number of contacts with respect to the parallel ones. If we compare the maps of the starting conformations with those obtained from MD simulations, then there appear to be a growing number of contacts for parallel models. The majority of the contacts involve alanine residues and are maintained during the whole simulation (σ_{max} is about 0.6 Å).

TABLE 1: Number of Contacts (NCs) and the Contact with the Minimum Mean Distance, in Å, (MMD) for the Initial Models and for the MD Simulations[a]

model	NC	residues	$C_\alpha-C_\alpha$ MMD	σ	min	max
PN	13	Ala106−Ala676	3.99			
PA	15	Ala40−Thr452	4.75			
APN	8	Ala106−Hpr590	4.75			
APA	8	Ala106−Hpr590	4.75			
APAG	6	Ala103−Pro569	5.20	0.23	4.57	6.03
APNF	8	Ala103−Pro569	4.72	0.24	4.07	5.81
PAG	18	Pro51−Pro465	4.18	0.42	4.69	7.31
PNF	25	Gly91−Hpr662	4.36	0.42	3.54	6.02
PNW	28	Gly42−Hpr570	3.92	0.21	3.36	4.70

[a]Minimum (min) and maximum (max) values together with the standard deviations (σ) are also reported.

Closer inspection of each individual case showed that both APNF and APAG complexes have a small increase in the C_α distances between $THS1_I$ and $THS5_{II}$, which are opposite one another, with a consequential loss of contacts especially in the central region of the THSs. The interhelical interfaces $THS2_I$ and $THS4_{II}$ experience instead much greater readjustments, a displacement of the contact regions at the middle of the helices and an enlargement of the N- and C-terminal distances (Figure 4). In the case of PAG, both $THS2_I-THS1_{II}$ and $THS1_I-THS2_{II}$ pairs maintain their contacts even if greater distances are observed in their terminal regions.

The PNF and PNW models show some similarities in their maps and behavior. In both models, $THS1_I$ and $THS4_{II}$ interfacial contacts are slightly affected by the environment whereas the other pair of triple helical segments ($THS2_I$ and $THS5_{II}$) exhibits major changes; larger $C_\alpha-C_\alpha$ distances appear at the ends of some helices, but new contacts are achieved in their middle section (Figure 4).

3.1.4. Hydrogen Bonds. Patterns of hydrogen bond formation between the two CMS bundles have been analyzed. To keep track of the potential hydrogen bond interaction, a cutoff distance of 3.3 Å between donor and acceptor atoms and a hydrogen donor−acceptor cutoff angle of 30° have been specified. The percentage of occupancy of a hydrogen bond was defined as the number of sampled structures with the hydrogen bond present divided by the total number of structures used for analysis. The lifetime of a hydrogen bond was calculated as the time elapsed from its first appearance until it was first broken. A small number of interhelical hydrogen bonds at the interbundle interface are formed (Table 2). These are often found between AA side chains, less frequently between the backbone atoms, and are mostly localized at the middle toward the end of the helices.

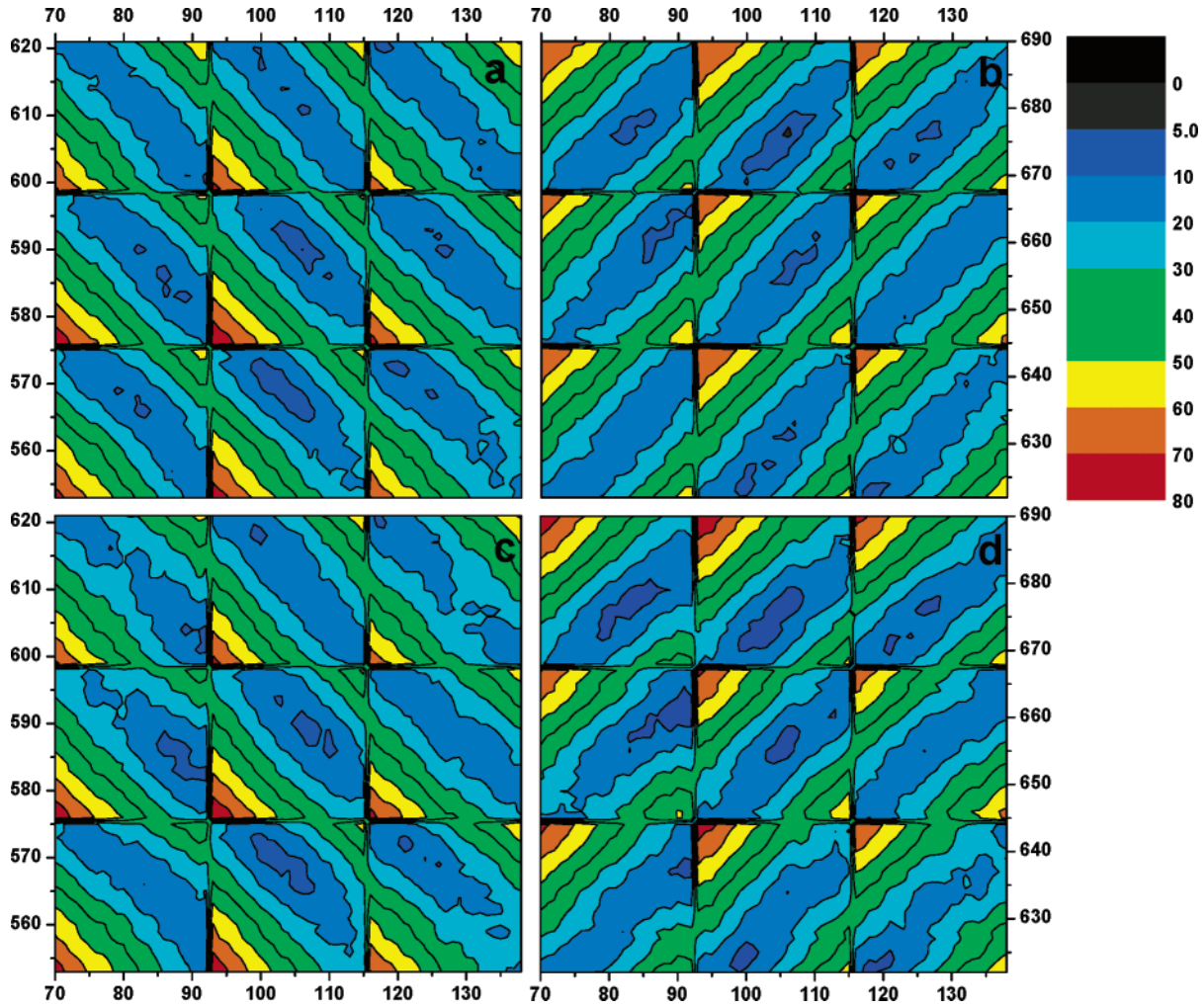

Figure 4. C_α distance maps. (a) APAG $THS2_I-THS4_{II}$ starting distances; (b) PNF $THS2_I-THS5_{II}$ starting distances; (c) APAG $THS2_I-THS4_{II}$ mean distances; (d) PNF $THS2_I-THS5_{II}$ mean distances. The axes are labeled with the residue numbers: $THS2_I$ residues 70−138; $THS4_{II}$ residues 553−621; $THS5_{II}$ residues 622−690. Each square refers to one helix.

TABLE 2: Total Number of Interhelical Hydrogen Bonds at the Interbundle Interface, Its Components Classified as Backbone−Backbone (bb), Side Chain−Side Chain (ss), and Backbone−Side Chain (bs)[a]

model	total	bb	ss	bs	% occupancy > 50%
APAG	11	3	4	4	5
APNF	17	1	3	13	4
PAG	18	5	7	6	2
PNF	21	5	2	14	5
PNW	21	6	1	14	3

[a] The number of hydrogen bonds with a percentage of occupancy greater than 50% is also shown.

In PNF and PNW models, a more consistent hydrogen-bonding pattern is seen, but statistical analysis of individual contributions shows that the PNF case has the highest number of hydrogen bonds that persist throughout the simulation. The PAG model has a higher number of hydrogen bonds than the APAG complex; however, only two of them, coming from backbone interactions, are maintained during the simulation. In the antiparallel case, five hydrogen bonds are preserved and involve both backbone and side-chain atoms.

3.2. Structure and Dynamics of Solvent Molecules. In addition to the conformational dynamics of the CMS bundles, it is of interest to analyze the *interaction* of the helices with their solvent environment made of cross-linking agent and water molecules. All solvent species, water, formaldehyde, and gallic acid, are found in the first solvation layer around the bundles where they are in direct contact with the protein residues. A quantitative measurement of the solvent distribution surrounding the CMSs can be given by the solvent−THS radial distribution functions (rdf's), which may provide information on the local structure of the solution (Figure 5 and Table 3). To describe the dynamic behavior of solvent molecules, it is useful to identify the protein-bound ones and to characterize the hydrogen-bonding networks with their lifetimes (Tables 4−6).

TABLE 3: Coordination Numbers Derived from Water (wat), Gallic Acid (GA), and Formaldehyde (FA) rdf's Taking 2.3 Å as the Coordination Radius for the Five THSs of Each CMS (I and II) in the Different Models under Examination

		APAG		PAG		APNF		PNF		PNW
CMS	THS	wat	GA	wat	GA	wat	FA	wat	FA	wat
I	1	86.5	7.9	110.6	7.3	89.6	10.4	89.7	10.2	103.6
	2	98.2	6.0	103.5	8.4	108.2	13.1	101.4	6.6	122.3
	3	105.8	17.7	115.7	6.3	125.9	7.0	118.7	8.1	136.8
	4	102.8	9.2	124.7	11.0	115.3	8.3	104.3	8.1	124.2
	5	129.6	9.6	128.4	13.0	143.1	8.4	120.8	9.2	149.3
II	1	115.1	13.3	115.7	8.5	100.6	9.8	105.8	9.2	130.6
	2	106.5	17.9	95.3	10.2	120.3	7.4	117.6	8.4	146.1
	3	121.9	7.5	119.7	8.2	131.9	8.5	126.2	7.7	134.1
	4	118.6	8.2	107.1	8.4	108.7	9.8	91.6	7.0	112.1
	5	95.4	9.9	111.7	13.4	110.2	8.3	105.1	5.4	126.8

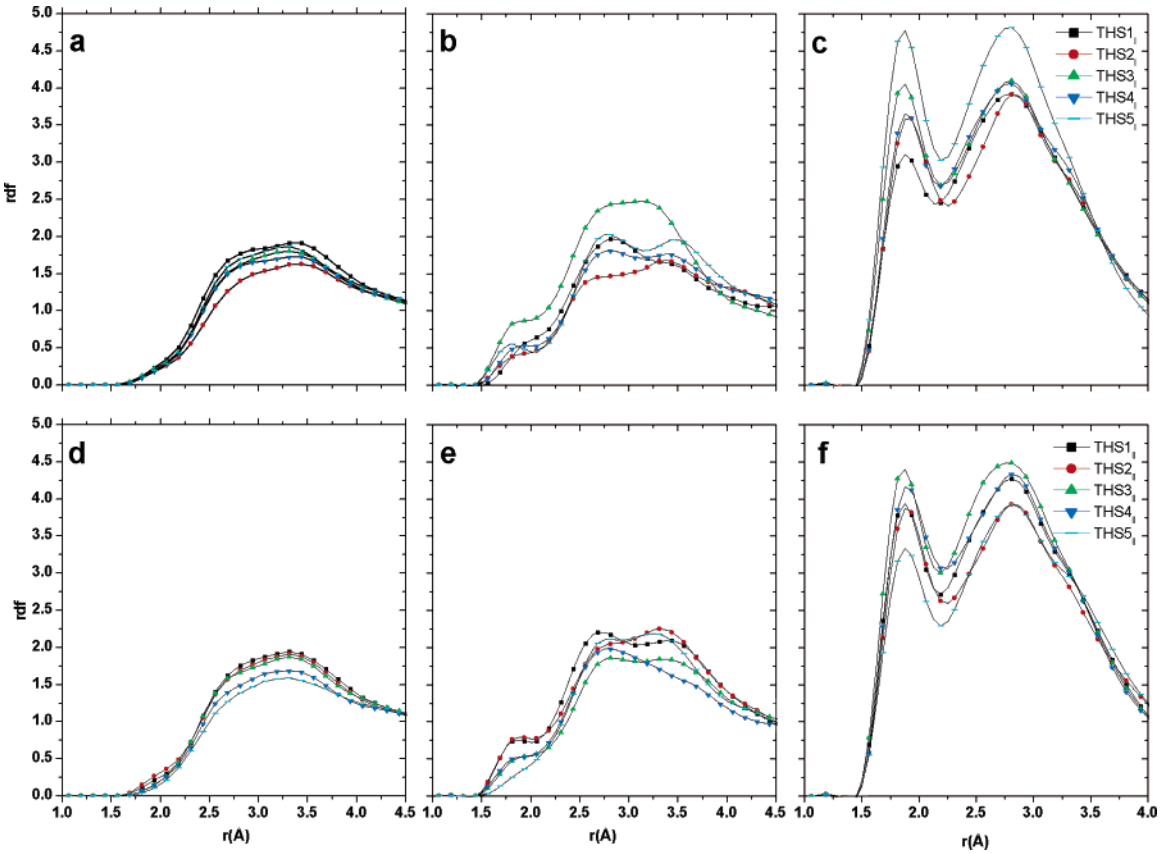

Figure 5. Radial distribution functions (rdf's) for each THS with (a and d) formaldehyde oxygen atoms (PNF model), (b and e) the entire group of gallic acid oxygen atoms (APAG model), and (c and f) water oxygen atoms (APAG model).

TABLE 4: Time Evolution of the Number of Hydrogen Bonds between Water Molecules and CMS Residues (wat) and Gallic Acid Molecules and CMS Residues (GA), with Percentages of Occupancy ≥10%, ≥25%, ≥50%, and ≥90%, for the APAG and PAG Models

	APAG percentage of occupancy								PAG percentage of occupancy							
time	≥10%		≥25%		≥50%		≥90%		≥10%		≥25%		≥50%		≥90%	
(ps)	GA	wat	GA	wat	GA	wat	GA	wat	GA	wat	GA	wat	GA	wat	GA	wat
50	58	508	41	213	28	107	13	34	66	597	38	274	27	132	9	33
100	63	470	44	193	29	95	9	20	69	616	41	246	27	106	9	21
150	75	468	44	176	28	78	9	18	64	608	43	221	24	89	9	21
200	75	449	49	183	26	63	9	16	67	573	42	213	23	80	8	19
250	77	425	50	167	22	52	9	15	72	545	39	197	20	84	8	13
300	78	412	49	163	24	49	8	10	71	508	35	184	18	78	6	8
350	78	397	48	145	26	49	8	10	74	488	35	187	17	61	7	10
400	77	387	48	134	26	50	7	10	76	470	37	186	17	57	6	9
450	79	383	49	129	25	51	7	10	74	467	36	183	17	52	6	9
500	78	379	47	124	25	48	7	7	79	433	40	180	16	47	6	9
550	81	367	46	119	28	46	8	7	82	424	40	176	15	50	6	9
600	78	345	46	122	27	48	8	8	82	409	41	169	15	53	6	10
650	81	339	47	120	28	47	7	8	83	401	41	158	14	53	6	10
700	83	329	49	120	26	43	7	7	83	386	43	154	15	52	7	10
750	84	332	46	120	27	43	7	6	84	369	43	144	15	49	7	10
800	83	317	48	122	27	45	7	8	88	367	42	140	15	53	7	10
850	84	307	50	124	26	38	8	7	89	369	39	135	15	50	6	7
900	85	304	51	124	26	42	8	7	90	366	37	128	15	48	6	6
950	86	297	48	122	26	42	7	7	91	361	37	127	13	50	5	6
1000	81	291	48	119	25	42	7	7	91	359	34	124	13	50	4	6

3.3. Distribution of Solvent Species. In Figure 5, the rdf's of THS with formaldehyde, gallic acid, and water are depicted. In particular, only the PNF and APAG models are shown because very similar behavior is observed in the other cases. In the whole set of simulations, water rdf's (Figures 5c and 5f) show a sharp first neighbor peak centered at ∼1.8 Å and a first minimum at ∼2.3 Å. The second peak is also well-defined but is broader than the first one, and its maximum is at ∼3.0 Å. The coordination numbers derived from water rdf's, taking 2.3 Å as the coordination radius, are reported in Table 3. Smaller coordination numbers are observed for inner THSs, i.e., those located at the interbundle interface. As expected, the PNW simulation shows the highest coordination numbers; in fact, in the other cases, even if water molecules equilibrate themselves

TABLE 5: Time Evolution of the Number of Hydrogen Bonds between Water Molecules and CMS Residues (wat) and Formaldehyde Molecules and CMS Residues (FA), with Percentages of Occupancy ≥10%, ≥25%, ≥50%, and ≥90%, for the APNF and PNF Models

	APNF percentage of occupancy								PNF percentage of occupancy							
time	≥10%		≥25%		≥50%		≥90%		≥10%		≥25%		≥50%		≥90%	
(ps)	FA	wat	FA	wat	FA	wat	FA	wat	FA	wat	FA	wat	FA	wat	FA	wat
50	42	822	17	371	10	168	1	42	34	725	15	345	8	170	1	55
100	40	810	20	323	10	143	1	27	33	740	16	286	10	131	2	35
150	45	765	19	287	8	122	1	22	30	700	17	254	7	112	2	30
200	41	743	18	290	7	108	1	15	26	662	16	238	7	104	0	26
250	43	696	17	263	7	107	1	12	26	633	15	219	7	93	0	23
300	40	675	16	247	6	96	1	12	29	619	15	205	4	89	0	19
350	44	652	17	245	6	88	1	11	30	600	15	198	5	80	0	20
400	40	620	13	252	6	81	1	13	32	556	16	198	3	83	0	17
450	42	615	13	237	5	76	1	13	33	540	17	192	5	77	0	15
500	45	597	13	232	5	75	1	12	35	510	17	189	4	71	0	15
550	44	567	12	222	5	68	1	12	34	501	19	184	4	67	0	15
600	44	543	12	214	4	64	1	12	34	494	15	194	5	68	0	14
650	44	530	12	206	4	62	1	12	34	472	13	194	5	65	0	14
700	41	517	11	199	4	60	1	11	32	458	13	188	5	68	0	13
750	39	501	13	187	5	61	1	10	32	444	12	181	5	68	0	12
800	38	494	13	186	5	60	1	10	33	437	13	176	3	70	0	13
850	37	490	12	180	4	61	1	10	33	434	13	174	3	67	0	14
900	36	491	12	175	4	56	1	10	32	423	13	175	2	65	0	14
950	33	486	11	174	4	57	1	8	31	423	13	180	2	59	0	13
1000	32	466	11	171	4	57	1	7	32	423	13	176	2	62	0	12

TABLE 6: Time Evolution of the Number of Hydrogen Bonds between Water Molecules and CMS Residues, with Percentages of Occupancy ≥10%, ≥25%, ≥50%, and ≥90%, for the PNW Model

	PNW percentage of occupancy			
time	≥10%	≥25%	≥50%	≥90%
(ps)	wat	wat	wat	wat
50	1079	467	221	51
100	1064	408	167	38
150	1004	342	135	30
200	938	334	126	28
250	891	311	116	21
300	832	292	101	14
350	798	282	94	14
400	764	287	93	15
450	738	281	86	16
500	713	272	82	13
550	710	260	82	11
600	694	258	83	10
650	671	254	87	11
700	665	245	84	9
750	658	247	85	8
800	637	239	84	8
850	626	237	84	6
900	628	228	83	6
950	626	227	72	5
1000	627	224	67	5

close to the relatively polar surface of the THSs, the cross-linking agent is present in this primary solvation layer and participates in direct interactions with the collagen segments.

As far as gallic acid rdf's are concerned (Figures 5b and 5e), a small and not well-defined first shell peak centered at ∼1.8 Å with a minimum at ∼2.3 Å is observed for all the THSs in both the parallel and the antiparallel arrangements of the CMSs. In some cases, it is only a small shoulder of very low amplitude extending between ∼1.5 and 2.3 Å. Apart from this shoulder, the main first peak is spread out over intervals between ∼2.3 and 4.0 Å.

In contrast, formaldehyde rdf's (Figures 5a and 5d) do not display any first peak at short intermolecular distances, but they gradually raise to a first maximum further away at ∼3.2 Å. To identify the number of cross-linking-agent molecules in the first hydration shell, the coordination numbers have been evaluated at the same distance (2.3 Å) chosen for water molecules (Table 3).

The role played by gallic acid molecules in stabilizing the model system structure was examined in more detail by analyzing separately the rdf's of their 3-, 4-, and 5-position hydroxylic oxygen atoms (OHp) and carboxylic oxygen atoms (O2) with various atoms and groups of atoms belonging to the side chains of the following amino acid residues: $Arg_{(N_{NH},N_{NH2})}$, $Asp_{(O_{COOH})}$, $Asn_{(N_{NH2})}$, $Glu_{(O_{COOH})}$, $Gln_{(N_{NH2})}$, $His_{(N_{NH},N)}$, $Hpr_{(O_{OH})}$, $Lys_{(N_{NH3+})}$, $Ser_{(O_{OH})}$, and $Thr_{(O_{OH})}$. These rdf's are shown in Figure 6.

The rdf's of OHp with $Arg_{(N_{NH},N_{NH2})}$, $Hpr_{(O_{OH})}$, $Lys_{(N_{NH3+})}$, $Glu_{(O_{COOH})}$, $Gln_{(N_{NH2})}$, and $Asn_{(N_{NH2})}$ exhibit a first peak centered at a distance in the range 2.8−3.3 Å in both APAG and PAG models; the coordination numbers of $Hpr_{(O_{OH})}$, $Lys_{(N_{NH3+})}$, $Gln_{(N_{NH2})}$, and $Asn_{(N_{NH2})}$ are 4, 2, 2, and 1, respectively, and are the same for the two simulations, whereas the APAG $Arg_{(N_{NH},N_{NH2})}$ and $Glu_{(O_{COOH})}$ rdf coordination numbers are smaller than the PAG ones (6 and 4 compared to 10 and 9). The OHp−$Asp_{(O_{COOH})}$ rdf displays a well-defined peak at 2.8 Å with coordination number of 4 in the PAG case; instead, it is centered at 3.6 Å with a coordination number of 2 in the APAG model.

The rdf's of O2 with $Arg_{(N_{NH},N_{NH2})}$, $Hpr_{(O_{OH})}$, $Lys_{(N_{NH3+})}$, $Gln_{(N_{NH2})}$, and $Ser_{(O_{OH})}$ show a first peak at a distance in the range 2.9−3.3 Å in both APAG and PAG models; the coordination numbers derived from the rdf's for O2−$Arg_{(N_{NH},N_{NH2})}$, O2−$Hpr_{(O_{OH})}$, O2−$Gln_{(N_{NH2})}$, and O2−$Ser_{(O_{OH})}$ are 13, 4, 20, and 7, respectively, in both simulations, while the O2−$Lys_{(N_{NH3+})}$ coordination number is 1 in the APAG model and 4 in the PAG one. The O2−$Asp_{(O_{COOH})}$ rdf displays, in the APAG case, a first peak at 3.2 Å with coordination number of 3, while in the other case (PAG) it is centered at 3.6 Å and has a smaller coordination number equal to 1. O2−$Glu_{(O_{COOH})}$ rdf's have first peaks at 3.4 and 3.1 Å with coordination numbers of 11 and 6 for APAG and PAG, respectively. Instead, a maximum in both APAG and PAG O2−$His_{(N_{NH},N)}$ rdf's is found at longer distances (3.5 Å) with coordination numbers of 3 and 2, respectively.

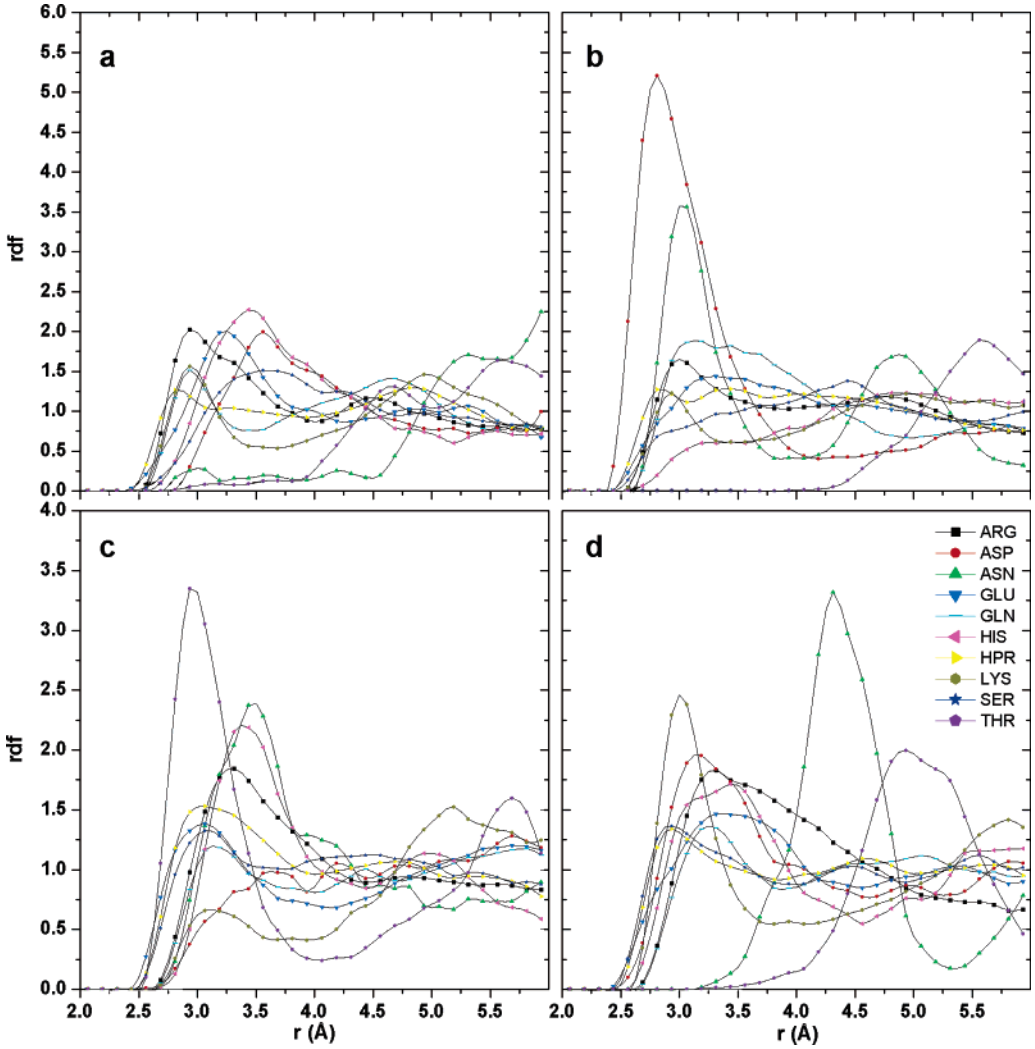

Figure 6. Radial distribution functions (rdf's) of gallic acid phenolic oxygen atoms (OHp) for APAG and PAG models (a and b) and gallic acid carboxylic oxygen atoms (O2) for APAG and PAG models (c and d) with various atoms or groups of atoms belonging to the side chains of the reported residues (see text).

This analysis suggests that both the phenolic OH groups and the carboxylic group of gallic acid molecules approach the Hpr, Arg, Gln, and Lys residues at hydrogen-bonding distance and are part of a well-defined first solvation shell coordinated with these amino acid side chains. The average total number of gallic acid molecules hydrogen bonded to the residues just mentioned is about 21 for Hpr, 19 for Arg, 15 for Glu, 6 for Gln, and 4 for Lys. These results are in agreement with our previous findings[23] and with experimental studies[36,37] suggesting that Hpr residues are among the preferred polyphenol binding sites and strong interactions also occur with Arg and Glu side chains.

3.4. Hydrogen Bonds: Formation and Mean Lifetime. Tables 4−6 show the time evolution of the number of hydrogen bonds between water molecules, gallic acid molecules, formaldehyde molecules, and protein residues during the molecular dynamics production runs. During the first 350 ps, the number of waters decreases rapidly and then reaches a steady state. This initial trend suggests that water molecules escape out of the bundles and at the same time penetrate into them. An imbalance between these two processes causes the fluctuations observed. As the molecular dynamics simulations proceed, the systems relax, and some water molecules access the protein interior making favorable interactions with its residues. The percentage of water molecules in the first solvation layer (3.5 Å), having hydrogen-bonding lifetimes shorter than 100 ps, is different for the various systems and reflects the presence and the nature of the cross-linking agent; indeed, in the PAG and APAG cases, this value is only ∼35%, whereas for the PNF and APNF models a higher percentage of ∼49%, closer to the pure water simulation (∼55%), is observed. Instead, the portion of water molecules with hydrogen-bonding lifetimes of about 1 ns is <3 for each studied system.

The behavior of formaldehyde molecules is similar to water, but stronger fluctuations are observed and a steady state is reached only after 700 ps of the production run. The percentage of formaldehyde molecules in the first solvation shell having hydrogen-bonding lifetimes shorter than 100 ps is ∼18%, while that with a hydrogen bond lifetime of 1 ns is very small (∼2%).

The number of hydrogen-bonded gallic acid molecules, however, increases initially during the first 100 ps, then, after a slight decrease, remains almost constant (only small fluctuations appear). About 70% of gallic acid molecules maintain their hydrogen bonds with collagen residues for time intervals shorter than 100 ps, and ∼10% of them have lifetimes shorter than 1 ns.

The presence of one carboxylic and three hydroxylic groups gives both proton-donor and proton-acceptor properties to gallic acid molecules, which as a result are involved in a greater and longer-lasting number of hydrogen bonds with the collagen bundles. Greater CMS surface areas are involved in the

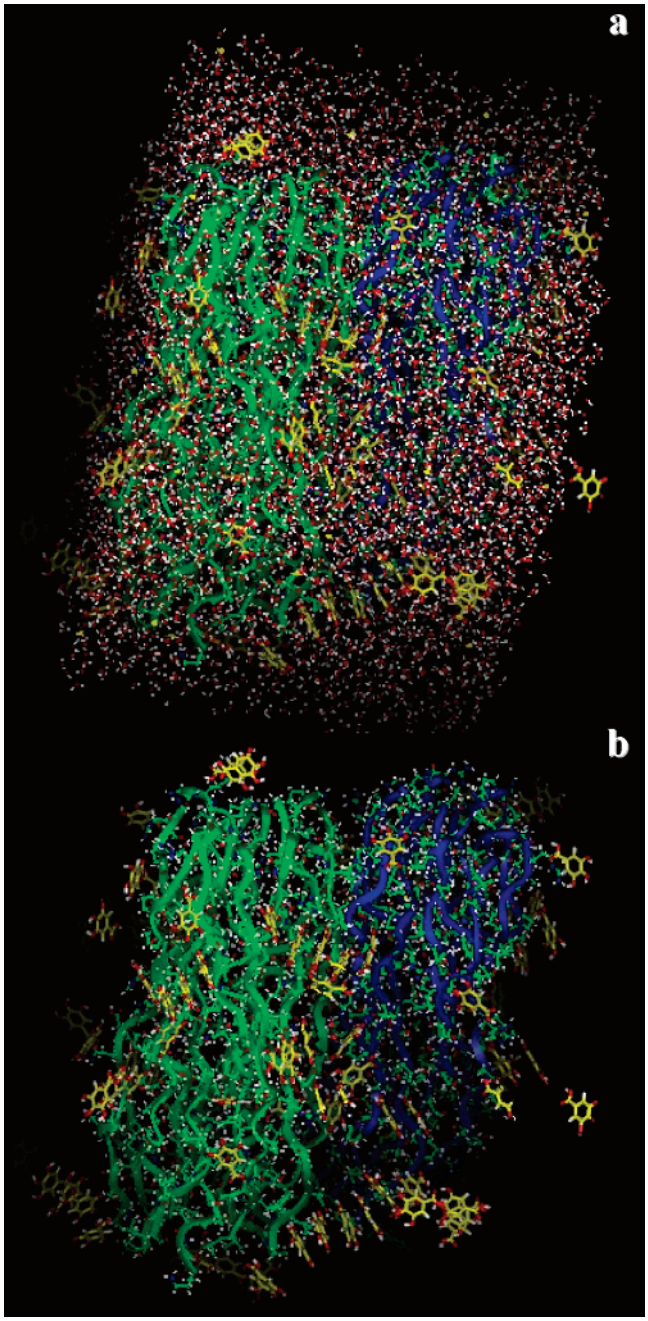

Figure 7. Snapshot of the PAG system with the two CMSs in ribbon format surrounded by (a) water and gallic acid molecules (shown as stick atoms) or (b) only gallic acid molecules (water has been omitted for clarity).

interaction and are thus inaccessible to water molecules. An example snapshot of the PAG system is shown in Figure 7.

4. Summary and Conclusions

In this paper, we have presented the first extensive atomistic molecular dynamics (MD) simulation study of the structure and dynamics of collagen-like microfibril segment (CMS) dimers in pure water and in aqueous solutions of formaldehyde and gallic acid.

Our findings show that by exploiting docking and energy minimization procedures it is possible to obtain stable supramolecular arrangements of collagen microfibril segments. All of the proposed models (APA, APN, PA, and PN), despite their structural differences and protonation states, share common features. In particular, the two CMSs are very close to each other, and the lateral chains are located at interacting distances.

The calculations also reveal that, as already pointed out in a previous paper of ours,[23] major changes in the conformation of the CMS dimer are observed when it is surrounded by pure water. By comparison of the simulations of CMS in gallic acid/water solution with those in formaldehyde/water, similar deviations from the starting conformations are observed in some cases. This suggests that the two chemicals perturb the CMS dimer in a similar way and the different behavior is mainly due to the relative arrangement of the two bundles together with their assumed protonation state. Within the time scale of the present simulations, the increase of the total solvent-exposed surface area, which is 20% for PNW, 22% for PNF, and 25% for APAG and PAG, is consistent with hydrogen-bonding capability, hydrophobic association properties, and the size and concentration of the cross-linking agent. Indeed, it is predominantly the charged and hydrophilic residues that increase their surface area in the pure water and formaldehyde/water simulations, whereas a contribution coming from the hydrophobic residues is also present in gallic acid/water solutions.

An exception to the described behavior is APNF, where larger C_α rmsd values from the starting conformation and a greater increase in the total exposed surface ($\sim$28%) are noted. These findings support the idea that the protonation state and the *complementarity* of the residues at the interbundle interface play an important role in the structure, packing, stability, and dynamics of the CMS α-helices.

The analysis of collagen/gallic acid interaction sites confirms, as already found in ref 23, that all of the oxygen atoms of the gallic acid molecules are involved in hydrogen-bonding interactions with the two CMSs. Again, as also demonstrated by experimental studies,[36,37] preferred binding to Hpr residues is shown, and strong interactions also occur between gallic acid and Arg and Glu side chains.

Acknowledgment. The authors thank Professor 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. Most of the calculations reported in this paper were performed with the resources of the CINECA supercomputer center made available to us through the "INFM Iniziativa Trasversale Calcolo Parallelo 2004".

References and Notes

(1) Smith, J. W. *Nature* **1968**, *219*, 157−158.
(2) Hulmes, D. J. S.; Miller, A. *Nature* **1979**, *282*, 878−880.
(3) Ruggeri, A.; Benazzo, F.; Reale, E. *J. Ultrastruct. Res.* **1979**, *68*, 101−108.
(4) Reale, E.; Benazzo, F.; Ruggeri, A. *J. Submicrosc. Cytol.* **1981**, *13*, 135−143.
(5) Ottani, V.; Martini, D.; Franchi, M.; Ruggeri, A.; Raspanti, M. *Micron* **2002**, *33*, 587−596.
(6) Marchini, M.; Morocutti, M.; Ruggeri, A.; Koch, M. H. J.; Bigi, A.; Roveri, N. *Connect. Tissue Res.* **1986**, *15*, 269−281.
(7) Folkhard, W.; Christmann, D.; Knoerzer, E.; Koch, M. H. J.; Mosler, E.; Nemetschek-Gansler, H.; Nemetschek, T. *Z. Naturforsch., C: J. Biosci.* **1987**, *42*, 1303−1306.
(8) Mechanic, G. L.; Katz, E. P.; Henmi, M.; Noyes, C.; Yamauchi, M. *Biochemistry* **1987**, *26*, 3500−3509.
(9) Yamauchi, M.; Chandler, G. S.; Tanzawa, H.; Katz, E. P. *Biochem. Biophys. Res. Commun.* **1996**, *219*, 311−315.
(10) Fraser, R. D. B.; MacRae, T. P.; Suzuki, E. *J. Mol. Biol.* **1979**, *129*, 463−481.
(11) Miller, A.; Tocchetti, D. *Int. J. Biol. Macromol.* **1981**, *3*, 9−18.
(12) Piez, K. A.; Trus, B. L. *Biosci. Rep.* **1981**, *1*, 801−810.
(13) Fraser, R. D. B.; MacRae, T. P.; Miller, A.; Suzuki, E. *J. Mol. Biol.* **1983**, *167*, 497−521.
(14) Hulmes, D. J. S.; Holmes, D. F.; Cummings, C. *J. Mol. Biol.* **1985**, *184*, 473−477.

(15) Fraser, R. D. B.; MacRae, T. P.; Miller, A. *J. Mol. Biol.* **1987**, *193*, 115−125.
(16) Raspanti, M.; Ottani, V.; Ruggeri, A. *J. Biol. Macromol.* **1989**, *11*, 367−371.
(17) Fratzl, P.; Fratzl-Zelman, N.; Klaushofer, K. *Biophys. J.* **1993**, *64*, 260−266.
(18) Vitagliano, L.; Nemethy, G.; Zagari, A.; Scheraga, H. A. *Biochemistry* **1993**, *32*, 7354−7359.
(19) Wess, T. J.; Hammersley, A.; Wess, L.; Miller, A. *J. Mol. Biol.* **1995**, *248*, 487−493.
(20) Hulmes, D. J. S.; Wess, T. J.; Prockop, D. J.; Fratzl, P. *Biophys. J.* **1995**, *68*, 1661−1670.
(21) Lee, J.; Scheraga, H. A.; Rackovsky, S. *Biopolymers* **1996**, *40*, 595−607.
(22) Wess, T. J.; Hammersley, A.; Wess, L.; Miller, A. *J. Mol. Biol.* **1998**, *275*, 255−267.
(23) Bronco, S.; Cappelli, C.; Monti, S. *J. Phys. Chem. B* **2004**, *108*, 10101−10112.
(24) King, G.; Brown, E. M.; Chen, J. M. *Protein. Eng.* **1996**, *9*, 43−49.
(25) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(26) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195−2199.
(27) Vakser, I. A. *Protein. Eng.* **1995**, *8*, 371−377.
(28) Vakser, I. A. *Protein. Eng.* **1996**, *9*, 37−41.
(29) Vakser, I. A. *Biopolymers* **1996**, *39*, 455−464.
(30) Martell, A. E.; Smith, R. M. *Critical Stability Constants Other Organic Ligands*; Plenum: New York, 1989; Vol. 3.
(31) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(32) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(33) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335−350.
(34) Case, D. A. et al. *AMBER 7*; University of California: San Francisco, 2002.
(35) Privalov, P. L. *Adv. Protein Chem.* **1982**, *35*, 1−104.
(36) Hagerman, A. E.; Butler, L. G. *J. Biol. Chem.* **1981**, *256*, 4494−4497.
(37) 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.