Theor Chem Acc (2004) 112: 145–157
DOI 10.1007/s00214-004-0565-4

Theoretical
Chemistry Accounts

Regular article

The influence of simulation conditions in molecular dynamics investigations of model β-sheet peptides

Luca Monticelli[1], **Giorgio Colombo**[2]

[1] Centre for Biomolecular Interdisciplinary Studies and Industrial Applications, University of Milan, 20131 Milan, Italy
[2] Istituto di Chimica del Riconoscimento Molecolare, CNR via Mario Bianco, 9, 20131 Milan, Italy

Received: 11 October 2003 / Accepted: 20 October 2003 / Published online: 29 April 2004

Abstract. The influence of simulation methods, cutoff based and particle mesh Ewald (PME) on the accuracy by which experimentally derived nuclear Overhauser effect (NOE) data are reproduced, has been investigated using 500-ns-long molecular dynamics simulations on a model β-sheet peptide in explicit solvent. The structural and conformational features under the different conditions were evaluated in terms of flexibility, secondary structure content, hydrogen-bonding pattern and percent of native contacts as a function of time. It was found that the different simulation methods strongly influence the dynamics of the peptide, confirming previous observations based on ideal peptide models simulated for much shorter times. Moreover, the results of our simulations prove once more that it is necessary to reach extremely long time scales to obtain enough statistics to accurately reproduce experimental NOE restraints even in the case of the PME method, despite its tendency to the stabilization of conformations which are structurally closely related to the ones derived through experiment. Possible implications regarding the stabilization and folding mechanisms, together with their relationship to the experimental study of peptide models, are discussed.

Keywords: Protein folding – Molecular dynamics – β-sheet – Peptide conformation– Electrostatics

1 Introduction

Most proteins must fold to a specific three-dimensional structure in order to perform their biological functions in living cells. The description of the folding process from first principles is one of the most important goals of molecular biology, and is particularly important in the postgenomic era: precise knowledge of the relationship between amino acid sequence and three-dimensional structure would allow the prediction of a protein secondary and tertiary structure from sequence information, and eventually its correlation to protein function and/or misfunction. Moreover, being able to describe the folding process at atomic level would contribute to elucidating the mechanisms underlying the conformational changes involved in several pathological disorders, like Alzheimer's disease and spongiform encephalopathies, that are characterized by the formation of amyloid fibrils most likely mediated by the interaction of β-sheets [1, 2, 3].

Molecular modeling and simulation techniques have been, and are, widely used to study problems related to the stability and folding of peptide and protein systems. In the case of α–helical peptides or small proteins, computer simulations could provide a detailed picture of the early events [4, 5, 6, 7, 8, 9] in the folding reaction, and reasonable agreement with experimentally derived nuclear Overhauser effect (NOE) restraints [10]. In contrast, the study of the principles underlying the formation and stability of β-sheet secondary structure elements has been hampered at the theoretical level by the need for long simulation times, at least comparable to the microsecond timescale typical of β-sheet folding kinetics, and by the difficulty, at the experimental level, in synthesizing model systems which could be characterized in the absence of aggregation phenomena. In the past few years a number of short model peptides that are monomeric in aqueous solution and display double- or triple-stranded anti-parallel β-sheet conformations have been reported [11, 12, 13, 14, 15] and have been studied via molecular simulations at various levels [16, 17, 18, 19, 20]. In very few cases, however, a systematic and quantitative comparison of the simulation data with NMR-derived restraints was attempted to benchmark the results of simulations against experimental data, and to compare the results obtained for β-sheet forming systems using different molecular dynamics (MD) simulation conditions, with particular attention to the treatment of long-range (electrostatic) interactions (forces). Calculation of long-range electrostatic interactions

Correspondence to: G. Colombo
e-mail: giorgio.colombo@icrm.cnr.it

is the most time-consuming part of MD simulations, and, up to recently, the most common approximation has been to neglect the effect of Coulomb interactions beyond a certain cutoff distance [21]. The particle mesh Ewald (PME) method, on the other hand, despite requiring about twice the computational time compared with the cutoff method, considers all electrostatic interactions over the complete system, employing interpolation of reciprocal space Ewald sums from a grid and computation of convolutions using fast Fourier transformation [22, 23]. Within the PME method the truncation of long-range forces is then eliminated.

Moreover, no continuous simulation in explicit solvent in conditions similar to the ones in the test tube has been run for timescales which are actually comparable to the experimental folding times (of the order of microseconds for β-sheets). The questions we want to address are how different are the results due to the use of cutoff methods versus PME methods for electrostatic treatment, how do the different simulations correlate with the experimentally derived NOE-restraints and, as a consequence of the validation of the simulations with experimental data, what can simulations indicate with regards to the factors stabilizing the folded conformation of the model β-sheet peptide? As a case study for investigating the relative stability of β-sheet peptides as a function of simulation conditions, the antiparallel three-stranded β-sheet peptide designed by deAlba et al. [13], called *beta3*, and characterized via NMR spectroscopy was chosen (Fig. 1a). This type of system was studied earlier with MD simulations by Ferrara and Caflish, who were able to derive a free-energy profile for the systems using high temperatures and an implicit solvent representation. In this work, the dynamics of the system has been investigated through three different very long time scale simulations (500 ns each), using an all-atom representation for the water solvent combined with an explicit representation of the polar H atoms of amino acid functional groups and of the C–H bonds on aromatic rings, and with a united-atom model on the remaining aliphatic groups, with periodic boundary conditions. The influence of the treatment of electrostatic interactions and simulation conditions on the conformations sampled and on the dynamics of the peptide in the simulations has been investigated, using two different electrostatic schemes, including cutoff and PME; for the cutoff scheme, two simulations were carried out, one with and one without counterions.

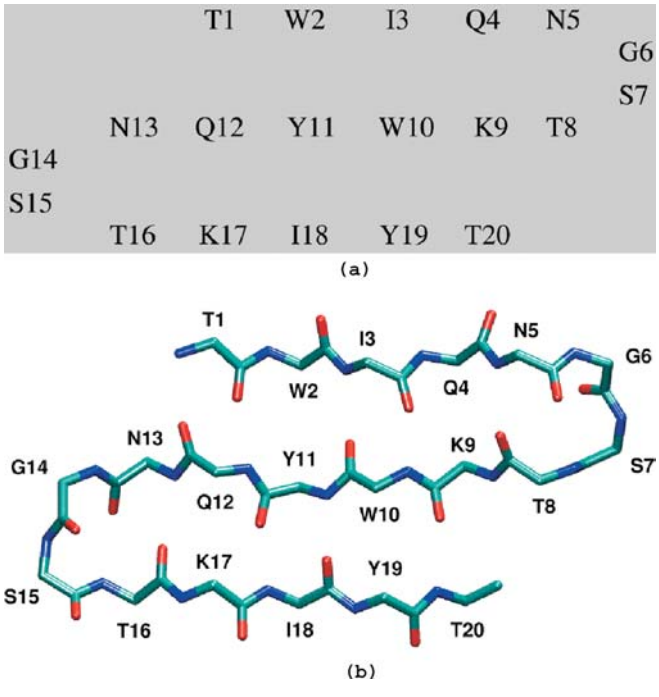

Fig. 1. **a** Amino acid sequence of the peptide by deAlba et al. [13]. **b** Hydrogen-bonding pattern in the ideal three-stranded β-sheet conformation. Only backbone atoms are shown. Hydrogen bonds are indicated in *red*

The three simulation methods were arbitrarily chosen among a wide number of other different possible electrostatic schemes, based fundamentally on two practical considerations. In fact a large number of the long time range MD studies of peptides and proteins were carried out using either the PME [6, 24, 25, 26, 27, 28, 29, 30, 31] or the cutoff method, with counterions [19, 26, 32, 33, 34, 35] or without counterions [17, 19, 36, 37, 38].

The treatment of electrostatic interactions, the effect of counterions and the artifacts induced by each method have been studied and discussed in depth by a number of authors [21, 26, 39, 40, 41, 42, 43, 44, 45, 46]. In most if not all cases, however, the peptides studied were either ideal models (e.g., polyalanines), or the proteins were studied in the crystal environment, and no systematic comparison was made with the data available for the peptides in solution. In the present study, the attention is focussed on the effect of the MD simulation protocols on the dynamics of the β-sheet peptide model with the objective of understanding at atomic level the stabilization and mobility determinants of β-sheets, fundamental factors in view of the importance that is increasingly being given to the transitions to β-structures as aggregation nuclei or seeds.

2 Materials and methods

The starting structure for the simulations was one of the NMR-derived structures [13] (kindly provided by M. A. Jimenez), with no violation of experimental restraints higher than 0.02 nm. The peptide was protonated to give a zwitterionic form (with N-terminal NH_3^+ and C-terminal COO^- groups), and the total charge was +2, owing to the presence of two lysine residues.

All simulations were carried out using the GROMACS package version 3.0 [47, 48], using the GROMOS96 43A1 force field [49]. All bond lengths were constrained to their equilibrium values, using the SETTLE algorithm [50] for water and the LINCS algorithm [51] for all other bonds. The aromatic rings in tyrosine and tryptophan side chains were kept rigidly planar using dummy atoms constructions, according to a published procedure [52], in order to remove degrees of freedom from the system. According to Feenstra et al. [52], removal of these degrees of freedom from the system allows the time step to be increased up to 7 fs with negligible influence on the thermodynamical and dynamical properties of the system. In our simulations, the timestep was 2 fs during solvent equilibration (initial 50 ps) and 5 fs in the following steps; the neighbor list for the calculation of non-bonded interactions was updated every ten timesteps in the first case, and every four in the latter.

The peptide was solvated with water in a dodecahedral box large enough to contain 1.2 nm of solvent around the peptide. The resulting system was composed of 2681 water molecules. The simple point charge water model was used [53].

Three simulations were carried out, using the same starting structures; box size and simulation parameters were identical except for the treatment of electrostatic interactions. In the first and second simulations, chloride ions replaced two water molecules to yield an electrically neutral system. The water molecules to be replaced were chosen randomly, using the GENION algorithm (included in the GROMACS package). In the first run, the calculation of electrostatic forces utilized the PME method [22, 54]; the real-space interactions were evaluated using a 0.9-nm cutoff, and the reciprocal-space interactions were evaluated on a 0.12-nm grid with fourth-order spline interpolation. A twin-range cutoff of 0.8–1.4 nm was used for the calculation of Lennard-Jones interactions. The relative accuracy of direct/reciprocal space is controlled by the parameter ewald_rtol (using GROMACS methodoloy), which is set to 1×10^{-5}.

In the second and third run, a twin-range cutoff of 0.8–1.4 nm was used for both Coulombic and Lennard-Jones interactions. The cutoff values are the same as those used for the parameterization of the GROMOS96 43A1 force field [49]. In the following tables and figures, the first simulation is referred to as B3-PME, the second as B3-ION and the third as B3-NOION. In all cases, periodic boundary conditions were used. The differences in the setup of the three simulations are summarized in Table 1.

The peptide, water and the counterions were coupled separately to a temperature bath at 300 K with $\tau_T = 0.1$ ps using the Berendsen algorithm [55]. The pressure was kept at 1 bar using weak pressure coupling with $\tau_P = 1.0$ ps [55].

The system was energy-minimized with a steepest descent method for 5000 steps. In all the simulations the solvent was equilibrated in a 50-ps MD run with position restraints on the peptide. The force constant on the peptide atoms was 1000 kJ mol^{-1} nm^{-2}. The solvent equilibration run was followed by another 100 ps run without position restraints on the peptide, in which all atoms were given an initial velocity obtained from a Maxwellian distribution at the desired initial temperature. Each production run, after equilibration, was 500-ns long, and peptide structures were sampled every 1 ps. To our knowledge, the present simulations are among the longest carried out on this kind of system using an all-atom representation of the solvent.

Conformational clustering analysis was performed on a subset of the conformations sampled in the trajectory, containing either 5000 structures (taken at 100-ps intervals) or 25000 structures (taken at 20-ps intervals), using the method described by Daura et al. [10] (as implemented in GROMACS). The positional root mean square deviation (RMSD) calculated on the backbone atoms of residues 3–18 was used as a similarity criterion, and a cluster radius of 0.1 nm was chosen.

The interproton distances were calculated from the simulations as $\left\langle d(t)^{-6} \right\rangle^{-1/6}$ averages [56], and violations were derived as $d_{viol} = \left\langle d(t)^{-6} \right\rangle^{-1/6} - d_{exp}$ (where d_{exp} is the experimentally derived distance). Since in the GROMOS96 43A1 force field the aliphatic hydrogen atoms are treated within a united-atom model, their position at each time in the trajectory was back-calculated assuming an ideal geometry.

In general, in the GROMOS force field, an atom–atom distance restraint which is derived from proton NMR experiments regarding hydrogen atoms on aliphatic carbons is treated by way of a virtual atom. In that case the distance restraint interaction refers to a nonatomic site as a center of interaction. These may also refer to nonatomic sites when a stereospecific assignment to a proton cannot be obtained. In these cases the distance restraint must refer to a pseudo-atom and a correction term must be added to the restraint distance. Calculating atom–atom distances this way gives the same results as using the procedure of building explicit hydrogens back in the trajectory, and does not affect the average calculated violation in any significant way: In this latter case, however, we have access to the comparison with stereospecific assignments, when they are available. When the NOE involved chemically equivalent aliphatic hydrogen atoms (e.g., Hδ and Hγ of isoleucine residues), the average distance [calculated as $\left\langle d(t)^{-6} \right\rangle^{-1/6}$] was taken into account.

Secondary structure assignments were based on the DSSP algorithm [57]. The graphical representations of the peptide were realized with the program VMD [58].

All calculations were performed on clusters of six PCs (dual processor AMD Athlon XP 1800+), on a HP V2500 computer (20 CPU PA-RISC PA8500 at 440 MHz) and on a HP SuperDome SD64000 computer (64 CPU PA-RISC PA8700 at 750 MHz).

3 Results

3.1 Backbone and secondary structure evolution

The RMSD relative to the starting structure is reported in Fig. 2 as a function of time for the three simulations. The average values are much higher in the simulations using the cutoff electrostatic scheme. This difference is also reflected in the evolution of the radius of gyration (R_g) values: while the average value over the whole simulation is very similar in the three simulations, as shown in Table 1, remarkable oscillations are present in the two cutoff simulations with values up to 1.15 and 0.99 nm in the B3-ION and B3-NOION simulations, respectively. The motional behavior of the peptide is thus determined by the particular simulation protocol employed, with sensitive differences in residue flexibility displayed by the system under different conditions: flexibility can be quantitatively probed with the calculation of the average positional fluctuation (root mean square fluctuation, RMSF), as reported in Fig. 3. In the case of B3-PME simulation, not only is the flexibility much lower than in the other two cases, but three minima are evident in the graph, corresponding to the central regions of the three extended strands (shaded areas in Fig. 3). The second strand shows the lowest RMSF values, owing to the higher number of hydrogen

Table 1. Summary of the simulations performed, differences in their setup and structural features of the peptide conformations sampled in the three simulations. The radius of gyration (R_g) was calculated on all atoms, while only backbone atoms in the 3–18 segment were used for the calculation of the positional root mean square deviation (RMSD)

System	Electrostatics	No. of H_2O molecules	No. of Cl^- ions	$< R_g >$ (nm)	RMSD (nm)
B3-PME	PME (0.9)[a]	2679	2	0.752 ± 0.023	0.247
B3-ION	Cutoff (0.8/1.4)[a]	2679	2	0.764 ± 0.054	0.541
B3-NOION	Cutoff (0.8/1.4)[a]	2681	0	0.754 ± 0.048	0.484

[a]Cutoff distances in nanometers

bonds formed by the central portion of the peptide, a feature not reproduced in the case of B3-ION and B3-NOION, indicating lower conformational stability in those conditions.

The nonnegligible influence of the simulation conditions on the peptide conformational evolution is clearly defined in terms of quantitation of ordered secondary structure and its evolution. The substantial differences among the three simulations are shown in Fig. 4: while in B3-PME the secondary structure is close to the ideal three-stranded β-sheet, in both other cases the percentage of β-sheet is very low and the secondary structure elements are found in positions different from those experimentally determined. In the B3-PME simulation the β-turns are found in positions 6–7 (77% of the simulation time) and 14–15 (87% of the simulation time), which is consistent with NMR data [13]; in contrast, in the other two MD runs (B3-ION and B3-NOION) the β-turns are not very stable. Analysis of dihedral angles indicates that, in B3-PME, both the β-turns belong to type II′. The population of the β-sheet structure calculated from this simulation is about 65%, significantly higher than that estimated from the experiments based on the NOE intensities (13–31%) but in better agreement with the more quantitative determination based on the variation of chemical shift data (30–55%) [13]. The second conformational ensemble (30% of the total conformations) includes a single β-hairpin (residues 11–18) and a disordered region.

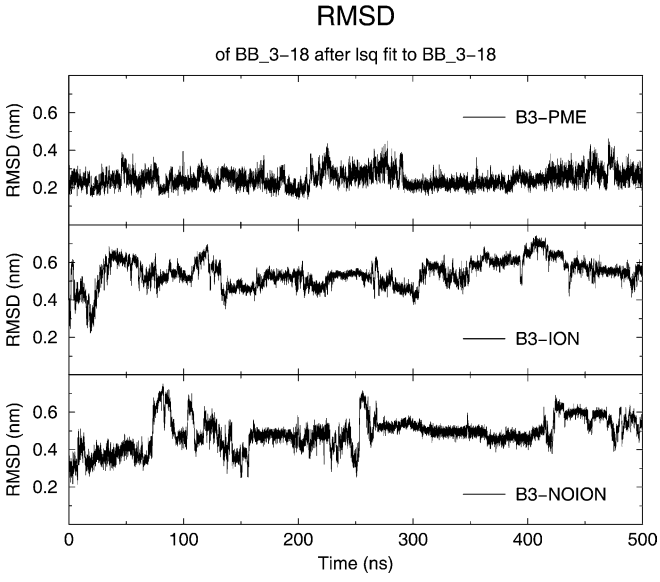

Fig. 2. Positional Root mean square deviation (*RMSD*) for the three simulations, calculated on backbone atoms of residues 3–18

In B3-ION the population of the β-turns is much lower than expected; nevertheless, the first β-turn (residues 6–7) is replaced by a larger bend, involving residues 6–9, for most of the simulation time (Fig. 4b); since residues 2–5 and 10–13 are found mainly in an extended conformation, the ensemble of structures sampled is quite similar to a single hairpin, followed by a disordered region. From 400 to 425 ns an extended structure is observed for residues 11–17, thus including the second turn region and yielding a conformation that is very different from the NMR-derived structures. Very few structures, among all those sampled in 500 ns, show a regular β-hairpin conformation (about 7%), while the population of three-stranded β-sheet structures is close to zero.

In B3-NOION the population of the β-turns is comparable to that in the B3-PME simulation, but a closer inspection of Fig. 4c reveals that, in the second half of

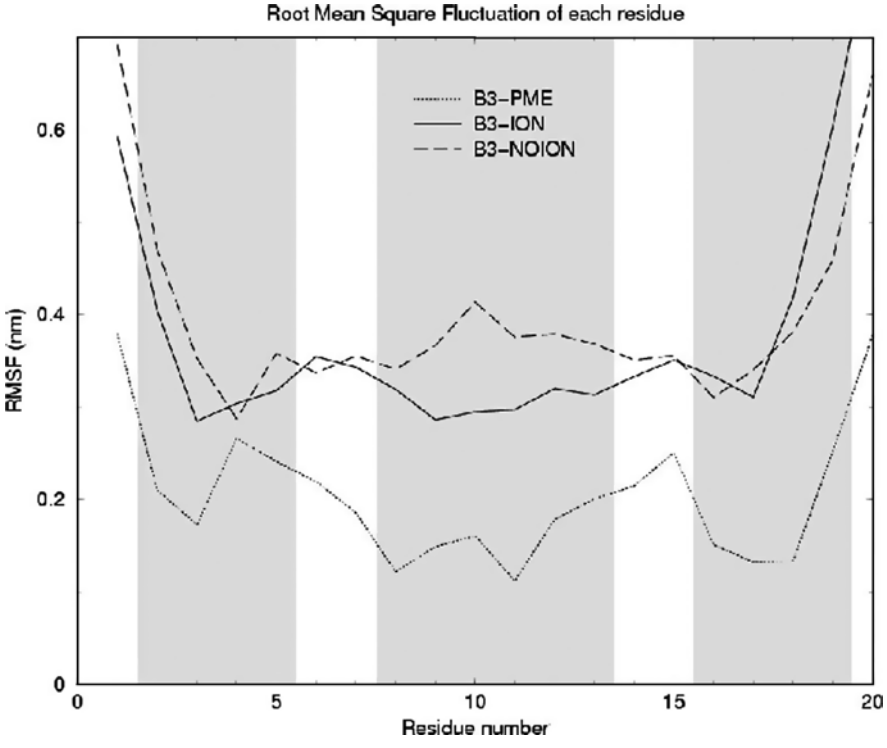

Fig. 3. Root mean square fluctuation (*RMSF*) for each residue in the three simulations. The fluctuations were calculated on the backbone atoms only, after least-squares fitting. *Shaded areas* indicate the residues in the extended strands (in the NMR-derived structures, by deAlba et al.)

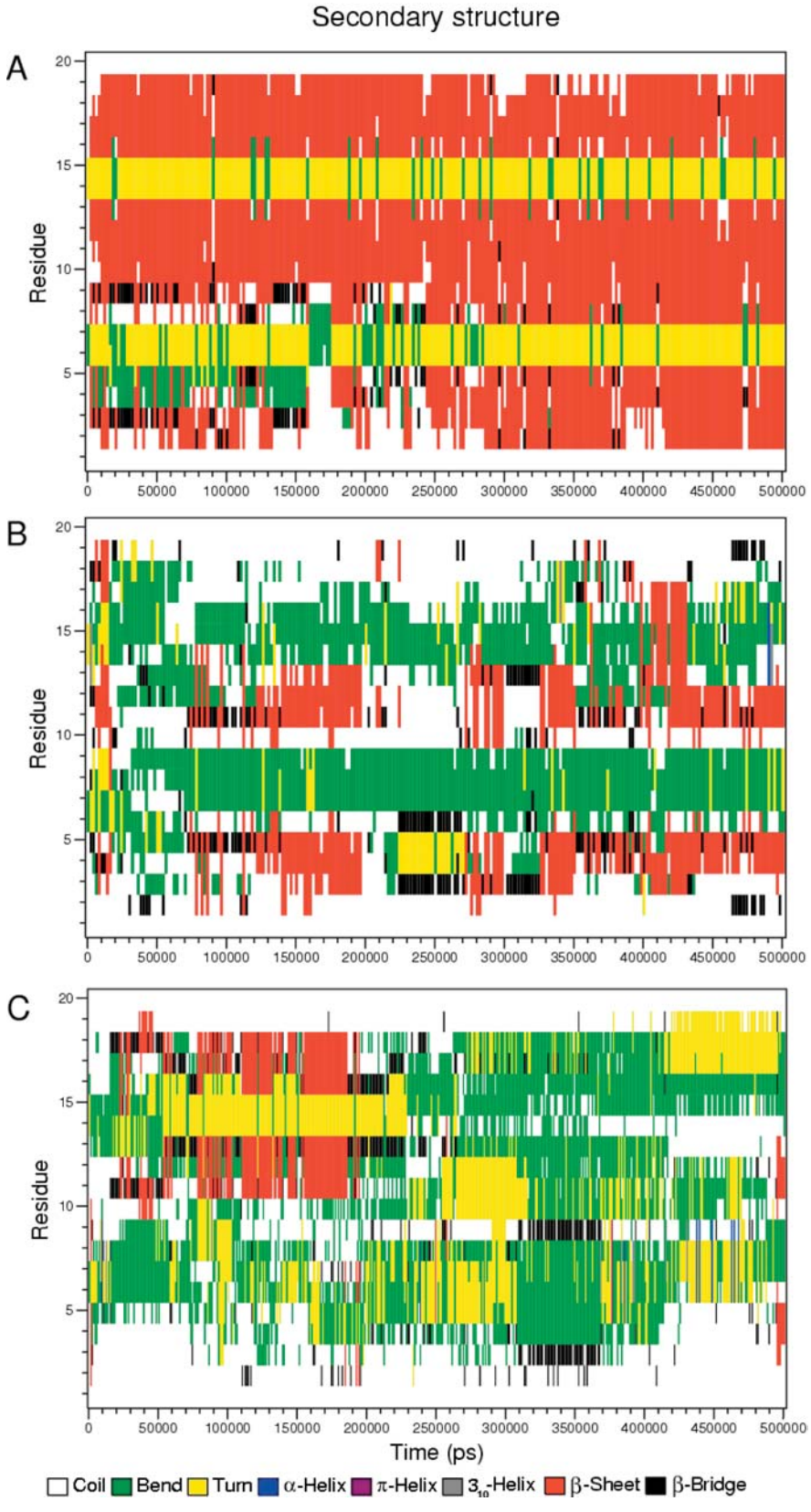

Fig. 4. Secondary structure as a function of the simulation time in each of the three simulations: **a** B3-PME, **b** B3-ION, **c** B3-NOION

the MD run, the position of the turns is different from that experimentally detected. The population of the β-sheet structure is high for the second hairpin (residues 11–18) only from 60 to 190 ns of the MD run, but it is close to zero from 190 ns on. Interestingly, the second β-turn (residues 14–15) is stable for a much longer time compared with the β-hairpin, until approximately 230 ns; moreover, a β-turn comprising residues 5–7 is

stable from approximately 230 to 310 ns; in both cases, the formation of a stable β-turn leads to the formation of a conformation that resembles a single β-hairpin structure, but is not sufficient to induce the formation of the hydrogen bonds characteristic of a regular β-sheet structure. Considering the whole trajectory, the population of the regular β-hairpin structure is approximately 25%, and the population of three-stranded β-sheet structures is only 2%, higher than in the B3-ION simulation. In all cases, analysis of the distributions of ϕ and ψ angles in the turn regions reveals that the type II′ turn is much more populated than all other types of β-turn.

Partially folded structures, showing one β-hairpin and a disordered region, were previously described by Ferrara and Caflisch [20], studying the same peptide, and also by Colombo et al. [17] studying Betanova. In our simulations, they are observed for significant parts of the simulations; in all cases, the second β-hairpin (residues 11–18) is formed preferentially, and its presence is always observed together with the formation of a salt bridge between the ammonium group on the side chain of Lys9 and the peptide C-terminal carboxy group. This salt bridge is observed for approximately 90% of the simulation time in B3-PME, and for 25% in B3-NOION. It is interesting to observe that the same salt bridge is also present in the B3-ION simulation and is highly populated (approximately 30% of the simulation time), but in this case regular β-hairpin structures are rarely formed (only 7% of the simulation time). There is no apparent correlation between the stability of the β-hairpin structure and the salt bridge interaction.

The estimate for the population of the β-sheet varies in the three simulations, and only in the B3-PME simulation is a predominant β-sheet conformation clearly defined for a dominant amount of time. The population calculated in this case (65%) is higher than the one calculated experimentally (see earlier), indicating that the use of the PME method can yield the stabilization energy necessary to drive the system into the three-stranded energy basin. This aspect is discussed in the next subsection.

3.2 Hydrogen-bonding pattern and native contacts

An important contribution to the stabilization energy in β-sheet systems comes from the formation of in-register hydrogen bonds. The populations of all the hydrogen bonds detected in simulation B3-PME are reported in Table 2. All hydrogen bonds characteristic of the ideal β-sheet conformation (Fig. 1b) are found, and their population is between 45% and 50% for the first hairpin (except for the one involving residue 1, which is very flexible), while the second hairpin is even stabler. A minor conformation is present in the trajectory in which the register of the hydrogen-bonding pattern displays a one-residue shift: hydrogen bonds are found between residues Trp2 and Trp10, and residues Gln4 and Thr8. The presence of this conformation is mainly due to the packing of the tryptophan side chains on the hydrophobic cluster present in the peptide. In simulation B3-ION, the hydrogen-bonding pattern characteristic of regular β-hairpin conformations is detected only less than 5% of the time. As emerged from the analysis of the secondary structure, the first β-turn (residues 6–7) is replaced by a larger bend (residues 6–9) for most of the simulation time, and a high population is registered for the hydrogen bonds between Asn5 and Tyr11 and between Ile3 and Asn13, which is not compatible with the NOE data from deAlba et al. [13]. In simulation B3-NOION, interstrand hydrogen bonds involving the first hairpin residues have a very low population, while those involving the second hairpin are quite stable for about 200 ns, and their pattern is the same as observed in the B3-PME simulation and is compatible with the NMR data. After 225 ns, however, a global misfolding of the peptide is observed and the conformations sampled are very different from those experimentally detected.

Table 2. Populations of the interstrand hydrogen bonds during B3-PME simulations. Only hydrogen bonds between backbone atoms are reported

Donor	Hydrogen	Acceptor	Population (%)
Thr1N	Thr1H1	Gln12O	9.4
Thr1N	Thr1H2	Gln12O	7.8
Thr1N	Thr1H3	Gln12O	5.8
Trp2N	Trp2H	Trp10O	9.2
Trp2N	Trp2H	Gln12O	1.0
Ile3N	Ile3H	Trp10O	47.3
Gln4N	Gln4H	Thr8O	27.9
Gln4N	Gln4H	Trp10O	1.4
Asn5N	Asn5H	Thr8O	52.7
Asn5N	Asn5H	Trp10O	1.8
Thr8N	Thr8H	Gln4O	1.4
Thr8N	Thr8H	Asn5O	49.9
Trp10N	Trp10H	Trp2O	28.1
Trp10N	Trp10H	Ile3O	49.5
Gln12N	Gln12H	Thr1O	37.5
Gln12N	Gln12H	Trp2O	2.4
Gln12N	Gln12H	Ile3O	0.6
LysH9N	LysH9H	Tyr19O	1.6
LysH9N	LysH9H	Thr20N	0.2
LysH9N	LysH9H	Thr20O1	17.8
LysH9N	LysH9H	Thr20O2	17.6
Tyr11N	Tyr11H	Ile18O	93.4
Asn13N	Asn13H	Thr16O	88.2
Thr16N	Thr16H	Asn13O	55.5
Ile18N	Ile18H	Tyr11O	94.0
Thr20N	Thr20H	LysH9O	72.1

The number of nativelike intramolecular interactions necessary to stabilize the peptide can be determined by the calculation of the time evolution of the fraction of native contacts. The fraction of native contacts was calculated on the Cα carbons as a function of simulation time in each MD run (Fig. 5). A Cα–Cα interaction is defined as a native contact if the Cα–Cα distance is shorter than 0.65 nm in an ideal β-hairpin conformation. In the B3-PME simulation, the fraction of native contacts is higher than 0.4 for all the simulation time, confirming that the peptide never unfolds completely. After a few picoseconds, its value drops below 0.6, and it takes about 200 ns of the MD run to reach the value of 0.9, indicating an ideal fold; fluctuations are large (between 0.5 and 0.9) until about 300 ns, when the value of 0.9 is

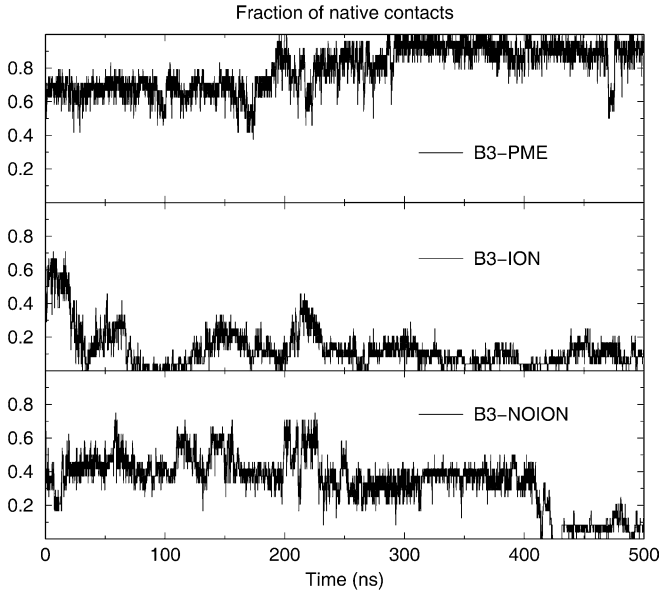

Fig. 5. Fraction of native contacts as a function of the simulation time in each of the three simulations. A Cα–Cα interaction is defined as a native contact if the Cα–Cα distance is shorter than 0.65 nm in an ideal β-hairpin conformation

reached and maintained for a long time (Fig. 5). A substantial convergence is obtained only after about 300 ns of simulation. This result is consistent with the secondary structure analysis, showing a very high and constant population for the β-sheet structure from 250 to 500 ns. Obviously, in both the B3-ION and the B3-NOION simulations the fraction of native contacts is much lower; however, it is interesting to observe that the analysis of native contacts evidences a clear difference between these two simulations. In B3-ION the average fraction of native contacts is 0.13, and it reaches a value greater than 0.4 only three times in 500 ns; in contrast, the average fraction of contacts in B3-NOION is 0.35 and it reaches the value of 0.6 several times during the MD run, indicating that the conformations sampled during this simulation are, on average, much more similar to β-sheet structures compared with those sampled during the B3-ION simulation.

It is important to note that the fraction of native contacts as a function of simulation time shows large variations in all MD runs, confirming that a very long simulation time is necessary in order to sample significant parts of the conformational space of β-sheet peptides and cross possible lag time phases, especially when using explicit solvent models.

3.3 Cluster analysis

A conformational-clustering algorithm was used to characterize the conformational space sampled in the simulations. The number of clusters as a function of simulation time is reported in Fig. 6; a cluster radius of 0.1 nm was used, and structures were taken at 20- or at 100-ps intervals to check the influence of sampling frequency on the results. The total number of clusters is

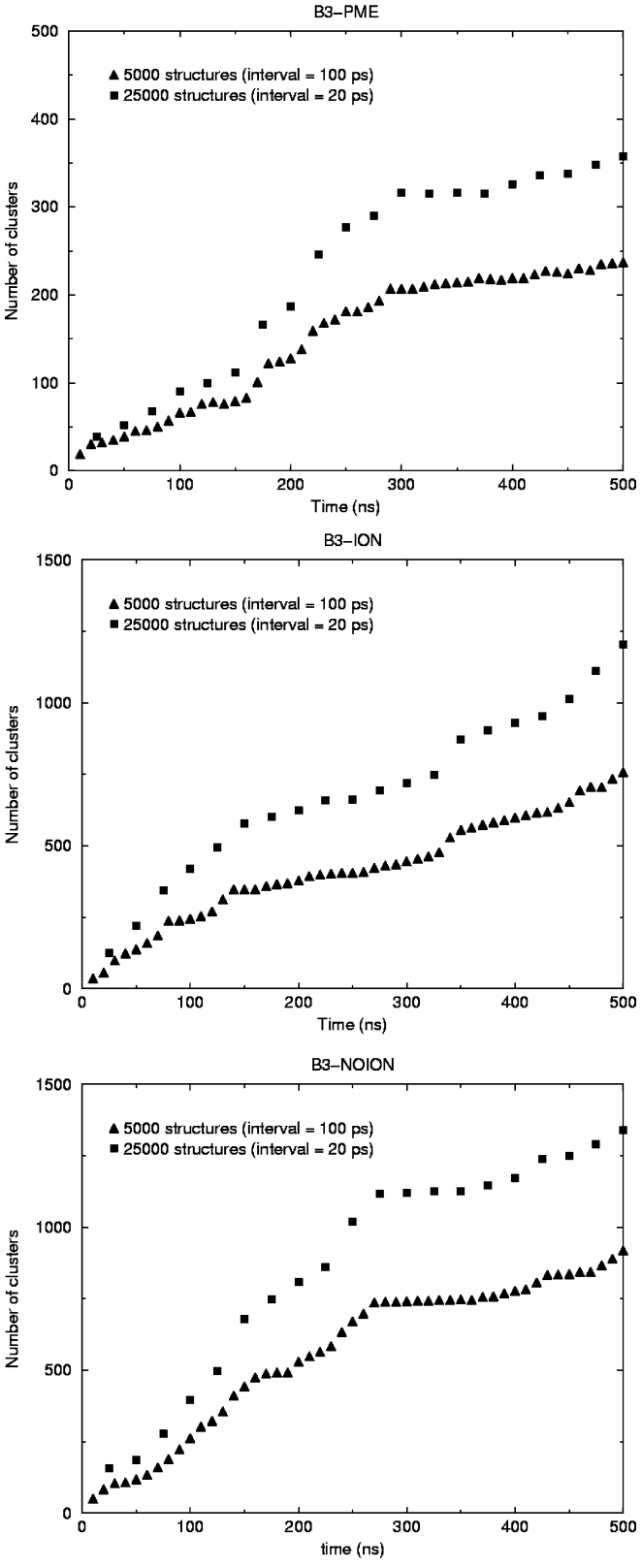

Fig. 6. Number of clusters as a function of simulation time. From *top* to *bottom* B3-PME, B3-ION, B3-NOION

reported in Table 3. In all cases, when structures are taken at shorter intervals the number of clusters found is higher; nevertheless, the trend of the number of clusters

Table 3. Number of clusters of peptide conformations and percentage population of the first cluster in each of the three simulations carried out. Clustering was performed on 5000 structures, taken at 100-ps intervals, using a cluster radius of 0.1 nm

System	No. of clusters (over 500 ns)	Population of cluster 1 (%)	Population of clusters 1–5 (%)
B3-PME	237	16.1	47.0
B3-ION	758	5.9	19.9
B3-NOION	918	10.9	32.1

as a function of simulation time is identical, and therefore independent of the time interval. The number of clusters increases during the simulation time in all three simulations, indicating clearly that convergence, i.e., complete sampling of conformational space, has not been reached. However, important differences are found among the three simulations. The number of clusters found in B3-PME is much lower than in B3-ION and B3-NOION, and increases only slightly during the last 200 ns, indicating, coherently with previous analysis, that most structures sampled in this simulation are found within the initial 300 ns. This is mainly due to the energetic stabilization of the three-stranded conformation determined by the explicit inclusion of the long-range interactions in the B3-PME simulation. The transitions among the different clusters as a function of simulation time are reported in Fig. 7: a circle is depicted if a certain cluster is populated at a given time. The two lower graphs, representing the two cutoff simulations, show a much higher number of intercluster transitions. Another interesting result is that the five most populated clusters account for about 47% of all structures in simulation B3-PME, and for a much lower fraction in the others (Table 3), confirming once again a higher flexibility in the simulations using the cutoff electrostatic scheme. In all cases, partially folded structures, with one hairpin and a disordered segment, are frequently sampled. The representative structures of the most populated clusters for B3-PME are reported in Fig. 8.

An alternative clustering algorithm, based on K-means using backbone dihedrals (defined by the Cα of four consecutive residues), was used as a more informative method on the conformational variation of the structures within each trajectory. The observation of the plot of means for each cluster (Fig. 9), where each point represents the mean value of a particular torsion in that cluster, proves that in the case of B3-PME the variability of the dihedrals is low: the single mean values do not present particular variability on going from one conformational family to another. Some degree of flexibility is displayed by variable 4, corresponding to the dihedral defined by the Cα atoms of residues Q4, N5, G6, and S7, containing a flexible glycine residue, and by the variables corresponding to the terminal parts of the strands. In B3-ION and B3-NOION, a much higher conformational variability is also present in the space of torsions. Moreover, in the case of B3-PME, most of the elements are concentrated in two big clusters containing, respectively, 28% and 25% of the total population, representing nativelike

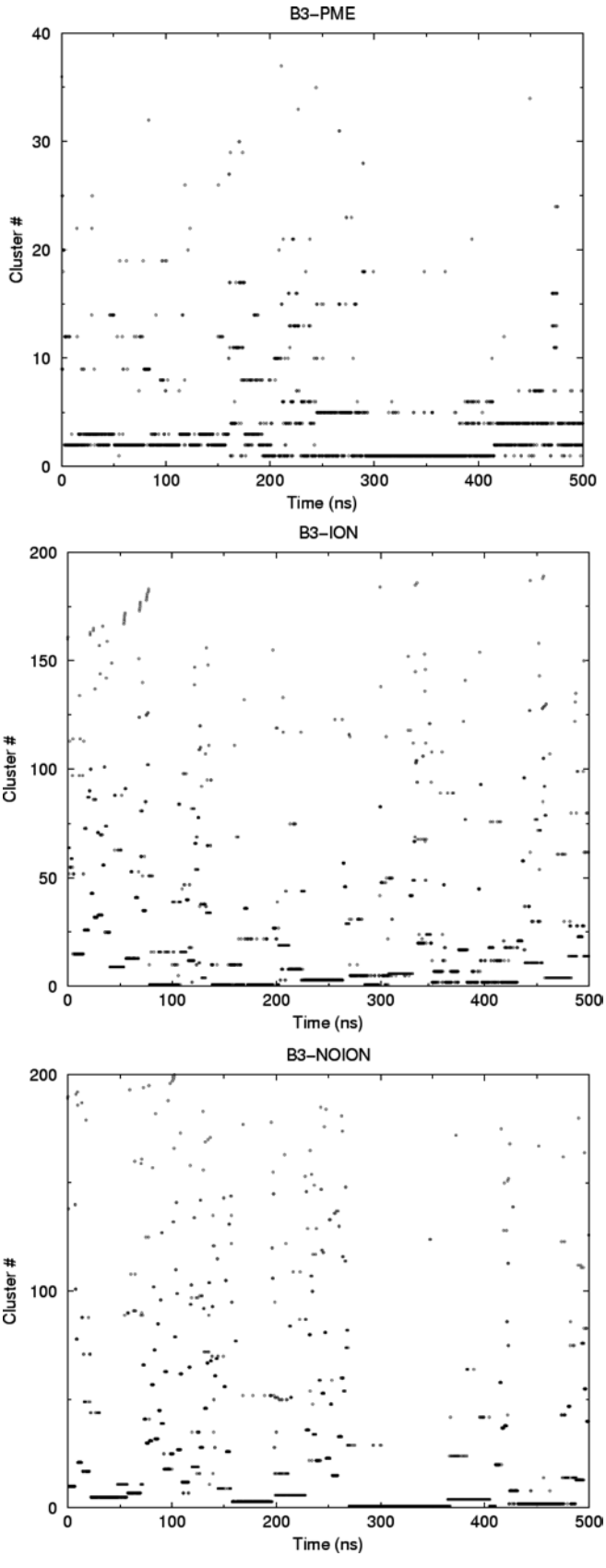

Fig. 7. Conformational transition among different clusters as a function of simulation time. A *circle* is present if a defined cluster is populated at a defined time. From *top* to *bottom* B3-PME, B3-ION, B3-NOION

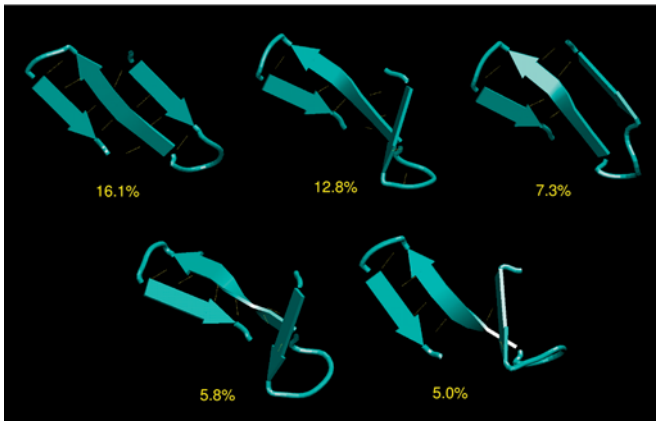

Fig. 8. Backbone structures of the five most populated clusters in the B3-PME simulation. Hydrogen bonds are colored in *yellow*. The population of each cluster is indicated

structures. The remaining objects are evenly distributed among the remaining ensembles. In the other two cases the two most populated ensembles account for only 12% and 15% of the total population.

3.4 Inter-proton distances

The results of the long simulations performed, considering the possible high amount of statistics and information they can yield, can be directly compared (and validated) with the experimentally derived NMR–NOE data, to assess their capability to sample the conformational space around the experimental conformation. If reasonable agreement is found between simulations and experiment, then simulations can actually be used to reveal factors and phenomena which are not accessible experimentally and to derive models for the description of the factors influencing β-sheet stability. In no cases, previously, was a direct comparison of the performance of each of the simulation methods made with the interproton distances derived from NOE calculations.

Interproton distances derived from experimental NOEs at 298 K were compared with the corresponding average distances in our simulations. The experimental NOE connectivities defining the β-sheet conformation of the peptide are reported (from Ref. [13]) in Table 4 together with the values of the violations calculated from each simulation, the number of violations and the value of the average violation.

As for simulation B3-PME, the results indicate that most of the NOE constraints are satisfied, when averaged over the whole trajectory. This is to be expected, because the molecule samples an ensemble of conformations characterized by the presence of three extended strands and two turn regions that are very similar to any of the NMR-derived structures. Using $\langle d^{-6}\rangle^{-1/6}$ distance weighting, significant violations (higher than 0.05 nm) are observed for four NOEs only, corresponding to entries 1, 3, 9 and 20 in Table 4. While entry 1 corresponds to a medium-range NOE between residues Thr1 and Ile3 (CαH T1 – CγH$_3$/CδH$_3$ I3), the other three violations involve long-range NOEs between strands 1 and 2 (CαH T1 - CεH Y11 and CγH$_3$/CδH$_3$ I3 – CαH N5, d_{viol} values of 0.080 and 0.096 nm respectively) and between strands 2 and 3 (CεH Y11–CγH$_3$ I18, d_{viol} value of 0.097 nm). The average violation, calculated over all the distances, is 0.023 nm, and is higher for the long-range NOEs within the first hairpin (residues 3–10) compared with those within the second hairpin (residues 11–18). This is consistent with the second hairpin being stabler than the first one in the B3–PME simulation. However, the most important results regard the violations of NOE distances in the two simulations carried out using the twin-range cutoff; the number of violations over 0.05 nm is 16 and 17 in B3-ION and B3-NOION respectively, about 4 times larger than in the B3-PME simulation, and also the average violation is about 5 times larger. It is interesting to observe that no substantial difference in the number of NOE violations and in the degree of the NOE violations is observed owing to the use of counterions; despite the conformations sampled in the B3-NOION simulation being more similar to the β-sheet structure, large violations are observed for NOEs involving both the first and the second hairpin in both MD runs using the cutoff scheme.

The number of violations was also analyzed as a function of simulation time; the simulations were split in five parts, each part 100-ns long, and the number of violations and the average violations were calculated and are reported in Fig. 10. Interestingly, both the number and the degree of the violations increase during the simulation time in B3-ION and in B3-NOION, but not in B3-PME. Moreover, the calculations show clearly that only averaging over the whole trajectory yields good agreement with the experiments, which highlights once more the need for multiple conformation as a structural model for peptides. It is important to observe that, in the B3-PME simulation, the central structure of the most populated cluster has many violations in all three MD simulations, despite the conformation of the backbone being very similar to the NMR structures and having all the hydrogen bonds characteristic of the β-hairpin. None of the structures sampled during the MD run fulfill all the NOE restraints; nevertheless, if we consider the collection of the central structures of the five most populated clusters (Fig. 8), only four violations are found; this ensemble of structures represents the conformation of the peptide in solution much better than any single structure.

4 Discussion

The benchmarking of the simulation results with the available experimental data is fundamental in the process of garnering atomic scale details on the stabilization process of peptides and in defining what the more suitable simulation conditions are for long time scale studies of β-sheet structures. The analysis of the three 500 ns MD trajectories presented in this work shows how sensitive the results of the simulations are to the conditions used, in particular with respect to reproducing the experimentally derived NOE-restraints.

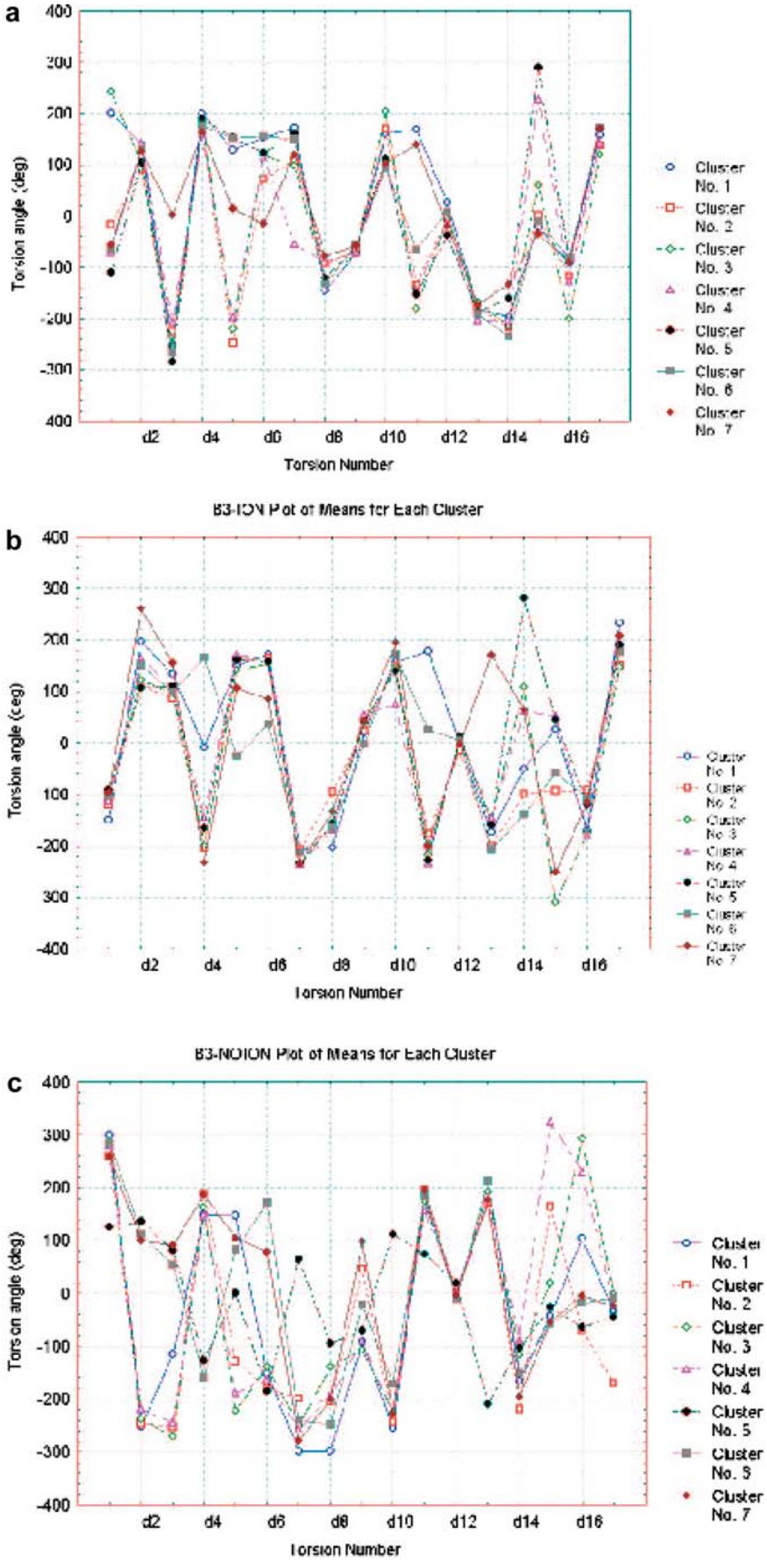

Fig. 9. Cluster mean values for each variable from the three simulations: **a** PME, **b** B3-ION, **c** B3-NOION

Table 4. Nuclear Overhauser effect (*NOE*) connectivities of the peptide from NMR experiments in water [13], and average violations of the same distances during the three simulations; both positive and negative values are reported, but negative values were not considered as violations. Violations smaller than 0.05 nm are highlighted with a + sign, those larger than 0.05 nm with two + signs

NOE no.		d_{exp} (nm)	d_{viol} (nm)		
			B3-PME	B3-ION	B3-NOION
1	CαH T1–CγH_3/CδH_3 I3	0.300	0.211 ++	0.274 ++	0.264 ++
2	Cβ′H T1–CγH_3/CδH_3 I3	0.300	0.023 +	0.057 ++	0.053 ++
3	CαH T1–CεH Y11	0.400	0.080 ++	0.342 ++	0.293 ++
4	Cββ′H W2–Cε$_3$H W10	0.450	−0.113	0.186 ++	−0.007
5	CαH W2–CαH Y11	0.300	−0.041	0.504 ++	0.391 ++
6	Cε$_3$H W2–CαH Y11	0.400	−0.078	0.058 ++	0.081 ++
7	Cε$_3$H W2–Cβ′H N13	0.400	−0.079	0.079 ++	0.192 ++
8	CαH Q4–CαH K9	0.250	0.000	0.341 ++	0.150 ++
9	CγH_3/CδH_3 I3–CαH N5	0.350	0.096 ++	0.142 ++	0.128 ++
10	CγH_3/CδH_3 I3–NH N5	0.400	−0.030	0.010 +	−0.057
11	CγH_3/CδH_3 I3–Cε$_3$H W10	0.400	−0.073	0.010 +	0.082 ++
12	CγH_3/CδH_3 I3–Cδ$_1$H W10	0.400	−0.049	0.015 +	0.028 +
13	Cδ$_1$H W10–Cβ′H Q12	0.400	−0.021	0.145 ++	0.153 ++
14	CαH W10–CαH Y19	0.300	−0.047	0.220 ++	0.252 ++
15	CδH Y11–CαH N13	0.400	0.041 +	0.001 +	0.093 ++
16	CδH Y11–Cγ′H I18	0.450	−0.110	0.058 ++	−0.016
17	CδH Y11–CγH_3 I18	0.350	0.009 +	0.056 ++	−0.006
18	CεH Y11–CαH N13	0.350	−0.015	−0.001	0.077 ++
19	CεH Y11–Cβ′H N13	0.350	−0.060	−0.023	0.117 ++
20	CεH Y11–CγH_3 I18	0.350	0.097 ++	0.111 ++	−0.010
21	CεH Y11–Cδδ′H K9	0.400	−0.073	−0.037	0.053 ++
22	CαH Q12–CαH K17	0.300	−0.068	0.308 ++	0.511 ++
23	NH T16–Cα′H G14	0.350	0.038 +	0.090 ++	0.047 +
24	Cββ′H K17–CεH Y19	0.350	−0.050	−0.052	0.032 +
25	NH T20–CγH_3 I18	0.450	−0.045	0.028 +	0.099 ++
Total number of violations (> 0.05 nm)			4	16	17
Average violation (nm)			0.023	0.121	0.123

When using cutoff schemes, as in B3-ION and B3-NOION, the starting NMR-based conformation is only marginally stable and, despite the very long simulation times used, no reversible folding–unfolding behavior is noticed. The peptide is forced to explore sets of conformations with a high number of violations of the experimentally determined NOE restraints. The situation is reverted if the PME method is used to calculate electrostatic interactions: most of the NOE restraints are indeed satisfied when averaging interproton distances over the trajectory; only four violations are observed, and three of these correspond to weak NOE signals (long-range interactions). This result is comparable to that obtained by Ferrara and Caflish, who simulated the folding of the same peptide at 360 K over 200 ns using a simplified implicit solvent model and short interatomic cutoffs [20]. It is very important to note that the highest quantitative agreement between simulation B3-PME and the experimental NOE restraints is obtained only when interproton distances are averaged over the whole set of conformations sampled during 500 ns. When we calculate the NOE distances in each of the five 100-ns portions of the trajectory, a definite increase in the number of violations and in the average violation is observed, even if three-stranded β-sheet structures are sampled during all five portions. The peptide is shown to populate a large ensemble of conformational families, which are all required to satisfy the experimental NOE restraints. An analogous result could have been obtained by using a combination of multiple trajectories of shorter time length. However, the very slow convergence observed for the B3-PME trajectory (see Sect. 3.2) implies the need for long time scale equilibration of the systems under examination. In this regard, we note once more the problem of inferring the stability and dynamics of β-sheet peptides from simulations of insufficient length compared with experimental folding times.

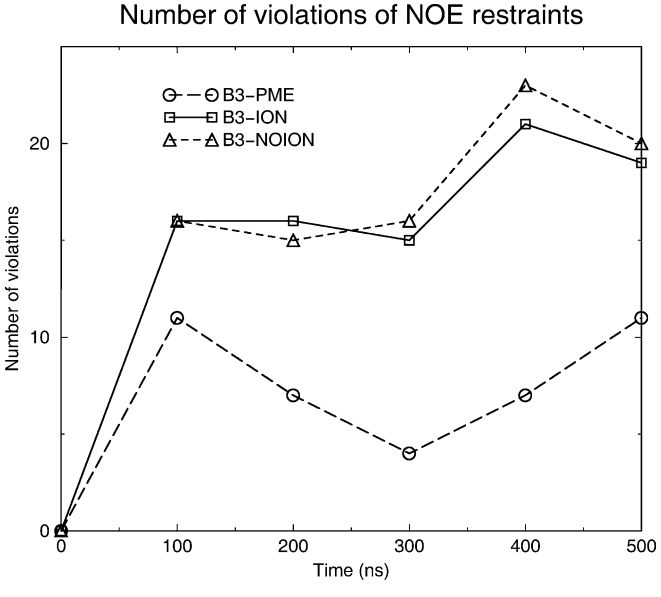

Fig. 10. Number of nuclear Overhauser effect restraint violations in different time intervals

Consistent with previous computational investigations of the peptide designed by deAlba [20] and of related three-stranded structures such as Betanova [16, 17, 59] the results of simulation B3-PME also show that in general one of the two hairpins is stabler than the other one: the solution structure of the peptide, thus, consists of a stable β-hairpin on which the temporary juxtaposition of the third strand determines the three-stranded geometry.

This situation can be discussed by thermodynamic considerations. Locking the third strand in place might be entropically too expensive, considering the high number of conformational degrees of freedom involved in the process, so a sufficient amount of stabilization energy has to be introduced into the system in the form of intramolecular interactions.

The explicit calculation of entropy would allow us to quantitatively calculate this enthalpy/entropy compensation phenomenon. This could however, be done only once fully reversible folding–unfolding simulations of these systems have been run: in this way, one would have access to the distinction between entropy and enthalpy, and to the kinetics of the process, being able to single out and quantify the different contributions to the stability of the peptide. This is currently out of reach, but the present fast increase in computer power should allow such simulations to be run in a reasonable time span [60].

5 Conclusions

In the present paper we have analyzed the dynamics of the three-stranded β-sheet peptide designed by deAlba et al. [13], via long time range (500 ns) MD simulations, and the influence of different electrostatic schemes on the results.

Definitively different behaviors are observed as a function of the simulation methodology: using a cutoff approach a loss of ordered secondary structure is observed, together with a high number of violations of the NOE-derived distance restraints. On the other hand, the use of the PME method contributes to the long-range stabilization energy necessary for the formation of a stable β-sheet structure, giving good agreement with NMR data, while leading to some overestimation of the β-sheet population. Low violations of NMR-derived distances are observed only when averaging over the whole ensemble of conformations sampled during the MD run, which is compatible with the presence in solution of several conformations at the same time.

From the experimental point of view, these considerations should be useful in the very many cases in which molecular simulations are used as a necessary tool in the design of new peptides with particular structure–function relationships.

Acknowledgements. We thank the "Consorzio Interuniversitario Lombardo per l'Elaborazione Automatica" (CILEA) for generous allocation of computer time on the HP machines. We are also grateful to Claudio Arlandini and Francesca Bonini for the installation and optimization of GROMACS on the HP computers at the CILEA. The C.N.R. is acknowledged for providing the facilities necessary to carry out this research. We also wish to thank M. A. Jimenez for providing the NMR-derived structures and A. Bernardi and L. Belvisi for helpful discussion.

References

1. Kelly J (1997) Structure 5: 595
2. Booth D (1997) Nature 385: 787
3. Ma BY, Nussinov R (2002) Proc Natl Acad Sci USA 99: 14126
4. McCammon JA, Harvey SC (1987) Dynamics of proteins and nucleic acids. Cambridge University Press, Cambridge
5. van Gunsteren WF, Luque FJ, Timms D, Torda AE (1994) Annu Rev Biophys Biomol Struct 23: 847
6. Duan Y, Kollman PA (1998) Science 282: 740
7. Karplus M, McCammon JA (2002) Nat Struct Biol 9: 646
8. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE (1998) J Mol Biol 280: 925
9. Daura X, van Gunsteren WF, Mark AE (1999) Proteins 34: 269
10. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE (1999) Angew Chem Int Ed 38: 236
11. Kortemme T, Ramirez-Alvarado M, Serrano L (1998) Science 281: 253
12. Gellman SH (1998) Curr Opin Chem Biol 2: 717
13. deAlba E, Santoro J, Rico M, Jimenez AA (1999) Protein Sci 8: 854
14. Serrano L (2000) Adv Protein Chem 53: 49
15. Cochran AG, Skelton NJ, Starovasnik MA (2001) Proc Natl Acad Sci USA 98: 5578
16. Bursulaya B, Brooks CL III (1999) J Am Chem Soc 121: 9947
17. Colombo G, Roccatano D, Mark AE (2002) Proteins Struct Funct Genet 46: 380
18. Colombo G, de Mori GMS, Roccatano D (2003) Protein Sci 12: 538
19. Roccatano D, Colombo G, Fioroni M, Mark AE (2002) Proc Natl Acad Sci USA 99: 12179
20. Ferrara P, Caflisch A (2000) Proc Natl Acad Sci USA 97: 10780
21. Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) J Chem Phys 102: 6199–6207
22. Darden T, York D, Pedersen L (1993) J Chem Phys 98: 10089
23. Darden T, Perera L, Li L, Pedersen L (1999) Struct Fold Des 7: R55
24. Zuegg J, Gready JE (1999) Biochemistry 38: 13862
25. Perera L, Foley C, Darden TA, Stafford D, Mather T, Esmon CT, Pedersen LG (2000) Biophys J 79: 2925
26. Norberg J, Nilsson L (2000) Biophys J 79: 1537
27. Lee MR, Tsai J, Baker D, Kollman PA (2001) J Mol Biol 313: 417
28. Massi F, Peng JW, Lee JP, Straub JE (2001) Biophys J 80: 31
29. Tieleman DP, Hess B, Sansom MS (2002) Biophys J 83: 2393
30. Im W, Roux B (2002) J Mol Biol 319: 1177
31. Massi F, Straub JE (2003) J Comput Chem 24: 143
32. Tieleman DP, Berendsen HJC (1998) Biophys J 74: 2786
33. Roccatano D, Amadei A, Di Nola A, Berendsen HJC (1999) Protein Sci 8: 2130
34. Tieleman DP, Berendsen HJC, Sansom MSP (2001) Biophys J 80: 331
35. Shrivastava IH, Tieleman DP, Biggin PC, Sansom MSP (2002) Biophys J 83: 633
36. Smith LJ, Mark AE, Dobson CM, van Gunsteren WF (1998) J Mol Biol 280: 703
37. De Filippis V, Colombo G, Russo I, Spadari B, Fontana A (2002) Biochemistry 41: 13556
38. Piserchio A, Shimizu N, Gardella TJ, Mierke DF (2002) Biochemistry 41: 13217
39. Schreiber H, Steinhauser O (1992) Chem Phys 168: 75
40. Auffinger P, Beveridge DL (1995) Chem Phys Lett 234: 413
41. Cheatham TEI, Miller JL, Fox T, Darden TA, Kollman PA (1995) J Am Chem Soc 117: 4193
42. Fox T, Kollman PA (1996) Proteins 25: 315

43. Hunenberger PH, McCammon JA (1999) Biophys Chem 78: 69
44. Hunenberger PH, McCammon JA (1999) J Chem Phys 110: 1856
45. Weber W, Hunenberger PH, McCammon JA (2000) J Phys Chem 104: 3668
46. Walser R, Hunenberger PH, van Gunsteren WF (2001) Proteins 44: 509
47. Berendsen HJC, Spoel DVD, Drunen RV (1995) Comput Phys Commun 91: 43
48. Lindahl E, Hess B, van der Spoel D (2001) J Mol Model 7: 306
49. van Gunsteren WF, Daura X, Mark AE (1998) Encyc Comput Chem 2: 1211
50. Miyamoto S, Kollman PA (1992) J Comput Chem 13: 952
51. Hess B, Bekker H, Fraaije J, Berendsen H (1997) J Comput Chem 18: 1463
52. Feenstra KA, Hess B, Berendsen HJC (1999) J Comput Chem 20: 786
53. Berendsen HJC, Grigera JR, Straatsma TP (1987) J Phys Chem 91: 6269
54. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen L (1995) J Chem Phys 103: 8577
55. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR (1984) J Chem Phys 81: 3684
56. Tropp J (1980) J Chem Phys 72: 6035
57. Kabsch W, Sander C (1983) Biopolymers 22: 2576
58. Humphrey W, Dalke A, Schulten K (1996) J Mol Graphics 14: 33
59. Wang H, Sung S (2000) J Am Chem Soc 122: 1999
60. Snow CD, Nguyen H, Pande VS, Gruebele M (2002) Nature 420: 102