# Ionic Peptide Aggregation: Exploration of Conformational Dynamics in Aqueous Solution by Computational Techniques

**Celia Duce,[†] Susanna Monti,*[,‡] Roberto Solaro,[†] and Maria Rosaria Tiné[†]**

*Dipartimento di Chimica e Chimica Industriale, Universita degli Studi 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: September 26, 2006; In Final Form: November 7, 2006*

The effects of end groups on KEK peptide conformational characteristics and self-assembling properties in water solution are investigated by using long lasting all-atom molecular dynamics simulations. The analysis of the structural macroscopic and microscopic properties and the examination of intra- and intermolecular interactions suggest, in agreement with experimental observations, the role played by side chains and terminal regions in determining the characteristic features of the assemblages. Competition between intra- and interchain interactions greatly affects the diffusivity of peptide molecules and the conformational space that they can sample, ultimately controlling the shape, size, and distribution of the aggregate configurations. Different peptide end groups influence peptide flexibility and seem to play a crucial role in determining the aggregates' supramolecular architectures.

## 1. Introduction

The combination of nanotechnology and biotechnology promises to be a revolutionary field where nanotechniques are used to design and develop synthetic biomaterials, with the aim of realizing new biologically compatible systems suitable for sensoristic, biomedical, and health care applications. Among the several types of not cytotoxic and biocompatible substances, an extremely interesting and versatile class of compounds is represented by spontaneously assembling self-complementary ionic oligopeptides,[1−4] which are easy to design and synthesize, show unique properties, such as the ability to adopt specific 3-D structures that can be fabricated into various geometric shapes,[5−10] and thus have many potential and effective applications in the biomedical as well as pharmaceutical sectors.[11−15] As reported in the literature, self-assembling peptides, usually 8- to 16-residues long, are characterized by periodic repeats of alternating hydrophobic and ionic hydrophilic amino acids whose positions along the sequence are used to classify them into families called moduli (I, II, III, IV, etc.). Recently, short 3- to 5-residue peptides, having only hydrophilic amino acids with alternating negative and positive charges, have been designed and experimentally studied[16,17] as ligands of the fibrin $\gamma$-(312−324) epitope. Calorimetry and atomic force microscopy (AFM) measurements[16] highlighted the tendency of these polypeptides to spontaneously self-assemble in aqueous solution to form supramolecular macrostructures.

In an attempt to explain the experimental findings and to understand the atomic level mechanisms by which these peptides self-organize, a computational study on the shortest segment having different end groups, namely H−KEK−$NH_2$ and ACE−KEK−$NH_2$ (where ACE = $-COCH_3$, K = lysine, and E = glutamic acid) was performed (Figure 1).

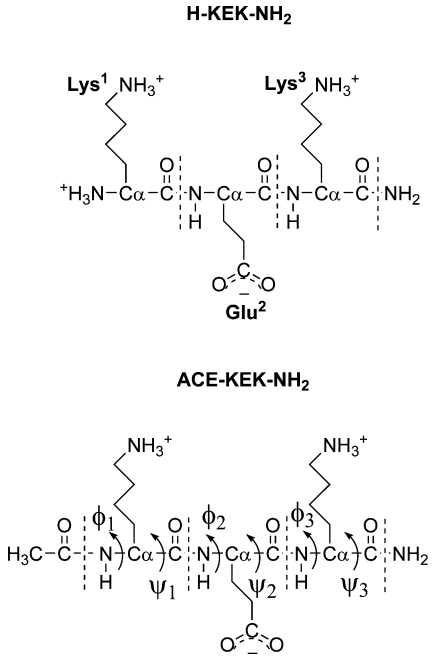


**Figure 1.** H−KEK−$NH_2$ and ACE−KEK−$NH_2$ structures and definition of the dihedral angles analyzed during the MD simulation runs.

Aggregation processes are complex dynamic events that can depend crucially on peptide specific amino acid sequence, the number of constituent units, peptide concentration, pH, and the solvent properties.[18−20] The formation of definite structures such as helical ribbons, globular aggregates, toroids, and $\beta$-sheet fibers can be preceded by formation of intermediates, oligomeric states, protofibrils, and other characteristic structures.[21] Determination of probable supramolecular architectures of self-assembling peptide molecules, intermediate structures formed during the aggregation process, and their dynamic behavior is a very challenging and extremely difficult task. The use of a concerted effort of computational chemistry methodologies and

* Corresponding author. Phone: +39-050-3152520. Fax: +39-050-3152442. E-mail: s.monti@ipcf.cnr.it.
[†] Dipartimento di Chimica e Chimica Industriale.
[‡] Istituto per i Processi Chimico−Fisici (IPCF−CNR).

10.1021/jp066307n CCC: $37.00 
Published on Web 01/11/2007

experimental techniques is a powerful means to gain a detailed molecular level understanding of the characteristics that lead to the formation of stable peptide aggregates. However, as suggested by in vivo and in vitro studies, an initial stage of an aggregation process occurs on time scales of about 1 s,[21,22] whereas further aggregation into supramolecular systems may span hours.[23] Unfortunately, these time scales cannot be simulated in a limited amount of computer time by using traditional all-atom molecular dynamics (MD) simulation techniques, which employ realistic force fields and try to describe the surrounding environment in an accurate way. The complete aggregation mechanism could only be monitored using a computationally fast and dynamically realistic technique able to extend the accessible simulation time and, at the same time, to allow rapid and accurate sampling of the conformational space of the molecules and their complexes. Recently, computationally efficient algorithms such as the discrete molecular dynamics approach (DMD),[24−26] combined with coarse-grained peptide and protein models, have been helpful and successful in revealing the general principle of folding and aggregation allowing for simulations of very long times and the study of multiprotein systems.[27−30] Unfortunately, the code is under development and is not yet available to the scientific community. However, the aim of this study is not to simulate the complete aggregation mechanism that is the large scale dynamics of peptide folding and clustering but to try to reveal detailed atomic level information both during the very early stage of the self-assembling phase as well as after the formation, in solution, of possible aggregates. Although the all-atom MD approach is computationally expensive, in comparison to simplified models, it is perhaps the only method able to capture specific aspects of amino acid interactions and to elucidate the role played by the solvent medium. Noncovalent interactions, such as hydrogen bonding, are of fundamental importance in folding and stability, and studies at the atomic level of solvated peptide systems might provide insight into structure arrangements during their aggregation and help to unravel the complex interplay between intra- and intermolecular interactions. A particularized description of all chemical species present in the simulated configurations, namely peptide molecules, counterions, and water should be able, in principle, to minimize any artificiality that might be included in the model and affect its dynamic behavior and morphology. Understanding all aforementioned effects might be of great help in the identification and design of improved self-assembling peptides that are suitable for specific applications.

## 2. Methods

**2.1. Set-Up of the Simulation System.** Modeling the system as realistically as possible and keeping it computationally tractable was the main aspect that was taken into account to choose the system size. On the basis of the experimental findings, all the carboxylic and amino groups of the studied peptides were assumed to be fully ionized.[16] The protonation states of ionizable residues were chosen according to the p$K_a$ of the isolated amino acids and a pH of 7.4. The N-terminus was considered ionized. Three different configurations were considered. The first one, **S1**, consisted of 10 H−KEK−$NH_2$ peptide molecules, 20 charge-balancing $Cl^-$ counterions, and 5512 water molecules; the concentration was about 0.1 M, and the box dimensions were 55 Å × 62 Å × 65 Å. The large number of water molecules allowed a more accurate simulation of the peptide environment and investigation of its behavior in a more dilute phase by providing a large distance between periodic images. The second configuration, **S2**, consisted of 42 H−KEK−$NH_2$ peptide molecules, 84 charge-balancing $Cl^-$ counterions, and 5680 water molecules; the concentration was about 0.3 M, and the box dimensions were 54 Å × 59 Å × 60 Å. The third configuration, **S3**, consisted of 42 ACE−KEK−$NH_2$ peptide molecules, 42 charge-balancing $Cl^-$ counterions, and 5842 water molecules; the concentration was about 0.3 M and the box dimensions were 57 Å × 59 Å × 60 Å. The initial coordinate of the tripeptides were first generated by a high-temperature MD simulation (at $T = 600$ K) in the gas phase starting from a fully extended conformation. One of the explored conformations, which did not have secondary structure (α-helix or $\beta$-sheet), was randomly selected and energy minimized by using the AMBER8 software.[31] The force field parameters for peptides were taken from the all-atom version of AMBER parm99,[32,33] which has proven to give reliable descriptions in our recent investigations on peptide−$TiO_2$ surface adsorption mechanisms.[34,35] The resulting structure was then surrounded by 9 peptides in the case of **S1** and 41 peptides in the other cases (**S2** and **S3**) in random orientations by using the xfit[36] procedure, available in the SYBYL software,[37] which finds the best fit between two molecules and can be employed to pack molecules in a cluster. For **S1**, the shift of the van der Waals distances was such that each peptide was not in contact with the other molecules, which were well separated from each other and thus formed a scattered starting configuration. On the contrary, in the more concentrated systems, peptide molecules were progressively accommodated one after the other in the space surrounding the first chosen conformation at a distance where favorable interactions took place. Each new added structure was positioned in an orientation that was the most advantageous with respect to the molecules already present so as to form a spherical cluster and to ameliorate the stability of the whole assembly. Peptide conformations that preferentially fit with the molecular bunch were extracted from a set of minimum energy structures identified by a preliminary conformational search. However, it should be taken into account that the resulting system is just one of the infinite configurations that can be used as a starting condition for the subsequent simulation runs. The built structures were then centered in a parallelepiped simulation box that was filled with TIP3P[38] water molecules whose relative orientations were obtained from the AMBER8 package. $Cl^-$ counterions were then included replacing water molecules at positions corresponding to the lowest Coulombic energy of the ion.

**2.2. MD Simulation Protocol.** The preliminary system preparation consisted in the relief of bad solute−solvent interactions. This was performed by keeping peptide molecules and $Cl^-$ counterions fixed by setting the forces acting on the atoms to zero and energy minimizing the rest of the system. In order to randomize water positions, a constant volume MD simulation (NVT ensemble) was performed, gradually increasing the temperature to 600 K during 50 ps and then gradually decreasing it to 300 K during 50 ps. The resulting configuration was equilibrated at 300 K through an additional 10 ps of a NVT molecular dynamics run. Peptide molecules and counterions were then allowed to move along with the solvent in a 40 ps NVT−MD run at 300 K. The final structure obtained was subjected to a further equilibration protocol coupling the system to a 300 K thermal bath and to a 1 atm pressure bath for a total time of 50 ps to allow for density adjustment. The dielectric constant was set equal to 1. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) method, and nonbonded interactions were truncated at 12 Å. Periodic boundary conditions were applied in all simulations to avoid

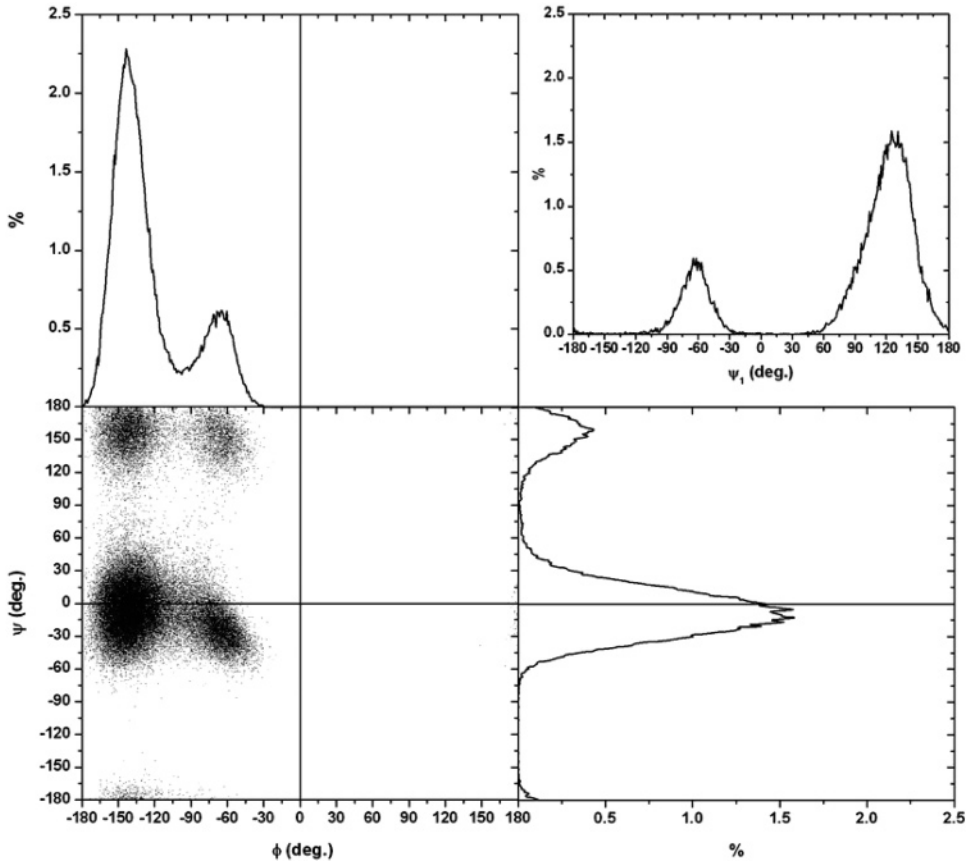


**Figure 2.** Total ($\phi$, $\psi$) Ramachandran map for **S1** peptide molecules flanked by $\phi$ and $\psi$ single probability plots. $\psi_1$ probability plot is shown in the top right corner.

edge effects, and high-frequency degrees of freedom were eliminated by constraining all bond lengths to their equilibrium values using the SHAKE[39] algorithm, which allowed a time step of 2 fs. Total simulation times were about 11 ns. The first nanosecond of each production run trajectory was considered as part of the equilibration phase and thus was excluded from the final analysis. During the production times the coordinates were saved every 10 ps and further analyzed.

The sampling methodology has been validated selecting 10 system configurations, every 10 ps, at the beginning of the production run and after 500 ps, and comparing them with the starting structure and one another. An overall measure of their drift from the initial configuration and of their reciprocal deviations was provided by the root mean square difference (rmsd) obtained by superimposing peptide heavy atoms. The average value of the rmsd was about 3.9 Å with large fluctuations, thus confirming that the system was not trapped into a local state. Consecutive sampled structures were quite different from each other (the rmsd was about 1.5 Å), suggesting that a 10 ps sampling interval was appropriate; moreover, they appeared to have little correlation with the starting configuration, suggesting that perhaps the starting configuration has little effect on the long-time dynamics of the systems.

## 3. Results and Discussion

To describe the behavior of H−KEK−$NH_2$ and ACE−KEK−$NH_2$ in solution, a set of structural parameters that could provide a general idea of their folding and assembling properties was analyzed in detail. Atomic positions collected every 10 ps were used to build radial pair distribution functions (RDFs) chosen to depict the arrangement of water and $Cl^-$ counterions around solute atoms.

To locate each peptide molecule with respect to the other peptide structures, center of mass pair distribution functions (peptide−peptide RDFs) were reported. Peptide structural behavior in solution was delineated through the ($\phi$, $\psi$) dihedral angles of the backbone chain, displayed both as single distribution probabilities and as Ramachandran plots ($\phi$ vs $\psi$), which are useful schemes to directly visualize the secondary structure content.

Size and shape of both peptides and peptide clusters were classified according to the values taken by the radius of gyration, $R_{gyr}$, defined as

$$R_{gyr} = \sqrt{\frac{\sum_i m_i (r_i - r_{cm})^2}{\sum_i m_i}} \quad (1)$$

where $m_i$ is the mass of the atom $i$ located at a radial distance $r_i$ and $r_{cm}$ is the radial position of the center of mass of the considered system; the ratio of the largest to the smallest principal moments of inertias, $I_{max}/I_{min}$ and the eccentricity, $\eta$, is defined as

$$\eta = 1 - \frac{I_{av}}{I_{max}} \quad (2)$$

where $I_{av}$ is the average of the three principal moments of inertia. These latest parameters provide a measure of the deviation from a spherical shape where $I_{max}/I_{min} = 1$ and $\eta = 0$, whereas for flat or needlelike profiles $\eta \rightarrow 1$.

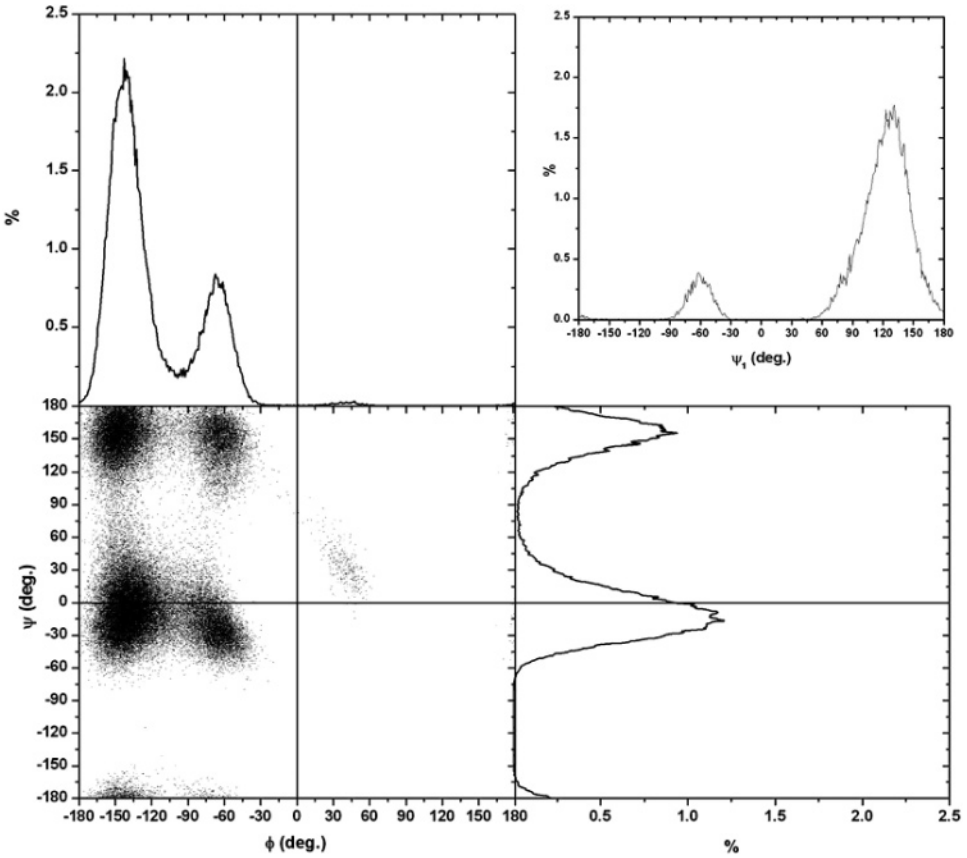


**Figure 3.** Total ($\phi$, $\psi$) Ramachandran map for **S2** peptide molecules flanked by $\phi$ and $\psi$ single probability plots. $\psi_1$ probability plot is shown in the top right corner.

Peptide self-association was characterized through the analysis of the hydrogen-bonding network, identifying potential hydrogen bonds when donor−acceptor distances were smaller than 3.3 Å and when the angle formed by the hydrogen-donor−acceptor triplet was smaller than 30°. Hydrogen-bond persistency was expressed via their percentage of occupancy, which was defined as the number of frames with the hydrogen bond present divided by the total number of frames used for the analysis. The lifetime of a hydrogen bond was calculated as the time elapsed from its first appearance until it was first broken. The average lifetime of a hydrogen bond during the simulation was then calculated as the average of all of its occurrences excluding those with a lifetime less than 30 ps.

As peptide molecules had several different ways to interact with each other, any definition of an aggregate in terms of number of contacts or distances between peptide centers of mass would have been less selective in identifying possible clusters. Therefore, we decided to use a more specific criterion based on potential hydrogen-bond formation, considering 4 Å as the maximum distance between the heavy atoms involved in hydrogen-bonding interactions. Two peptides were defined to be in the same aggregate if the distance between any one of their hydrogen-bond donors and acceptors was smaller than the maximum distance for a period longer than 30 ps. The aggregate was considered to have fallen apart into separate clusters or units if the distance was greater than the chosen cutoff distance for more than 30 ps. The distance and lasting time period were chosen on the basis of data found in the literature[40,41] and visual inspection of the trajectories, which showed both aggregation and disaggregation events. The chosen parameters proved to reproduce satisfactorily the visualized data.

On the basis of this definition, both **S2** and **S3** starting configurations consisted of a big spherical aggregate formed by all the 42-peptide molecules present in the system. The total surface area available for interaction with water was quantified by summing contact areas obtained by rolling over the surface of the cluster a probe molecule with a radius equal to 1.4 Å.[42]

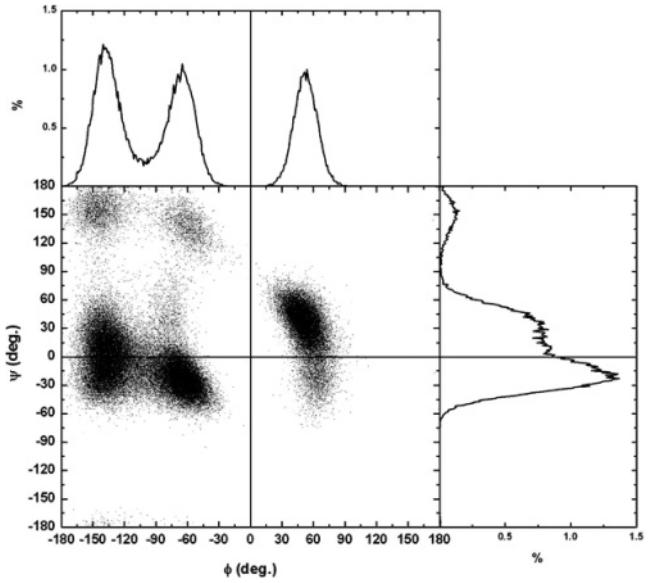


**Figure 4.** Total ($\phi$, $\psi$) Ramachandran map for **S3** peptide molecules flanked by $\phi$ and $\psi$ single probability plots.

**3.1. Dihedral Angle Distributions.** Peptide molecules were not conformationally rigid but slowly evolved in their conformational space folding into different secondary structures, which were identified by the analysis of ($\phi$, $\psi$) Ramachandran maps.

As can be seen in the total Ramachandran plots (Figures 2−4), H−KEK−$NH_2$ and ACE−KEK−$NH_2$ tripeptides exhibit

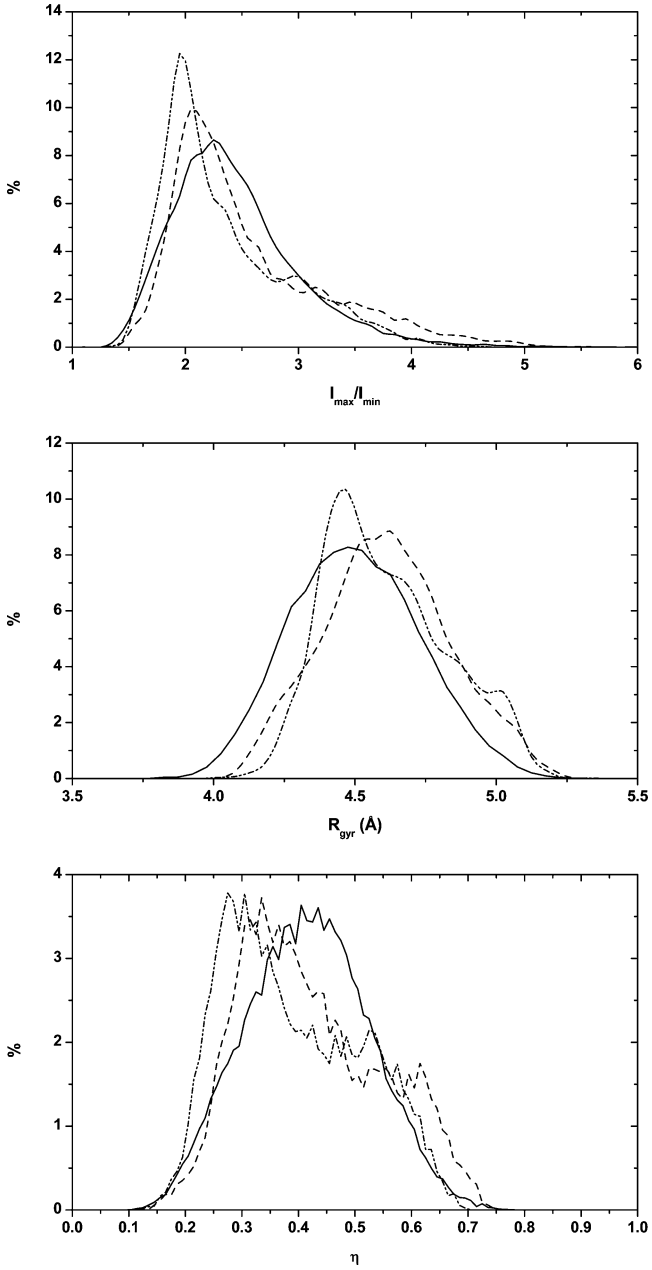


**Figure 5.** Normalized probability distributions of $I_{max}/I_{min}$, $R_{gyr}$ (Å), and $\eta$ for **S1**(solid line), **S2**(dashed line), and **S3**(dash dot dot line) systems.

a different degree of flexibility. ACE−KEK−$NH_2$ presents a considerably larger conformational variability than H−KEK−$NH_2$, occupying the $\beta$, right-handed α-helix ($\alpha_R$) and left-handed α-helix ($\alpha_L$) regions of the map, whereas only $\beta$ and $\alpha_R$ zones are explored in the case of H−KEK−$NH_2$.

The final analysis, displaying single residue plots (Supporting Information), shows that, for **S2**, $Glu^2$ and $Lys^3$ remain confined to two quadrants of the map throughout the simulation, namely the $\beta$-region and the $\alpha_R$-quadrant. A very similar behavior is noticed for the dilute system (**S1**), which presents a Ramachandran map almost equal to the **S2** one with four distinct high probability peaks in the [(−180°, −120°); (−60°, 45°)], [(−90°, −45°); (−60°, 15°)], [(−180°, −120°); (120°, 180°)], and [(−90°, −45°); (120°, 180°)] intervals. The only small difference is due to the presence, in **S2** maps, of a very limited number of structures with ($\phi_3$, $\psi_3$) pairs in the $\alpha_L$ region, which are presumably generated during self-interaction processes. As far as **S3** is concerned, all of the residues ($Lys^1$, $Glu^2$, and $Lys^3$) explore the four quadrants of the Ramachandran map concentrating their values in the $\beta$- and α-structured regions but with a marked preference for the [(−150°, −120°); (−45°, 45°)], [(−90°, −30°); (−60°, 15°)], and [(30°, 75°); (15°, 75°)] areas.

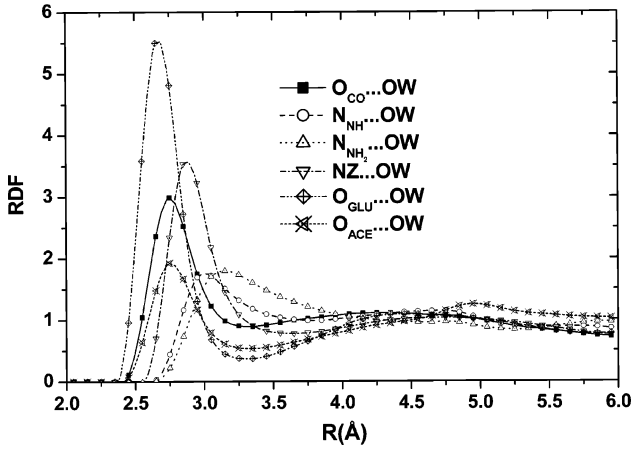


**Figure 6.** Carbonyl oxygen ($O_{CO}$)···water oxygen (OW), carboxyl oxygen ($O_{GLU}$)···OW, amidic nitrogen ($N_{NH}$)···OW, Lys side-chain nitrogen (NZ)···OW, terminus nitrogen $N_{NH_2}$···OW, and terminus carbonyl oxygen ($O_{ACE}$)···OW RDFs for the simulated systems.

Although it has been postulated that left-handed helical amino acids would have unfavorable local steric interactions, an appreciable population of $\alpha_L$-residues has been found in the **S3** simulation but not in the **S2** one, suggesting that the uncharged ACE termination could be responsible of the formation of favorable left-handed arrangements, whose existence has been demonstrated in the literature.[43,44]

The analysis of the ($\phi$, $\psi$) pairs indicates that although the studied peptides adopt a great variety of conformations none of them corresponds closely to an ordered structure, suggesting that these data should be complemented by other global or local descriptors that will be discussed in the following sections.

**3.2. Global Descriptors: $R_{gyr}$, $I_{max}/I_{min}$, $\eta$.** Macroscopic descriptors, able to provide a comprehensive picture of size and shape variation of these short peptides, are represented by the radius of gyration, $R_{gyr}$, the mass weighted rms average distance of all atoms in a peptide from its center of mass; the ratio $I_{max}/I_{min}$; and the eccentricity $\eta$. The normalized probability distributions are shown in Figure 5 for the three systems. In the **S1** case, peptide molecules are well surrounded by water during the whole simulation time and get in touch with each other only on sporadic occasions and for a few picoseconds. The **S1** probability distribution can be fitted with single broad Gaussian functions with peaks at 2.3 and 0.4 for $I_{max}/I_{min}$ and $\eta$, respectively. This result indicates that a great variety of ellipsoidal shapes, which are generated by the highly mobile charged Lys and Glu side chains, can be found in dilute solutions. As expected, $R_{gyr}$ remains centered at 4.6 Å in all simulated systems. However, at higher peptide concentration (**S2** and **S3** systems), packing effects and intermolecular hydrogen bonds affect the shape of single tripeptide molecules. Contrary to the dilute system, $I_{max}/I_{min}$, as well as $\eta$ probability distributions peaks, are shifted to lower values (about 2.0 for $I_{max}/I_{min}$ and about 0.3 for $\eta$), indicating the tendency to form less elongated structures.

**3.3. RDFs.** The atom−atom radial distribution functions of water around each hydrogen-bond donor and acceptor atom of the investigated peptides are shown in Figure 6. Pronounced first neighbor peaks and a weak tendency for second neighbors peaks can be observed in all plots. The first shell peak is located at different distances for positively, negatively, and uncharged

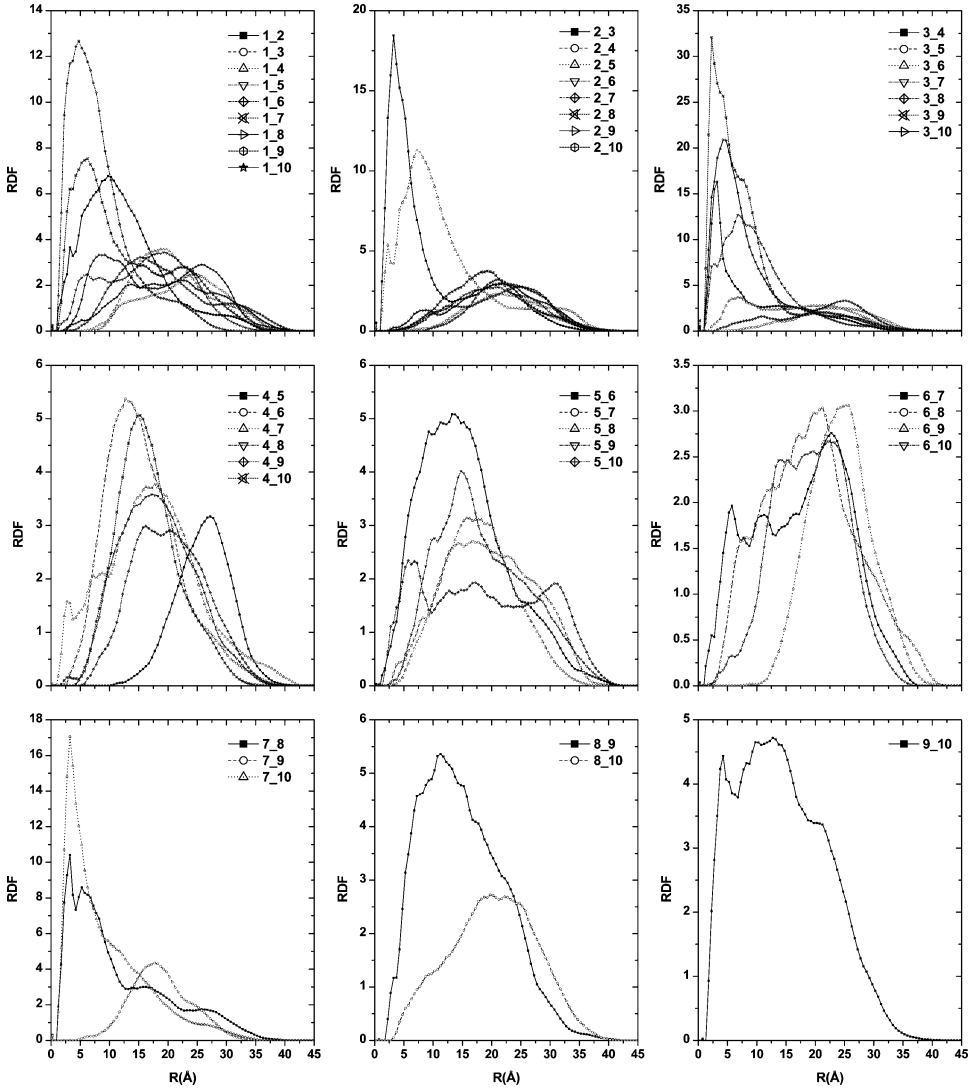


**Figure 7.** Peptide−peptide radial distribution functions for **S1**. (For convenience, the ten peptide molecules have been numbered from 1 to 10.)

groups of peptide, in agreement with the results of Komeiji et al.,[45] who proved the aforementioned effect in the case of protein structures. The carbonyl oxygen ($O_{CO}$)···water oxygen (OW) RDF has a first peak centered at 2.75 Å with an average coordination number of 1.7, whereas the carboxyl oxygen ($O_{GLU}$)···OW RDF shows a sharper first neighbor peak at shorter distances, 2.67 Å, with a greater coordination number of 2.5. As far as hydrogen-bond donors are concerned, the first peak of amidic nitrogen ($N_{NH}$)···OW and of the lysine side-chain nitrogen (NZ)···OW RDFs is located at 3.00 and 2.87 Å with coordination numbers of 1.5 and 4.3, respectively. The $N_{NH_2}$···OW RDF displays a trend similar to the $N_{NH}$···OW one, but the center of the broad first hydration shell peak is shifted at a larger distance (3.15 Å), and integration of this peak up to the first minimum gives an average coordination number of 7.5 water molecules. The simulation results indicate that the majority of $Cl^-$ ion density is located very close (∼3.35 Å) to peptide positively charged groups at higher concentration; the $Cl^-$ density distribution is more homogeneous in dilute solution. In order to identify possible aggregated peptide structures, in the **S1** case, peptide−peptide RDFs were computed and analyzed (Figure 7). To obtain the RDF plots, the two peptides under examination were treated as solute and solvent, respectively, and then for each solvent atom, the distance to the closest solute atom was measured. As a consequence, when first neighbor peaks are at distances shorter than 5 Å, peptides are in contact with each other, and most probably, they are hydrogen-bonded. Examination of the RDF plots in Figure 7 suggests that the vast majority of peptide pairs are well separated from each other during the whole simulation time. Only a few RDFs display sharp peaks at hydrogen-bonding distances, namely 1−2, 1−6, 2−3, 3−10, 4−7, 6−7, and 7−10. However, these contacts are not stable and have a large degree of fluctuation, indicating that the aggregation is a highly dynamic process in which the peptides constantly associate and dissociate. Moreover, these results suggest that the initial formation of disordered peptide aggregates is principally driven by side-chain−side-chain electrostatic interactions and to a minor extent by backbone−side-chain hydrogen bonds, whereas backbone−backbone interactions are rarely observed.

**3.4. Intramolecular Hydrogen Bonds.** The analysis of intramolecular hydrogen bonds in solution can be useful to understand preorganization that would yield favorable or unfavorable arrangements to the formation of supramolecular aggregates. Examination of the **S1** simulation reveals that all peptide molecules have intramolecular hydrogen bonds, which are mainly formed between backbone and side-chain atoms (67%) with an average percentage of occupancy (APO) of about 13%. About 22% of the total intramolecular hydrogen bonds are formed by Lys and Glu sidechains (with APO ≈ 9%), whereas only 15% involve the two terminal regions ($-NH_3^+$

**TABLE 1: Cluster Structure Descriptors**[a]

| system | cluster | Pn | ASA | H-bonds | $I_{max}/I_{min}$ | $\eta$ | $R_{gyr}$ |
|---|---|---|---|---|---|---|---|
| **S1** | I | 2 | 774 | 2 | 2.8 | 0.51 | 7.0 |
| | II | 2 | 824 | 1 | 2.4 | 0.45 | 6.6 |
| | III | 2 | 814 | 1 | 3.1 | 0.56 | 6.8 |
| | IV | 2 | 779 | 2 | 3.0 | 0.55 | 7.4 |
| | V | 2 | 795 | 1 | 1.9 | 0.35 | 6.2 |
| | VI | 2 | 805 | 2 | 2.9 | 0.52 | 6.9 |
| | VII | 2 | 803 | 1 | 2.4 | 0.45 | 6.6 |
| | VIII | 2 | 777 | 2 | 4.2 | 0.67 | 7.5 |
| | $t = 0$ | 33 | 12472 | 57 | 1.3 | 0.11 | 18.1 |
| **S2** | I | 16 | 5448 | 38 | 1.7 | 0.26 | 14.5 |
| | II | 8 | 2968 | 9 | 4.3 | 0.68 | 14.8 |
| | III | 3 | 1149 | 3 | 2.6 | 0.47 | 9.2 |
| | IV | 2 | 725 | 1 | 1.4 | 0.18 | 5.6 |
| | $t = 0$ | 28 | 9411 | 68 | 1.4 | 0.19 | 17.8 |
| **S3** | I | 13 | 4509 | 34 | 2.2 | 0.43 | 14.9 |
| | II | 7 | 2367 | 20 | 1.4 | 0.17 | 9.6 |
| | III | 3 | 1075 | 6 | 3.6 | 0.62 | 8.1 |
| | IV | 3 | 1089 | 7 | 6.7 | 0.79 | 9.8 |
| | V | 2 | 715 | 5 | 1.4 | 0.17 | 5.4 |
| | VI | 2 | 695 | 5 | 1.4 | 0.15 | 5.2 |

[a] Number of peptide molecules forming the aggregate (Pn); accessible surface area (ASA) in Å$^2$; number of hydrogen bonds (H-bonds) with a percentage of occupancy greater than 10%; ratio of the largest to the smallest principal moments of inertia ($I_{max}/I_{min}$); eccentricity ($\eta$); radius of gyration ($R_{gyr}$) in Å. All the reported quantities are average values.

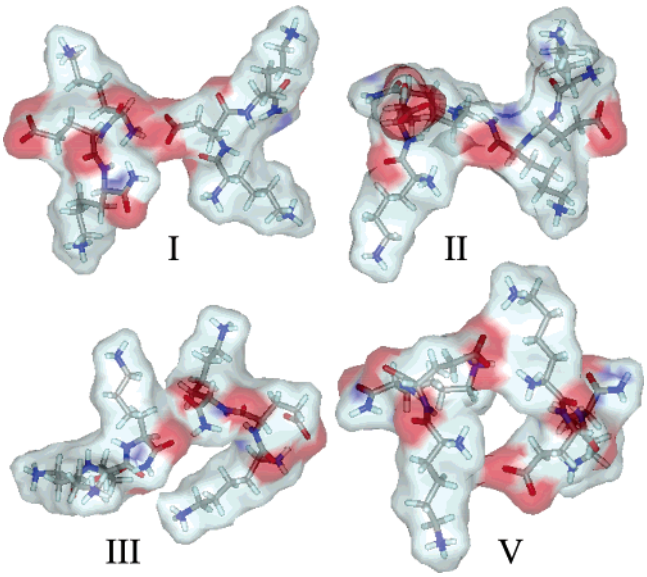


**Figure 8.** Molecular surface representation of the aggregate structures found during the **S1** production run (for their descriptors, see Table 1) colored by atom type (oxygen in red, nitrogen in blue, carbon in gray, and hydrogen in cyan). Water is omitted for clarity.

and $-NH_2$ groups). All the other H-bonds are less persistent, having APO values lower than 7%.

In more concentrated solutions (**S2**), 57% of the peptides show intramolecular hydrogen bonds, and as in the **S1** case, backbone−side-chain interactions are decidedly preferred (65%); their persistence (29%) is greater than the value found for the **S1** system because of the lower mobility of side-chain groups and of the reduced number of interactions with water molecules. The analysis shows that the side chains change their orientation and, as a result, remain in confined regions for quite a long time, intermittently breaking and reforming hydrogen-bonding contacts with neighboring backbone atoms. A lower number of intramolecular H-bonds (29%) are formed instead between $NH_2$ terminus and backbone or side-chain atoms with a marked preference for backbone carbonyl oxygens, which have an APO equal to 12%. Intramolecular side-chain−side-chain and backbone−backbone H-bonds are less populated and short lasting. Both of them are about 3% of the total population, and their APO is lower than 7%.

A different behavior is observed for the ACE−KEK−$NH_2$ sequence, where 50% of peptide structures have intramolecular H-bonds, and more favored interactions (52%) are those involving both terminal regions. They can be terminus−side-chain, terminus−backbone, and terminus−terminus interactions, but their duration is short (APO = 15%). Intramolecular H-bonds formed between backbone atoms such as Lys(N)···Lys(O) interactions are populated enough (26%), but their APO is lower than 8%. Longer lasting intramolecular H-bonds (APO = 29%) are those formed by side-chain−side-chain motion, but they represent only 7% of the total.

Fairly strong and persistent intramolecular H-bonds are observed in **S2** and **S3** simulated systems. However, considering population and persistency at the same time, it emerges that the **S2** system has a higher tendency to intramolecular group association and thus has reduced conformational freedom. Due to their structural features, Glu and Lys sidechains are able to form intramolecular H-bonds with each other and with the polar moieties of the backbone. Once such bonds are formed, the hydrogen-bonded backbone centers are no longer available for the through-backbone interactions necessary to stabilize particular conformations. Such H-bonds are almost always present, and occasionally many of them exist at the same time. The most common ones are those formed between backbone and side-chain atoms in the case of H−KEK−$NH_2$ peptides, whereas those formed by terminal groups and side chains or backbone atoms and to a lesser extent by backbone−backbone interactions are preferred for ACE−KEK−$NH_2$ peptides. All other types are equally marginal. Intramolecular H-bonds are constantly formed and broken, and at any given time, only a fraction of all possible bonds of a certain type exists as a result of the competition with surrounding peptides and above all with the surrounding water molecules that can be engaged in intermolecular hydrogen bonds. However, these bonds are short-lived because of the high water diffusivity.

**3.5. Cluster Structures and Intermolecular Hydrogen Bonds.** In the scattered starting configuration (**S1**), all peptide molecules appear somewhat randomly dispersed within the simulation box during the whole production run. Although the tendency to evolve into aggregate configurations is evidenced by the presence of small and unstable clusters, mainly made of two and sometimes three tripeptide molecules, the simulation time was not sufficiently long to observe appreciable aggregation. However, the preliminary formation of small clusters indicates the beginning of peptide aggregation, which seems a very slow process driven, according to the obtained results, by long-range attractive interactions of the oppositely charged Lys and Glu side chains.

Examination of the sampled configurations highlights the frequent formation of rather intermittent side-chain−side-chain intermolecular hydrogen bonds, which cause a rearrangement of peptide reciprocal positions and the appearance of new intermolecular interactions among charged side-chain groups and polar backbone atoms, which represent 50% of the total hydrogen-bond population. Backbone−backbone hydrogen bonds are not observed; instead, competition with water molecules leads to recurring bond formation and breakage and hence to the observed intermittency in the hydrogen-bonding characteristics. Morphology parameters of the aggregates identified during the course of the simulations are reported in Table 1, and some

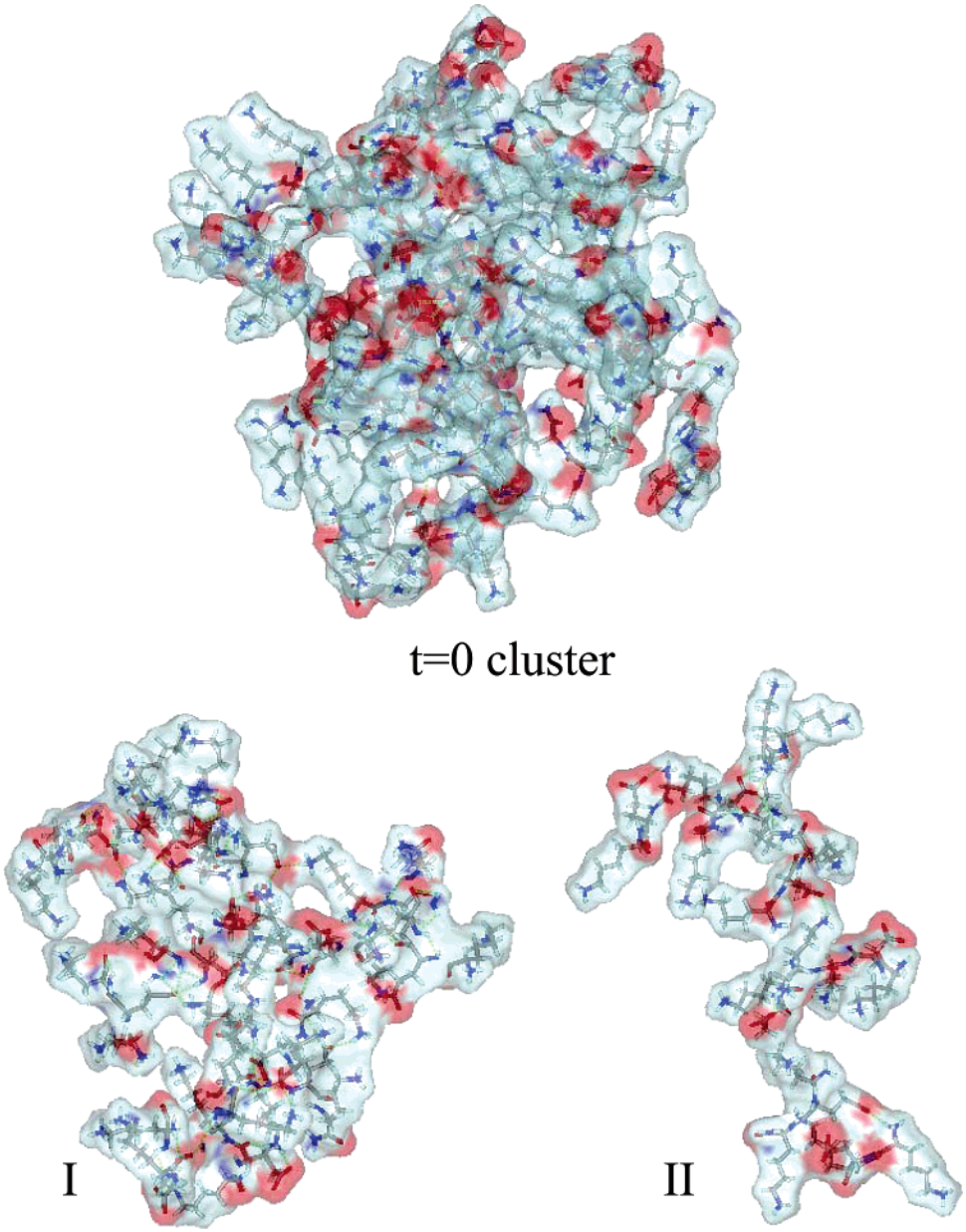


**Figure 9.** Molecular surface representation of the aggregate structure found at the beginning of the **S2** production run ($\mathbf{S2}_{t=0}$) and equilibrium structures of the two biggest identified clusters (I and II), which appeared after 500 ps and remained stable during the remainder of the simulation time (for average descriptors, see Table 1). Molecular surfaces are colored by atom type (oxygen in red, nitrogen in blue, carbon in gray, and hydrogen in cyan). Water is omitted for clarity.

of them are displayed in Figures 8−10. All sampled clusters are dimers with marked ellipsoidal shape, and no preferred spatial arrangement of the constituting monomers, which are organized in a variety of orientations due to their high flexibility in water solution, was observed. The average radius of gyration is 6.9 Å, whereas visual inspection of the conformations and examination of eccentricity and $I_{max}/I_{min}$ values determine that different ellipsoidal shapes far from a perfectly smooth ellipsoidal body can be assumed by the self-assembled structures. Indeed, the selected clusters have a large amount of surface roughness able to host water molecules and appear to be well solvated. This effect does not favor their respective association and thus the formation of bigger aggregates. The prevalence of dimers suggests that these species play the potential role of initial nuclei for the formation of larger aggregates, as confirmed by the occasional presence of trimers. Assemblages constantly formed and dissolved, enabling peptides to repack and allowing the individual peptides to undergo conformational changes. However, the rare appearance of larger assemblies suggests that the two peptides in close proximity to each other were not yet correctly associated to activate cooperative aggregation. On the other hand, reorganization and reorientation in active nuclei can be a very slow process, especially in this case (**S1**) where, due to the total charge of each peptide ($+2e$), specific intermolecular interactions develop more slowly. Noticeable aggregation could not be observed in the chosen time scale.

As far as **S2** and **S3** systems are concerned, the examination of peptide intermolecular hydrogen-bonding networks reveals that the aggregated starting configurations evolve into different aggregated structures dynamically changing their peptide content and location. Starting assemblages appear to have little correlation with the final clusters, and despite the higher peptide-to-water ratio with respect to **S1**, dispersion of peptides as monomers and also dimers is observed. Four and six aggregates

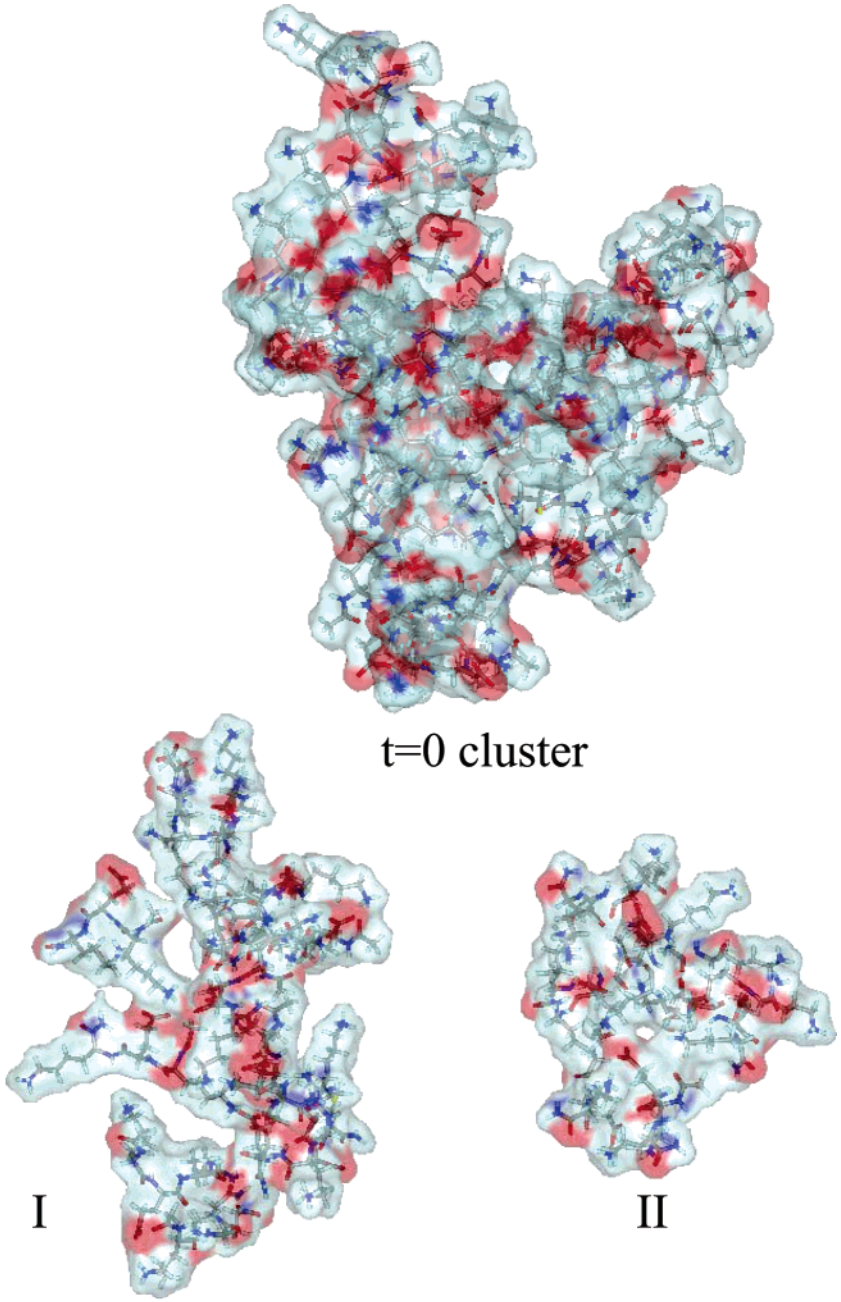


**Figure 10.** Molecular surface representation of the aggregate structure found at the beginning of the **S3** production run ($\mathbf{S3}_{t=0}$) and equilibrium structures of the two biggest identified clusters (I and II), which appeared after 500 ps and remained stable during the remainder of the simulation time (for average descriptors, see Table 1). Molecular surfaces are colored by atom type (oxygen in red, nitrogen in blue, carbon in gray, and hydrogen in cyan). Water is omitted for clarity.

of various size have been identified in **S2** and **S3**, respectively, and their average descriptors are reported in Table 1 together with those of the clusters present at the beginning of the production run (Figures 9 and 10). Peptide motion and diffusion into the solvent can be observed in Figure 11 and are better estimated in Figure 12 where the time evolution of the distance (center of mass−center of mass distance) between peptide molecules belonging to the core of $\mathbf{S2}_{\mathrm{I}}$ is compared with the time evolution of the distance between one peptide of the core and a neighboring molecule. The majority of intermolecular H-bonds are of the side-chain−side-chain type, being 50% of the total intermolecular hydrogen-bond population in both **S2** and **S3** systems. Backbone−backbone interactions are 15%, whereas side-chain−backbone interactions represent 32% and 18% of the detected intermolecular H-bonds for **S2** and **S3**, respectively. The normalized probability distributions of the percentage of occupancy of those interpeptide hydrogen bonds show that the most probable values are in the 10−30% range, suggesting that the pattern of H-bonding fluctuates dynamically, as expected, given the fluctuating nature of the environment. This behavior confirms that also the aggregated starting structures (**S2** and **S3**) are quite mobile, evolving continuously under the influence of hydrodynamic effects. The bigger clusters found at the beginning of the production runs are made of about 30 peptide molecules each; they are almost spherical and have a greater number of interpeptide hydrogen bonds with respect to the aggregate configurations identified during the course of the simulations. The aggregate found at the beginning of the **S3** production run ($\mathbf{S3}_{t=0}$) is more compact than the one detected in the **S2** system ($\mathbf{S2}_{t=0}$); indeed, peptide molecules are closer

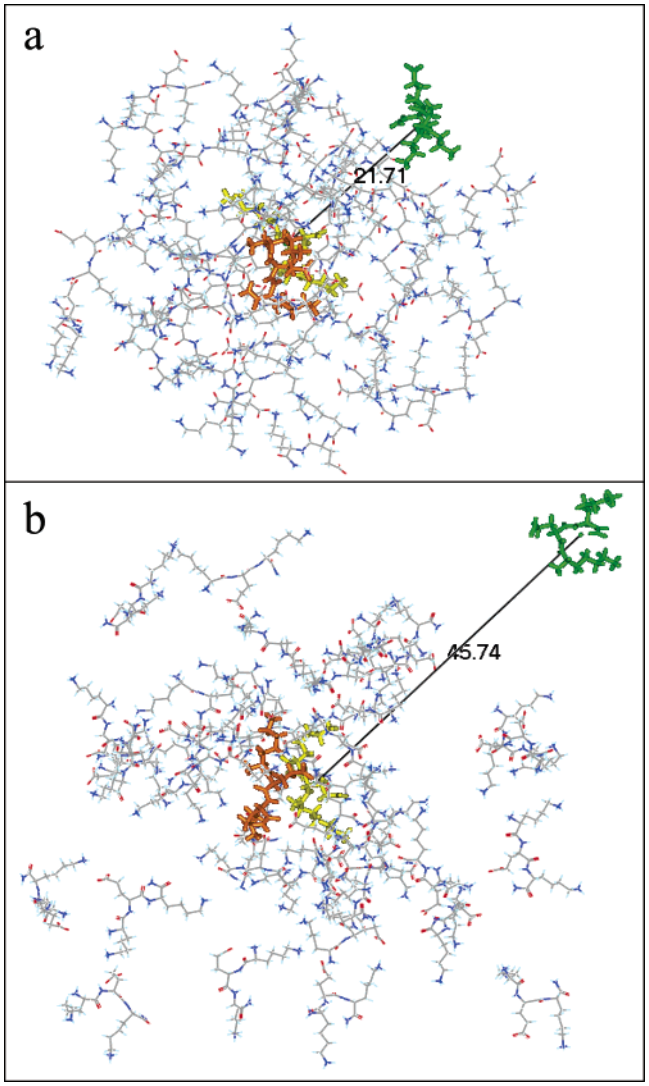


**Figure 11.** Initial (a) and final (b) configurations of the **S2** system. Two peptides belonging to the core of the larger cluster (**S2**$_l$; see Table 1), identified during the simulation, are represented as yellow and orange stick models. One of the neighboring peptide molecules, which moves away from the aggregate and diffuses into the solvent, is evidenced by green sticks, and the distance of its center of mass from the center of mass of the yellow peptide is reported. Water is omitted for clarity.

to each other and more tightly packed due to a denser network of intermolecular hydrogen bonds. The degree of resemblance between the **S3**$_{t=0}$ cluster and a spherical body is lower than that of the **S2**$_{t=0}$ assembly, which has a greater accessible solvent area (by about 3000 Å$^2$) due to frequent gaps between its peptide units, able to accommodate water and $Cl^-$ counterions. The biggest aggregates (**S2**$_l$, **S3**$_l$) originated from the **S***X*$_{t=0}$ clusters after about 500 ps, remain stable during the remainder of the simulation time, have ellipsoidal shapes that do not deviate too much from sphericity, and their peptide members are attached in disordered orientations. However, the arrangement of **S3** monomers is more compact than that of **S2** ones. A tendency to increase peptide spatial organization by reorienting the constituting units such as to form a greater number of main-chain intermolecular hydrogen bonds is also observed. On the contrary, smaller clusters tend to form elongated structures that slowly grow in size due to the dynamic process of association and dissociation. From the simulation results, it can be speculated that ordered ACE−KEK−$NH_2$ aggregates might be

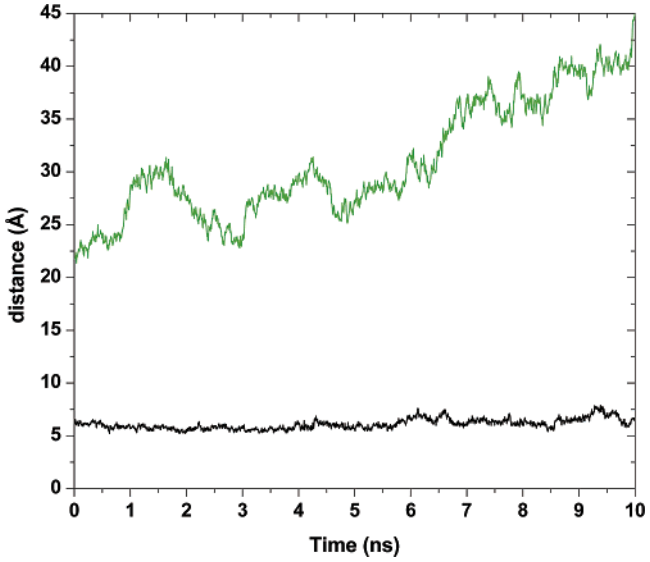


**Figure 12.** Distance between the centers of mass of peptide molecules belonging to the core of the **S2**$_l$ cluster as a function of the simulation time (black line). Their structures are showed as yellow and orange stick models in Figure 11. The green line represents time evolution of the distance between the center of mass of one of the aforementioned peptides (yellow sticks in Figure 11) and the center of mass of a neighboring peptide molecule (green sticks in Figure 11), which escapes into the solvent.

obtained through side-to-side association of peptide molecules by spatial rearrangement of the disordered oligomers.

## 4. Conclusions

Atomistic MD simulations of few peptide molecules give a possible picture of the very early steps of the aggregation process and suggest the role played by side chains and terminal regions in determining the characteristic features of the assemblages.

Several insights have been provided into the nature of peptide−peptide interactions. From the simulation results, we can speculate that intra- and interpeptide hydrogen bonds limit the diffusivity of peptide molecules and the conformational space that they can sample, ultimately controlling the shape, size, and distribution of the aggregate configurations. Peptide backbone secondary structures fluctuate through generation and breakage of intramolecular hydrogen bonds, which enable the formation of intermolecular hydrogen bonds by the side-chain charged groups. Both H−KEK−$NH_2$ and ACE−KEK−$NH_2$ have the tendency to form mainly side-chain-side-chain intermolecular hydrogen bonds, and charged groups interactions seem to dictate the structure, orientation, and position of each peptide molecule inside the aggregates, which are characterized by ellipsoidal shapes. Asphericity of the clusters, peptide mobility, and dynamic fluctuation of the H-bonding interactions appears to contribute to the creation of larger assemblages. The small clusters that appear during the simulations need much time to relax toward their equilibrium structures, which represent possible intermediate configurations of larger supramolecular associations. The nature of the peptide end groups and their interactions seem to be fairly important in determining the aggregates characteristics. The computational results, in agreement with the experimental data,[16,19,20] suggest that different terminations can affect the degree of flexibility of peptide molecules and thus their plasticity and ability to accommodate conformational changes upon binding, determining self-organization in different supramolecular architectures.

However, identification of final stable macroaggregates should require simulation times too long to be achievable through all-atom MD techniques, and more simplified and efficient models, like the one used by Dokholyan and colleagues, will further help us, hopefully, in future work, to extend system size and simulation time, allowing the individualization of the features necessary to design specific self-assembling peptide sequences.

**Acknowledgment.** Most of the calculations reported in this paper were done on the resources of the CINECA Supercomputer Center made available to us through the Progetti Supercalcolo 2006-Fisica della Materia.

**Supporting Information Available:** Ramachandran Maps. This material is available free of charge via the Internet at http://pubs.acs.org.

## References and Notes

(1) Zhang, S.; Holmes, T. A.; Lockshin, C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3334−3338.

(2) Zhang, S.; Lockshin, C.; Cook, R.; Rich, A. *Biopolymers* **1994**, *34*, 663−672.

(3) Zhang, S.; Holmes, T. A.; DiPersio, C. M.; Hynes, R. O.; Su, X.; Rich, A. *Biomaterials* **1995**, *16*, 1385−1393.

(4) Zhang, S.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 23−28.

(5) León, E.; Verma, N.; Zhang, S.; Lauffenburger, D. A.; Kamm, R. M. *J. Biomater. Sci., Polym. Ed.* **1998**, *9*, 297−312.

(6) Thirumalai, D.; Klimov, D. K.; Dina, R. I. *Curr. Opin. Struct. Biol.* **2003**, *13*, 146−159.

(7) Caplan, M. R.; Lauffenburger, D. A. *Ind. Eng. Chem. Res.* **2002**, *41*, 403−412.

(8) Fernandez-López, S.; Kim, H. S.; Choi, E. C.; Delgado, M.; Granja, J. R.; Khasanov, A.; Kraehenbuehl, K.; Long, G.; Weinberger, D. A.; Wilcoxen, K. M.; Ghadiri, M. R. *Nature* **2001**, *412*, 452−455.

(9) Zhang, S.; Marini, D. M.; Hwang, W.; Santoso, S. *Curr. Opin. Chem. Biol.* **2002**, *6*, 865−871.

(10) Yokoi, H.; Kinoshita, T.; Zhang, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8414−8419.

(11) Nowak, A. P.; Breedveld, V.; Pakstis, L.; Ozbas, B.; Пne, D. J.; Pochan, D.; Deming, T. J. *Nature* **2002**, *417*, 424−428.

(12) Hartgerink, D. J.; Beniash, E.; Stupp, S. I. *Science* **2001**, *294*, 1684−1687.

(13) Vanthey, S.; Santos, S.; Gong, H.; Watson, N.; Zhang, S. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5355−5360.

(14) Ellis-Behnke, R. G.; Liang, Y.-X.; You, S.-W.; Tay, D. K. C.; Zhang, S.; So, K.-F. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5054−5059.

(15) Keye-Baig, C.; Fung, J. D. S. Y.; Bezaire, J.; Chen, P. *J. Am. Chem. Soc.* **2004**, *126*, 7522−7532.

(16) Duce, C. *Physical chemical methods in the rational design of new materials: QSAR and calorimetric approaches*; Ph.D. thesis: University of Pisa, 2005.

(17) Bianucci, A. M.; Massarelli, I.; Chiellini, F.; Eidelman, C.; Chiellini, E. *J. Biomater. Sci., Polym. Ed.* **2004**, *15*, 1203−1222.

(18) Fung, S. Y.; Keyes, C.; Duhamel, J.; Chen, P. *Biophys. J.* **2003**, *85*, 537−548.

(19) Hong, Y.; Legge, R.; Zhang, S.; Chen, P. *Biomacromolecules* **2003**, *4*, 1433−1442.

(20) Jun, S.; Hong, Y.; Imamura, H.; Ha, B.-Y.; Bechhoefer, J.; Chen, P. *Biophys. J.* **2004**, *87*, 1249−1259.

(21) Hwang, W.; Marini, D. M.; Kamm, R. D.; Zhang, S. *J. Chem. Phys.* **2003**, *118*, 389−397.

(22) Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 330−335.

(23) Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *300*, 486−489.

(24) Ding, F.; Dokholyan, N. V. *Trends Biotechnol.* **2005**, *23*, 450−455.

(25) Dokholyan, N. V.; Borreguero, J. M.; Buldyrev, S. V.; Ding, F.; Stanley, H. E.; Shakhnovich, E. I. *Methods Enzymol.* **2003**, *374*, 616−638.

(26) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Folding Des.* **1998**, *3*, 577−587.

(27) Peng, S.; Ding, F.; Urbanc, B.; Buldyrev, S. V.; Cruz, L.; Stanley, H. E.; Dokholyan, N. V. *Phys. Rev. E* **2004**, *69*, 41908−1914.

(28) Urbanc, B.; Cruz, L.; Ding, F.; Sammond, D.; Khare, S.; Buldyrev, S. V.; Stanley, H. E.; Dokholyan, N. V. *Biophys. J.* **2004**, *87*, 2310−2321.

(29) Dokholyan, N. V. *Curr. Opin. Struct. Biol.* **2006**, *16*, 79−85.

(30) Smith, A. V.; Hall, C. K. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 376−391.

(31) Case, D. A. et al. *AMBER 8*; University of California: San Francisco, 2002.

(32) 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.

(33) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(34) Carravetta, V.; Monti, S. *J. Phys. Chem. B* **2006**, *110*, 6160−6169.

(35) Polzonetti, G.; Battocchio, C.; Dettin, M.; Gambaretto, R.; Bello, C. D.; Carravetta, V.; Monti, S.; Iucci, G. Submitted for publication.

(36) Blanco, M. *J. Comp. Chem.* **1991**, *12*, 237−247.

(37) *SYBYL Molecular Modelling Software, Version 7.2*; TRIPOS Associates: St. Louis, MO, 2005.

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

(39) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327−341.

(40) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin Heidelberg, 1991.

(41) Sugita, S.; Okamoto, Y. *Biophys. J.* **2005**, *88*, 3180−3190.

(42) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.

(43) Takano, K.; Yamagata, Y.; Yutani, K. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 274−280.

(44) Takano, K.; Yamagata, Y.; Yutani, K. *Proteins* **2001**, *44*, 233−243.

(45) Komeiji, Y.; Uebayasi, M.; Someya, J.; Yamato, I. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 268−277.