PROTEINS: Structure, Function, and Genetics 48:487–496 (2002)

Assessing Equilibration and Convergence in Biomolecular Simulations

Lorna J. Smith,[1]* Xavier Daura,[2] and Wilfred F. van Gunsteren[2]
[1]*Oxford Centre for Molecular Sciences, Central Chemistry Laboratory, University of Oxford, Oxford, United Kingdom*
[2]*Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich, ETH-Hönggerberg, Zürich, Switzerland*

ABSTRACT **If molecular dynamics simulations are used to characterize the folding of peptides or proteins, a wide range of conformational states needs to be sampled. This study reports an analysis of peptide simulations to identify the best methods for assessing equilibration and sampling in these systems where there is significant conformational disorder. Four trajectories of a β peptide in methanol and four trajectories of an α peptide in water, each of 5 ns in length, have been studied. Comparisons have also been made with two 50-ns trajectories of the β peptide in methanol. The convergence rates of quantities that probe both the extent of conformational sampling and the local dynamical properties have been characterized. These include the numbers of hydrogen bonds populated, clusters identified, and main chain torsion angle transitions in the trajectories. The relative equilibrium rates of different quantities are found to vary significantly between the two systems studied reflecting both the differences in peptide primary structure and the different solvents used. A cluster analysis of the simulation trajectories is identified as a very effective method for judging the convergence of the simulations. This is particularly the case if the analysis includes a comparison of multiple trajectories calculated for the same system from different starting structures. Proteins 2002;48:487–496.**

Key words: molecular dynamics; peptide; unfolded conformations; GROMOS; protein folding

INTRODUCTION

Rapid developments in computer power are giving increased possibilities for using molecular dynamics (MD) simulation techniques to provide important insights into protein folding. Recently MD simulations have been reported that characterize at an atomic level the folding of peptides[1–5] and a small protein[6] in explicit solvent. MD simulation techniques are also being used to generate models for folding free energy landscapes and for denatured and partially folded states of proteins.[7–10] In all these simulations compared with those of proteins in their native state, much wider conformational ensembles need to be explored if meaningful results are to be obtained. Therefore, reliable assessments of the equilibration and the extent of sampling within simulations of these conformationally disordered states are required. This is the issue we address here in a study which considers the most effective ways in which the quality of such MD simulations can be judged.

Two peptide systems are analyzed in this work. One of these is a 7-residue β peptide in methanol and the other an 11-residue α peptide in water. The β peptide [Fig. 1(A)] is a non-natural peptide that forms a stable left-handed 3_{14} helix in methanol.[11] MD simulations of this system show a temperature-dependent equilibrium between the folded 3_{14} helix and unfolded conformations.[12] The α peptide sequence [Fig. 1(B)] corresponds to residues 105–115 of the protein hen lysozyme (with Cys 115 changed to serine). These residues form an α helix (helix D) in the native protein.[13] Experimental studies of the isolated peptide, however, show that it is unstructured in aqueous solution.[14] The β and α peptides studied here have a very similar number of rotatable (i.e., nonpeptidic) main chain torsion angles (21 for the β peptide and 22 for the α peptide). However, the characteristics of their sequences differ considerably, all the side chains in the β peptide being aliphatic, whereas the α peptide contains amino acids with hydrophobic, polar, and charged side chains. The contrasting primary structures and the different solvents used in the simulations results in two systems that differ in flexibility and dynamics.

Two 50-ns trajectories and four 5-ns trajectories of the β peptide in methanol are studied in this work. One of the 50-ns simulations was started from the folded 3_{14} helical conformation and was run at 340K (^{340}β1), whereas for the other the initial structure was an extended conformation. A temperature of 360K was used for this second simulation (^{360}β1). Two 5-ns trajectories were taken at different time points from each of the longer 50-ns simulations for analysis (^{340}β2, ^{340}β3, ^{360}β2, ^{360}β3). Each of the 5-ns trajectories of the β peptide therefore has a different

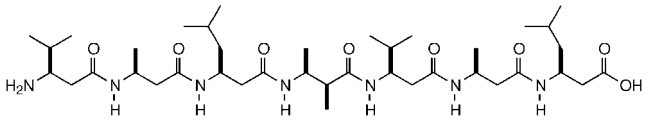
Abbreviations: MD, molecular dynamics; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation.

Grant sponsor: Schweizerischer NationFonds; Grant number: 21-57069.99

*Correspondence to: L.J. Smith, Central Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, OX1 3QH, UK. E-mail: lorna.smith@chem.ox.ac.uk

Received 23 August 2001; Accepted 21 February 2002

A



B

MET–ASN–ALA–TRP–VAL–ALA–TRP–ARG–ASN–ARG–SER
1 2 3 4 5 6 7 8 9 10 11

Fig. 1. (A) Structural formula of the seven residue β peptide studied. (Sequence H-β-HVal-β-HAla-β-HLeu-(S,S)-β-HAla(αMe)-β-HVal-β-HAla-β-HLeu-OH). (B) Amino acid sequence of the 11-residue α peptide studied.

starting structure [Fig. 2(A)]. For the α peptide four 5-ns trajectories are analyzed. Two of these were calculated at a temperature of 300K ($^{300}\alpha 1$ and $^{300}\alpha 2$) and the others at 450K ($^{450}\alpha 1$ and $^{450}\alpha 2$). Trajectories $^{300}\alpha 1$ and $^{450}\alpha 1$ were taken at different time points from a simulation that was started from the α helical conformation this sequence adopts in the X-ray structure of lysozyme. The other two trajectories, $^{300}\alpha 2$ and $^{450}\alpha 2$, come from different points in a simulation in which alternating residues in the starting structure adopted α and β main chain conformations. As for the β peptide, there is a different initial conformation used for each of the four α peptide trajectories [Fig. 2(B)].

The convergence and equilibration of six different quantities in these peptide trajectories is determined. The quantities are selected to characterize a range of properties of the peptides including their energy, conformation, hydrogen bonding, and local dynamics. Comparisons between the trajectories used enable the dependence of the convergence on the inherent flexibility of the covalently bonded chain, the solvent characteristics, the temperature, and the starting conformation of the trajectory to be probed. Through this, insight is gained into the most appropriate manner of analyzing MD simulations in order to assess their convergence and quality.

METHODS

All simulations were carried out using the GROMOS96 package of programs and the GROMOS96 43A1 force field.[15] The simulations were run in the presence of explicit solvent (methanol for the β peptide and water for the α peptide) with periodic boundary conditions. For the β peptide the initial structure of the peptide for the simulation at 340K ($^{340}\beta 1$) was the 3_{14} helical structure.[11] The system contained the β-heptapeptide and 962 methanol molecules in a rectangular box. The simulation was run for 50 ns, and the time blocks from 0 to 5 ns and 25 to 30 ns were used for trajectories $^{340}\beta 2$ and $^{340}\beta 3$, respectively. For the 360K simulation ($^{360}\beta 1$) the peptide was initially fully extended (all backbone dihedral angles set to 180°), and the system contained the β-heptapeptide and 1778 methanol molecules in a truncated octahedral box. The simulation was run for 50 ns, and the time blocks from 0 to 5 ns and 35 to 40 ns were used for trajectories $^{360}\beta 2$ and $^{360}\beta 3$, respectively. In both the β peptide simulations the dimensions of the periodic box were chosen so that the minimum distance from the peptide to the box wall was 1.4 nm in the starting configuration.[1] Details of the β peptide simulations have been reported previously.[1,12] The diffusion constant for the methanol model used is at 300K and 1 atm a factor of 1.3 larger than the experimental value.

For the α peptide, for simulation A the starting peptide coordinates were those of residues 105–115 in the crystal structure of triclinic hen lysozyme (pdb code 2LZT).[16] In the α peptide the C-terminal residue, which is a cysteine in hen lysozyme, was changed to serine, and low pH conditions were modeled by protonating the C-terminus of the peptide. In simulations B the starting coordinates were generated by modifying those used to start simulation A from the crystal structure, so that the main chain of residues 2, 4, 6, 8, and 10 adopted β conformations (ϕ −113°, ψ 123°) instead of α conformations (ϕ −57°, ψ −47°). Truncated octahedron periodic boundary conditions were used in all the simulations, and the initial dimensions of the periodic box were chosen so that the minimum distance from the peptide to the box wall in the starting configuration was 1.5 nm in simulation A and 1.05 nm in simulations B. In each case the system contained the α peptide and 2105 simple point charge water molecules.[17] In simulation A the following temperatures were used: 300K for 0–1 ns, 350K for 1–1.3 ns, 400K for 1.3–2.3 ns, 450K for 2.3–5.3 ns, 400K for 5.3–6.3 ns, 350K for 6.3–6.4 ns and 300K for 6.4–11.4 ns. In addition, branching from the simulation at 5.3 ns, the trajectory was continued out to 7.3 ns at 450K. The simulation from 6.4 to 11.4 ns at 300K was used as trajectory $^{300}\alpha 1$ in this work and that from 2.3 to 7.3 ns at 450K as trajectory $^{450}\alpha 1$. One simulation B was run from 0 to 6 ns at 300K. The time block from 1 to 5 ns was used as trajectory $^{300}\alpha 2$. A second simulation B was run from 0 to 5 ns at 450K, and this was used as trajectory $^{450}\alpha 2$. The diffusion constant of the SPC water model used is at 300K and 1 atm a factor of 1.9 larger than the experimental value.

In all simulations of the β and α peptides a time step of 2 fs was used. The simulations were performed at constant pressure (1 atm), the temperature and pressure being maintained by weak coupling to an external bath[18] [temperature coupling relaxation time 0.1 ps; pressure coupling relaxation time 0.5 ps; isothermal compressibility κ_T $4.575\ 10^{-4}$(kJ mol^{-1} $nm^{-3})^{-1}$]. Throughout the simulations bond lengths were constrained to ideal values using the SHAKE procedure with a geometric accuracy of 10^{-4}.[19] Nonbonded interactions were treated using a twin range method.[20] Within a short-range cutoff of 0.8 nm, all interactions were determined at every step. Longer range (electrostatic and van der Waals) interactions within a cutoff range of 1.4 nm were updated at the same time as the pair list was generated (every 10 fs). A reaction field was applied (ϵ 54.0)[21] beyond a cutoff of 1.4 nm in the simulations of the α peptide in water.

Analysis of the trajectories was performed using programs from GROMOS96.[15] The following quantities were determined for each trajectory:

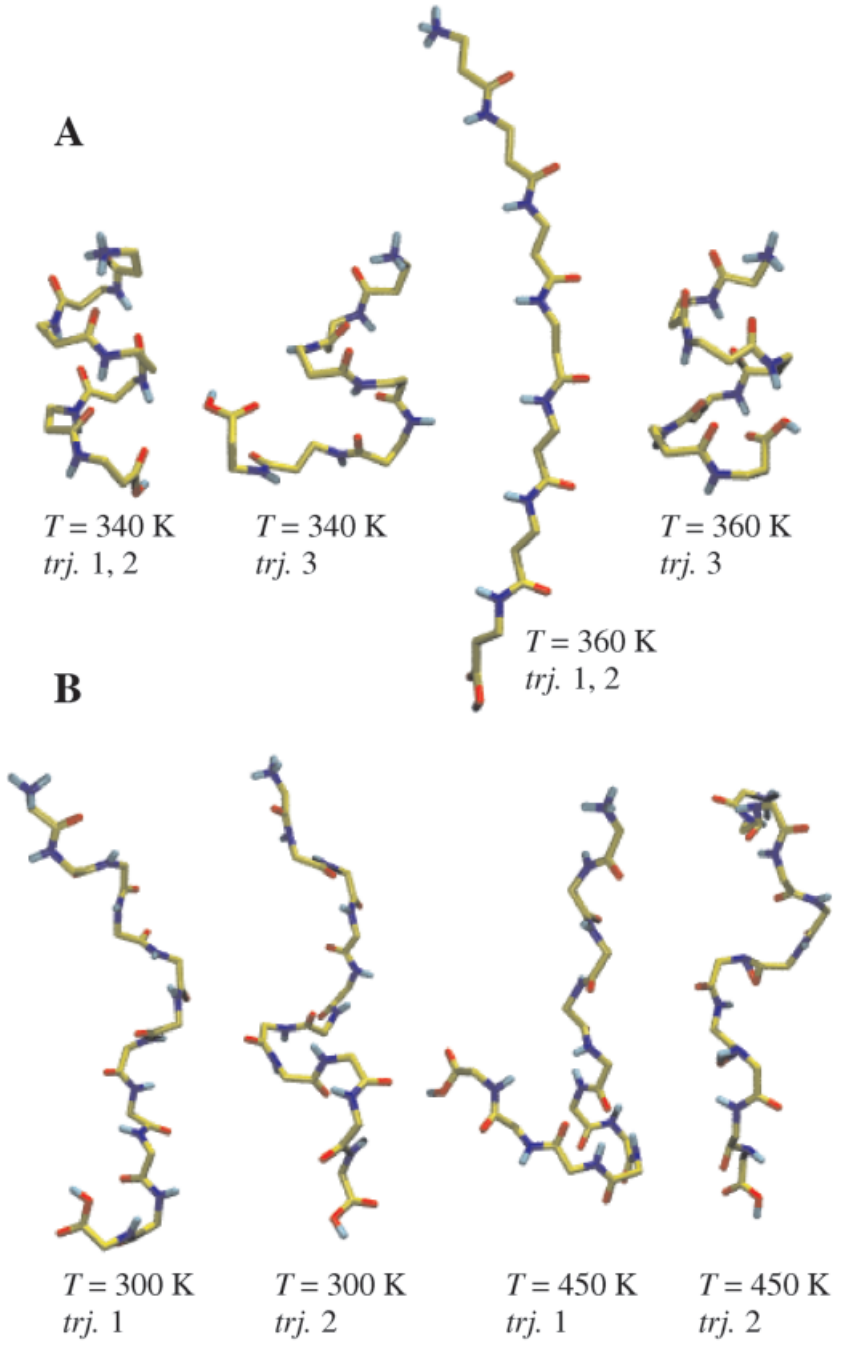


Fig. 2. Backbone conformation of the first structure of each of the trajectories analyzed. (A) β peptide, from left to right: first structure of trajectories $^{340}\beta 1$ and $^{340}\beta 2$ at 340K; first structure of trajectory $^{340}\beta 3$ at 340K; first structure of trajectories $^{360}\beta 1$ and $^{360}\beta 2$ at 360K; first structure of trajectory $^{360}\beta 3$ at 360K. (B) α peptide, from left to right: first structure of trajectory $^{300}\alpha 1$ at 300K; first structure of trajectory $^{300}\alpha 2$ at 300K; first structure of trajectory $^{450}\alpha 1$ at 450K; first structure of trajectory $^{450}\alpha 2$ at 450K.

1. The intramolecular interaction energy of the peptide (Fig. 3).
2. The main chain atom positional root-mean-square deviation with respect to the initial structure of the trajectory (Fig. 4). In calculating these values the N, Cβ, Cα, and C atoms of residues 2–6 were used for the β peptide and the N, Cα, and C atoms of residues 2–10 were used for the α peptide.
3. The number of clusters present (Fig. 5). For the cluster analysis structures of the peptide were extracted from the trajectories at 0.01-ns intervals. The clustering was performed in cartesian space. For each pair of structures a least squares translational and rotational fit was performed. This for the β peptide used the peptide backbone atoms (N, Cβ, Cα, C) of residues 2–6 and for the α peptide the backbone atoms (N, Cα, C) of residues

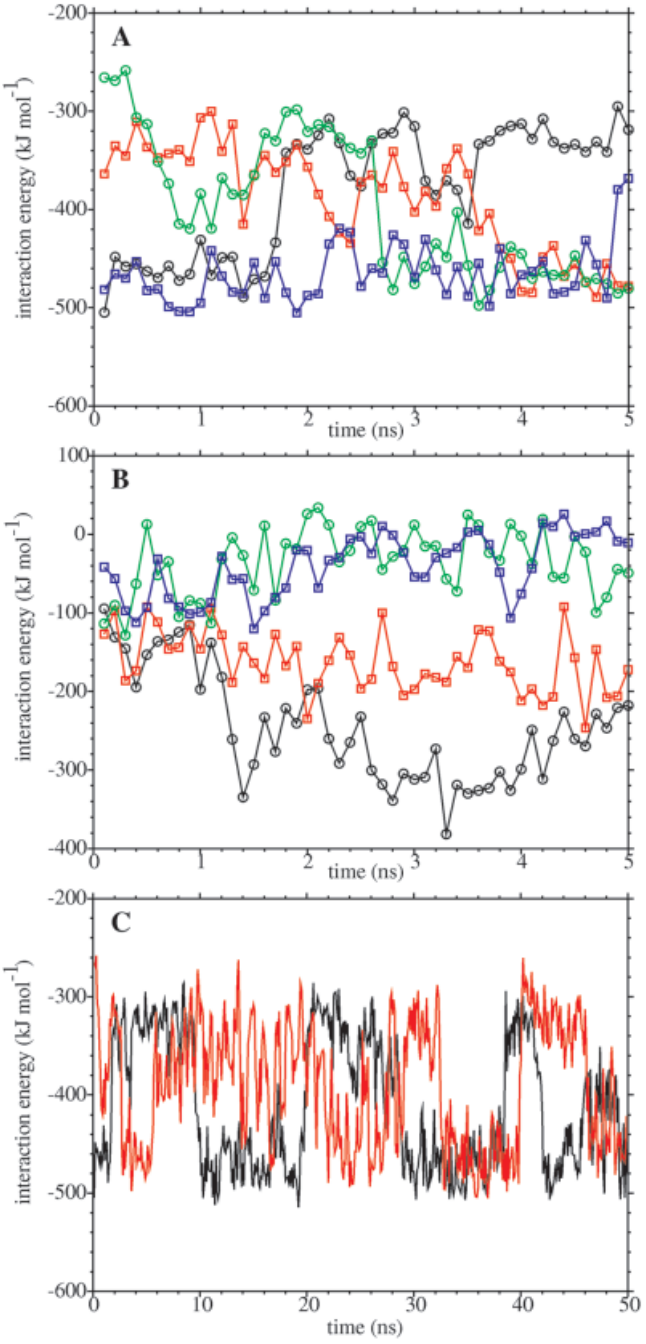


Fig. 3. Intramolecular interaction energy of the peptide as a function of time, averaged over 0.1-ns windows. (A) β peptide: data from trajectories ^{340}β2 (black) and ^{340}β3 (red) at 340K, and from trajectories ^{360}β2 (green) and ^{360}β3 (blue) at 360K. (B) α peptide: data from trajectories ^{300}α1 (black) and ^{300}α2 (red) at 300K, and from trajectories ^{450}α1 (green) and ^{450}α2 (blue) at 450K. (C) β peptide: data from trajectory ^{340}β1 at 340 K (black) and trajectory ^{360}β1 at 360K (red).

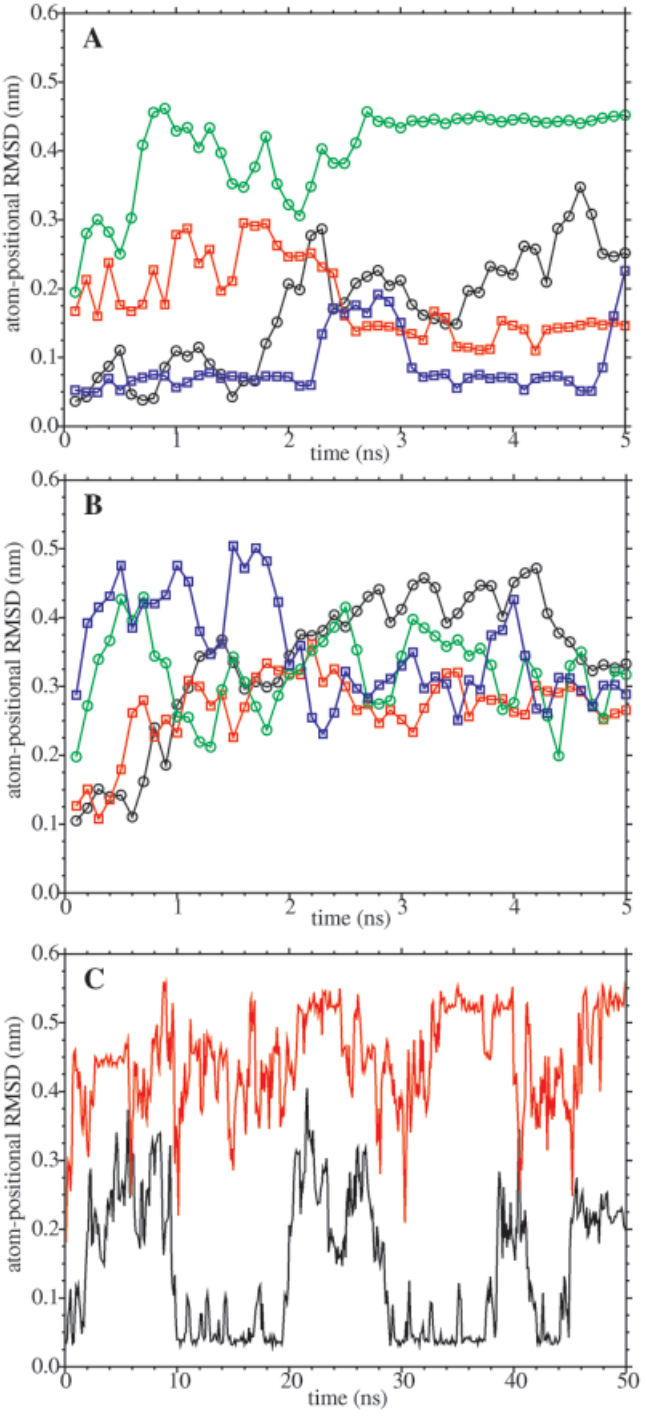


Fig. 4: Peptide main chain (excluding N- and C-terminal residues) atom positional root-mean-square deviation (RMSD) from the initial structure (see Fig. 2) as a function of time, averaged over 0.1-ns windows. (A) β peptide: data from trajectories ^{340}β2 (black) and ^{340}β3 (red) at 340K, and from trajectories ^{360}β2 (green) and ^{360}β3 (blue) at 360K. (B) α peptide: data from trajectories ^{300}α1 (black) and ^{300}α2 (red) at 300K, and from trajectories ^{450}α1 (green) and ^{450}α2 (blue) at 450K. (C) β peptide: data from trajectory ^{340}β1 at 340K (black) and trajectory ^{360}β1 at 360K (red).

2–10. The matrix of atom positional root-mean-square deviations (RMSD) between pairs of structures was calculated for these sets of atoms. The criteria of similarity for two structures were a backbone atom positional RMSD ≤0.10 nm for residues 2–6 for the β peptide and ≤0.15 nm for residues 2–10 for the α peptide. In each case with these criteria the number of structural neighbors was determined for each structure. The structure with the highest number of struc-

tural neighbors was taken as the centre member of the first cluster of structures. All the structures belonging to this cluster were thereafter removed from the pool. For each of the remaining structures the number of structural neighbors was again calculated. The structure with the most neighbors then became the centre member of the second cluster of structures. Structures belonging to this second cluster were then also removed from the pool. The process was iterated until all structures were assigned to a cluster. Further details of this procedure are given in Reference 12.

4. The number of different intramolecular hydrogen bonds populated in the peptide (Fig. 6). Hydrogen bonds were identified using the cutoff criteria of a maximum distance between the hydrogen atom and the acceptor atom of 0.25 nm and a minimum angle between the donor atom, the hydrogen and the acceptor atom of 135°.
5. The atom positional root-mean-square fluctuations (Fig. 7). These values were averaged over all peptide atoms.
6. The total number of torsion angle transitions and the number of main chain torsion angle transitions (excluding the two main chain torsion angles at either end of the peptides, Table I). A transition is counted when the torsion angle passes through the minimum of an adjacent well in the torsion angle energy function.

In addition, for each peptide the two 5-ns trajectories at the same temperature were merged, and the hydrogen bond and cluster analyses were performed on the combined trajectory. To do this at each time point the initial x ns of the first trajectory were combined with the initial x ns of the second trajectory, and the quantities were calculated using the $2x$-ns time period. The values calculated are plotted at the x-ns time point in Figures 5 and 6 (triangle symbols).

RESULTS AND DISCUSSION

General Characteristics

We concentrate first on the eight trajectories of 5 ns. Four of these are of the β peptide in methanol ($^{340}\beta2$, $^{340}\beta3$, $^{360}\beta2$, $^{360}\beta3$) and four are of the α peptide in water ($^{300}\alpha1$, $^{300}\alpha2$, $^{450}\alpha1$, $^{450}\alpha2$). Variations in the intramolecular interaction energy of the peptide, the main chain atom positional RMSD with respect to the initial structure, the number of clusters present, and the number of different intrapeptide hydrogen bonds populated through these 5-ns trajectories are shown in Figures 3–6, respectively (panel A for the β peptide and panel B for the α peptide). For the hydrogen bonds and clusters in addition to determining the quantities for the individual trajectories the values for a merged trajectory at each temperature have been calculated as described in the Methods section. The analyses of the merged trajectories give insight into the similarity of the conformations that are sampled in the individual trajectories. They therefore probe the dependence of the results on the initial structure used for the simulation.

Various overall characteristics are immediately apparent from the data. The peptide interaction energy and

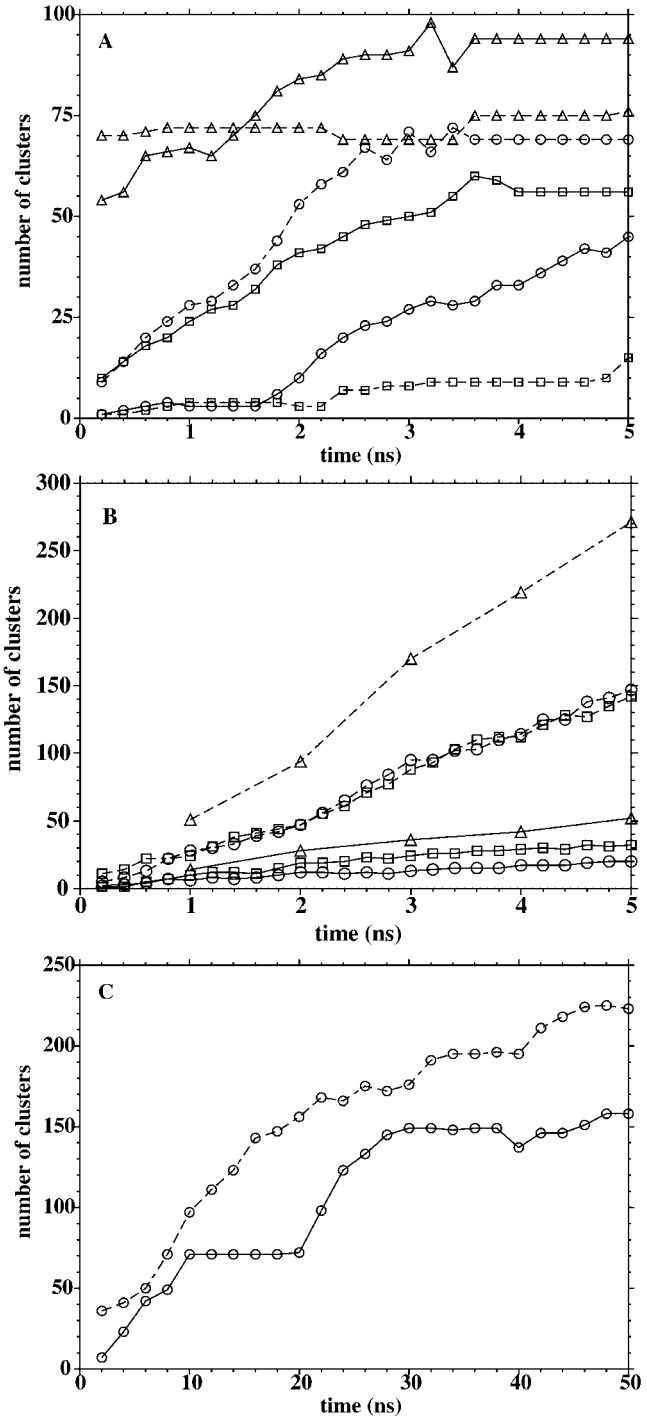


Fig. 5. Number of clusters as a function of cumulative time. (A) β peptide: data from trajectories $^{340}\beta2$ (solid line, circles) and $^{340}\beta3$ (solid line, squares) at 340K and from a merged trajectory including the previous two (solid line, triangles); data from trajectories $^{360}\beta2$ (dashed line, circles) and $^{360}\beta3$ (dashed line, squares) at 360K and from a merged trajectory including the previous two (dashed line, triangles). (B) α peptide: data from trajectories $^{300}\alpha1$ (solid line, circles) and $^{300}\alpha2$ (solid line, squares) at 300K and from a merged trajectory including the previous two (solid line, triangles); data from trajectories $^{450}\alpha1$ (dashed line, circles) and $^{450}\alpha2$ (dashed line, squares) at 450K and from a merged trajectory including the previous two (dashed line, triangles). (C) β peptide: data from trajectory $^{340}\beta1$ at 340K (solid line) and trajectory $^{360}\beta1$ at 360K (dashed line).

atom positional RMSD values in general equilibrate quickly, and for a given trajectory similar equilibration times are observed for these two quantities (Figs. 3 and 4). This is seen, for example, in the $^{360}\beta 2$ trajectory, which starts from an extended peptide conformation, where both these quantities equilibrate within 2.5–3 ns [Figs. 3(A) and 4(A)]. The length of time required for equilibration for these two quantities has a clear temperature dependence, slower oscillations in the values being seen at low temperatures. For example, for the α peptide at 450K equilibration for the peptide interaction energy and atom positional RMSD occurs within 2 ns, whereas at 300K equilibration is not observed for the $^{300}\alpha 1$ trajectory within 5 ns [Figs. 3(B) and 4(B)].

Compared with the equilibration times for the peptide interaction energy and atom positional RMSD, much longer times are required for the total number of clusters present to equilibrate (Fig. 5), with the clusters defined with the backbone RMSD cut-off criteria used here (≤0.10 nm for residues 2–6 for the β peptide and ≤0.15 nm for residues 2–10 the α peptide). For example, for the two α peptide trajectories at 450K, $^{450}\alpha 1$ and $^{450}\alpha 2$, the peptide interaction energy equilibrates within 2 ns [Fig. 3(B)], but the total number of clusters present [Fig. 5(B)] continues to rise throughout the 5-ns trajectories. Interestingly, the relative equilibration times for the number of hydrogen bonds populated (Fig. 6) differ significantly between the two systems studied. For the β peptide in methanol initial equilibration occurs within 3 ns [Fig. 6(A)], and therefore with a similar timescale to the peptide interaction energy and atom positional RMSD. In contrast, equilibration of the number of different hydrogen bonds populated for the α peptide in water is not observed in any of the trajectories within the 5 ns analyzed [Fig. 6(B)].

Quantities such as the total number of clusters and the number of different hydrogen bonds populated provide an assessment of how much conformational space has been populated in the simulations. We have also analyzed the equilibration of two quantities that give insight into the local dynamics of the peptide chain. These are the atom positional root-mean-square fluctuations averaged over all peptide atoms (Fig. 7) and the number of torsion angle transitions (all and main chain) in the trajectories (Table I). The values of both these quantities after equilibration depend on the simulation temperature. For example, for the $^{300}\alpha 1$ trajectory 342 main-chain torsion angle transitions are observed from 4 to 5 ns, compared with 1532 transitions for the $^{450}\alpha 1$ trajectory (Table I). However, as for the hydrogen bonds, significant differences are observed between the times required for equilibration for the α and β peptide systems. Indeed for the β peptide in methanol equilibration is not reached within 5 ns [Fig. 7(A)]. We consider this in more detail later.

The values of each of the quantities considered have been compared in the different trajectories after equilibration. Some of the quantities equilibrate to very similar values in all the trajectories, whereas other quantities are much more conformationally sensitive. These later quantities therefore distinguish much more effectively between

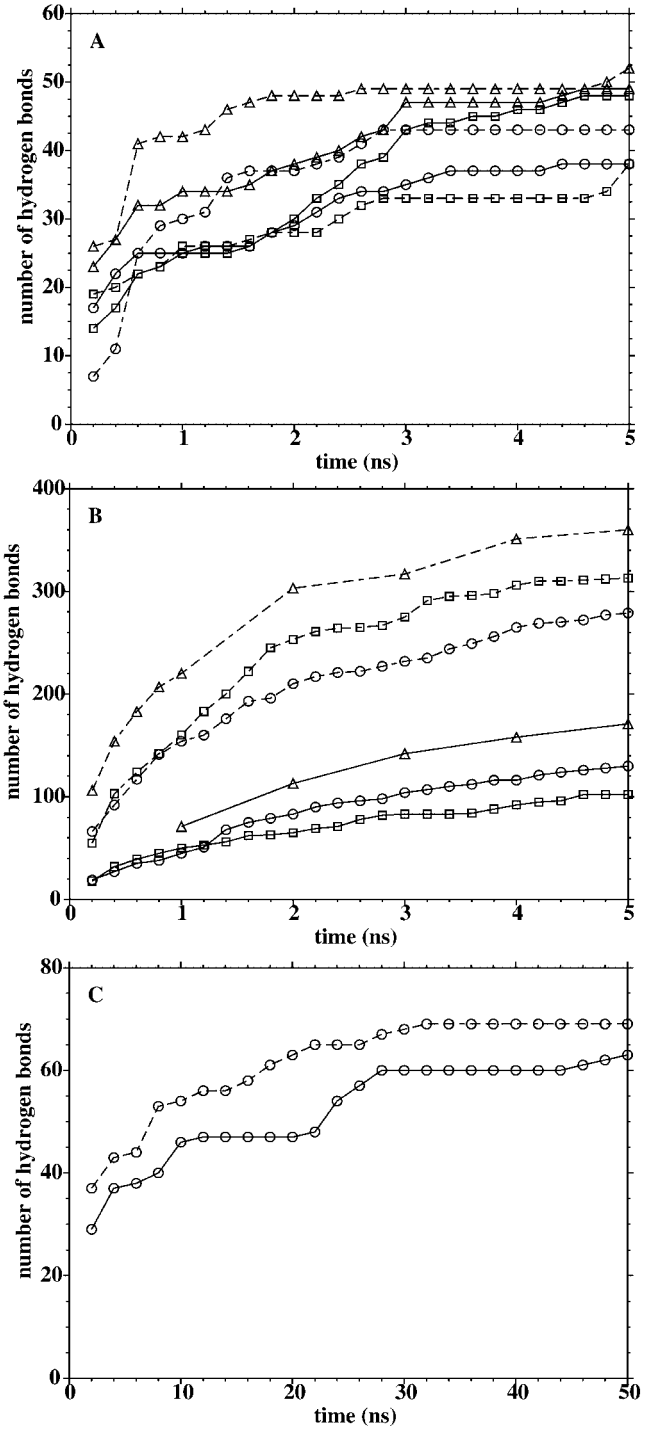


Fig. 6. Number of different intrapeptide hydrogen bonds as a function of cumulative time. (A) β peptide: data from trajectories $^{340}\beta 2$ (solid line, circles) and $^{340}\beta 3$ (solid line, squares) at 340K and from a merged trajectory including the previous two (solid line, triangles); data from trajectories $^{360}\beta 2$ (dashed line, circles) and $^{360}\beta 3$ (dashed line, squares) at 360K and from a merged trajectory including the previous two (dashed line, triangles). (B) α peptide: data from trajectories $^{300}\alpha 1$ (solid line, circles) and $^{300}\alpha 2$ (solid line, squares) at 300K and from a merged trajectory including the previous two (solid line, triangles); data from trajectories $^{450}\alpha 1$ (dashed line, circles) and $^{450}\alpha 2$ (dashed line, squares) at 450K and from a merged trajectory including the previous two (dashed line, triangles). (C) β peptide: data from trajectory $^{340}\beta 1$ at 340K (solid line) and trajectory $^{360}\beta 1$ at 360K (dashed line).

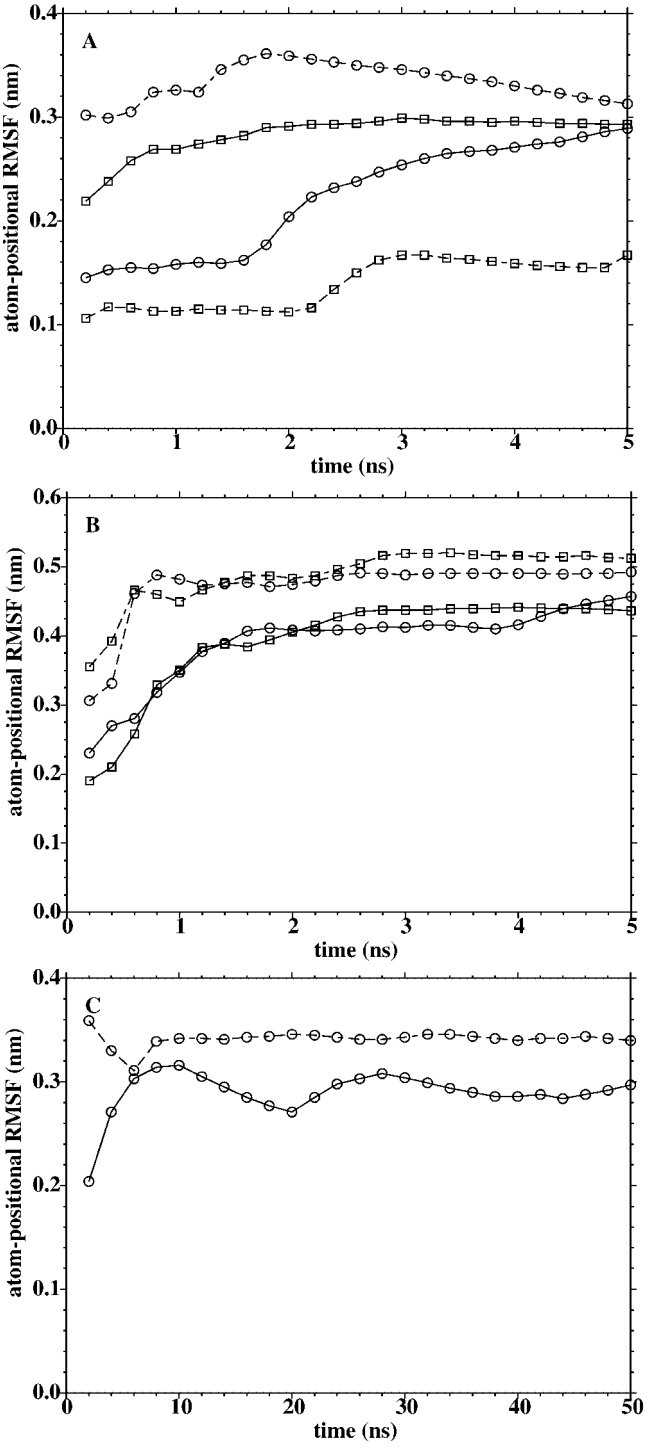


Fig. 7: Atom positional root-mean-square-fluctuations (RMSF), averaged over all atoms, as a function of cumulative time. (A) β peptide: data from trajectories $^{340}\beta 2$ (solid line, circles) and $^{340}\beta 3$ (solid line, squares) at 340K; data from trajectories $^{360}\beta 2$ (dashed line, circles) and $^{360}\beta 3$ (dashed line, squares) at 360K. (B) α peptide: data from trajectories $^{300}\alpha 1$ (solid line, circles) and $^{300}\alpha 2$ (solid line, squares) at 300K; data from trajectories $^{450}\alpha 1$ (dashed line, circles) and $^{450}\alpha 2$ (dashed line, squares) at 450K. (C) β peptide: data from trajectory $^{340}\beta 1$ at 340K (solid line) and trajectory $^{360}\beta 1$ at 360K (dashed line).

the behavior of the peptides in the different trajectories. Most notably the number of clusters present is much more structurally informative than the atom positional RMSD from the starting structure. For example, after ~2 ns all the α peptide simulations have similar atom positional RMSD values [Fig. 4(B)]. However, Figure 5(B) demonstrates that each of the trajectories at 450K has by 5 ns populated more than four times the total number of clusters present at 300K. This comparison of the clusters also highlights the temperature dependence of the behavior of the simulations.

Trajectories with Limited Sampling of Conformational Space

To analyze the data in more detail, we concentrate first on the trajectories where the sampling of conformational space within 5 ns is limited. This limited sampling is either because of a low temperature or because the starting conformation for the simulation is stabilized by features such as persistent intrapeptide hydrogen bonds. The simulations in this category are $^{360}\beta 3$, $^{300}\alpha 1$, and $^{300}\alpha 2$. For the $^{360}\beta 3$ trajectory the equilibration of almost all the quantities considered is quite fast. Thus the number of different hydrogen bonds [Fig. 6A)] and the total number of clusters [Fig. 5(A)] stabilize within 3 ns (excluding the final 200 ps where the helix starts to unfold). Equilibration for the α peptide trajectories is slower. This is likely to reflect at least in part the lower temperature of the α peptide simulation (300K compared with 360K for the β peptide).

For all these trajectories, however, the main characteristic is a strong dependence of the quantities on the region of conformational space that is being explored and hence on the starting structure. For both the α peptide trajectories there is a low total number of clusters (≤32 at 5 ns) and no overlap between the clusters populated in the $^{300}\alpha 1$ and $^{300}\alpha 2$ runs [Fig. 5(B)]. Similarly the $^{360}\beta 3$ trajectory has the lowest number of clusters and the lowest total number of hydrogen bonds of any of the four β peptide simulations [Fig. 5(A)]. Additionally the number of torsion angle transitions is significantly different between the two α peptide simulations at 300K even at the end of the trajectories analyzed (Table I). From 4 to 5 ns there are 342 main chain torsion angle transitions for $^{300}\alpha 1$ compared with 559 transitions for $^{300}\alpha 2$. This contrasts with the 450K α peptide simulations where a closely similar number of main chain torsion angle transitions is observed for the two trajectories over the same 4- to 5-ns time period (1532 transitions for $^{450}\alpha 1$ and 1575 transitions for $^{450}\alpha 2$).

Trajectories Without a Dominant Conformation

We now consider the trajectories in which there appears to be no significantly preferred conformation over the 5-ns time period. In particular this class includes the trajectories $^{340}\beta 3$, $^{360}\beta 2$, $^{450}\alpha 1$, and $^{450}\alpha 2$. Comparison of the two β peptide trajectories at 360K ($^{360}\beta 2$ and $^{360}\beta 3$) shows that the apparent equilibration times are similar for the trajectories where there is no preferred conformation and those where there is limited sampling of conformational space

TABLE I. The Number of Torsion Angle Transitions in the 5-ns Trajectories of the β Peptide in Methanol and the α Peptide in Water

Trajectory	Transitions[a]	0–1 ns	1–2 ns	2–3 ns	3–4 ns	4–5 ns	0–5 ns
			β Peptide in Methanol				
$^{340}\beta 2$	All	651	669	780	756	753	3597
	Main chain	237	262	373	393	388	1655
$^{340}\beta 3$	All	845	790	759	645	583	3245
	Main chain	487	436	400	311	174	1806
$^{360}\beta 2$	All	1005	778	857	598	615	2962
	Main chain	568	539	403	199	189	1892
$^{360}\beta 3$	All	513	392	506	408	551	2411
	Main chain	217	166	238	147	286	1068
			α Peptide in Water				
$^{300}\alpha 1$	All	1222	858	831	967	883	4761
	Main chain	543	316	355	443	342	1980
$^{300}\alpha 2$	All	1091	1266	1090	1050	1221	5742
	Main chain	499	583	543	493	559	2577
$^{450}\alpha 1$	All	3782	3402	3659	3396	2861	17100
	Main chain	1661	1493	1597	1428	1532	7711
$^{450}\alpha 2$	All	5498	3589	3492	3602	3574	19755
	Main chain	3057	1515	1500	1578	1575	9225

[a]The total number of torsion angle transitions and the number of main chain transitions (excluding the two main chain dihedral angles at either end of the peptides) are listed.

under the same temperature conditions. However, the equilibrated values of the quantities are significantly higher for the trajectories with no dominant conformation and their dependence on the starting conformation of the peptide is much more limited. For example, the total number of hydrogen bonds for both the $^{360}\beta 2$ and $^{360}\beta 3$ trajectories equilibrates within 3 ns [Fig. 6(A)]. However, the final value for the $^{360}\beta 2$ simulation is 43 hydrogen bonds, significantly higher than the final value of 34 hydrogen bonds for the $^{360}\beta 3$ trajectory where there is limited conformational sampling. Similarly, the high number of clusters populated by these trajectories with no dominant conformation (Fig. 5) demonstrates how much more conformational space is explored in these trajectories. Thus for the $^{450}\alpha 1$ and $^{450}\alpha 2$ trajectories at 5 ns 147 and 142 clusters, respectively, have been populated compared with the $^{300}\alpha 1$ and $^{300}\alpha 2$ trajectories where 20 and 32 clusters respectively have been populated.

It is interesting to note that for the α peptide at 450K there is still almost no overlap between the clusters that are populated in the $^{450}\alpha 1$ and $^{450}\alpha 2$ trajectories [271 clusters for the combined two 5-ns trajectories compared with 147 and 142 clusters for the individual trajectories, Fig. 5(B)]. However, the total number of hydrogen bonds if the two 5-ns trajectories are combined (360) is only slightly higher than the hydrogen bond numbers for the trajectories alone [271 $^{450}\alpha 1$, 313 $^{450}\alpha 2$, Fig. 6(B)]. This contrast between the hydrogen bonds and the clusters reflects the fact that the same hydrogen bond can be adopted in one part of the peptide, whereas a significantly different conformation is being adopted in another region of the sequence.

Comparison of the β Peptide in Methanol and the α Peptide in Water

For some quantities such as the atom positional RMSD very similar values are seen after equilibration in the trajectories of the β peptide in methanol and the α peptide in water. However, comparison of the atom positional root-mean-square fluctuation (RMSF) values (Fig. 7) and the number of torsion angle transitions (Table I) in the α and β peptide trajectories indicates that there are significant differences in the local dynamics of these two systems. Thus, for the α peptide at 300K the atom positional RMSF values are higher (0.457 nm for $^{300}\alpha 1$ at 5 ns) than for the β peptide at 340K (0.289 nm for $^{340}\beta 2$ at 5 ns). In a similar way from 4 to 5 ns 883 and 1221 torsion angle transitions are observed for the $^{300}\alpha 1$ and $^{300}\alpha 2$ trajectories, respectively, compared with 753 and 583 torsion angle transitions for the $^{340}\beta 2$ and $^{340}\beta 3$ trajectories, respectively. In addition, equilibration is observed much more quickly for these quantities for the α peptide trajectories. For example, the atom positional RMSF values equilibrate within 3 ns for the α peptide in water [Fig. 7(B)], but equilibration is not seen for the β peptide within 5 ns [Fig. 7(A)].

Another interesting difference between the two sets of simulations is seen when the total number of intrapeptide hydrogen bonds populated is compared. Significantly more different hydrogen bonds are observed for the α peptide simulations in water [Fig. 6(B)] than for those for the β peptide in methanol [Fig. 6(A)]. Indeed the combined total number of hydrogen bonds in the $^{450}\alpha 1$ and $^{450}\alpha 2$ simulations within 5 ns (360) is more than five times larger that the combined total (69) for the two β peptide trajectories at

360K over 50 ns. In addition, as mentioned previously, equilibration of the total number of hydrogen bonds is much slower in the α peptide simulations than in those of the β peptide. In this case these differences are likely to result from the larger number of potential hydrogen bond donors and acceptors present in the α peptide (44) compared with the β peptide (15). There may also be some effects from the different hydrogen bonding capabilities of the two solvents used, although this will have a much more significant effect on the peptide-solvent hydrogen bonds than on the intrapeptide hydrogen bonds considered here. The significant number of possible hydrogen bonds and the slow equilibration of the total number of hydrogen bonds populated for the α peptide is reflected in the long simulation times required for the folding of even short peptides such as this in MD simulations in aqueous solution.

Comparison With 50-ns Trajectories

The results from the analysis of the 50-ns trajectories of the β peptide in methanol (^{340}β1 and ^{360}β1) have been compared with those for the 5-ns trajectories of this system at 340K and 360K. This comparison is particularly important as it gives insight into the reliability of the conclusions drawn from the shorter simulation times. Variations in the peptide interaction energy in the range −280 to −520 kJ mol^{-1} are seen through the 50 ns in both the β peptide trajectories with faster fluctuations occurring at 360K than at 340K [Fig. 3(C)]. Similar fluctuations are seen in the atom-positional RMSD values [Fig. 4(C)], the large changes in this quantity reflecting transitions from the folded helical conformation to unfolded conformers through the 50 ns. For both the peptide interaction energy and the atom positional RMSD similar values of the quantities are seen in the 5-ns trajectories analyzed and throughout the 50-ns trajectories.

In contrast to the fast fluctuations in the interaction energy and atom-positional RMSD, analysis of the 50-ns trajectories shows that very long simulation times are needed for the number of different hydrogen bonds present [Fig. 6(C)] and the total number of clusters populated [Fig. 5(C)] to equilibrate. The values of both these quantities rise slowly in the simulations. The number of hydrogen bonds present levels off after ~30 ns, but the total number of clusters increases throughout the 50 ns. Somewhat faster equilibration is seen for the atom positional RMSF, equilibration at 360K occurring within 10 ns [Fig. 7(C)]. Therefore, for these quantities the apparent equilibration observed in many of the 5-ns trajectories analyzed was only a temporary one. For example, the increase in the number of different hydrogen bonds populated in the ^{360}β2 trajectory flattens off at a total of 43 hydrogen bonds within 3 ns [Fig. 6(A)]. However, analysis of the ^{360}β1 trajectory shows that a higher equilibrated value of 69 hydrogen bonds is reached after ~35 ns of simulation time [Fig. 6(C)]. Thus to obtain realistic results for simulations where a large region of conformational space needs to be explored, as for an unfolded peptide, very long simulation times are needed. This is particularly important because, of the quantities considered here, those that are most conformationally informative, such as the number of hydrogen bonds and the total number of clusters present, take the longest times to equilibrate.

CONCLUSIONS

This study has analyzed a range of different quantities. For assessing the quality of simulations of unfolded peptides and other disordered systems we have identified that it is important to probe both the conformational and dynamical properties of the peptide. These properties can have very different convergence times. For the two peptides considered here the convergence of the conformational characteristics is seen to be much slower for the α peptide in water than for the β peptide in methanol. However, the local dynamical properties of the α peptide equilibrate much more quickly than those of the β peptide. These differences reflect the characteristics of the peptide main chain in α and β peptides, the side chain groups in the peptide sequences and the different viscosities and hydrogen bonding capabilities of the methanol and water solvents.

Quantities such as the peptide interaction energy and atom positional RMSD from the starting structure are often used in the initial analysis of MD simulations as criteria for convergence. The values of these quantities have been found to stabilize quickly for the peptides studied here, equilibration being achieved in general within ~3 ns. For simulations of native folded proteins the atom positional RMSD values from the starting structure often provide a good method for judging the convergence of the simulations. However, comparisons here show that simulations with very different characteristics can have very similar atom positional RMSD values. Instead, when there is significant conformational disorder in the trajectory, we have shown that the number of clusters populated is a much more informative quantity for assessing the convergence of the simulations. Thus the atom positional RMSD fluctuates within the range 0.2–0.5 nm for the α peptide simulations at 450K, but the number of clusters present continues to rise steadily through the 5-ns trajectories. This reflects the fact that new regions of conformational space are being sampled throughout the simulations.

Comparisons between the trajectories studied here indicated that the similarity of the quantities for two trajectories of the same peptide calculated from different starting structures provides a very objective way of assessing simulation convergence. This is particularly the case if the results of a cluster analysis of the two trajectories merged together are compared with the cluster analyses of the individual trajectories. Such analyses highlight the fact that very long simulations times are required for adequate sampling in systems where there is significant conformational disorder. This is the case even for short peptide fragments and elevated temperatures as illustrated by the data for the 11 residue α peptide at 450K presented here. Indeed, there is almost no overlap between the clusters populated in the two 5-ns α peptide trajectories at 450K. At present, detailed characterisation using multiple trajec-

tories calculated from different initial conformations is in general only feasible for small system such as short peptides. However, developments in computing power should enable this type of study to become increasingly a realistic aim for simulations of larger systems, such as partially folded proteins in explicit solvent.

ACKNOWLEDGMENTS

L.J. Smith is a Royal Society University Research Fellow. Financial support was obtained from the Schweizerischer NationalFonds, project number 21-57069.99, which is gratefully acknowledged.

REFERENCES

1. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Reversible peptide folding in solution by molecular dynamics simulation. J Mol Biol 1998;280:925–932.
2. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. Angew Chem Int Ed Eng 1999;38:236–240.
3. Takano M, Yamato T, Higo J, Suyama A, Nagayama K. Molecular dynamics of a 15-residue poly(L-alanine) in water: helix formation and energetics. J Am Chem Soc 1999;121:605–612.
4. Bonvin AMJJ, van Gunsteren WF. β-Hairpin stability and folding: Molecular dynamics studies of the first β-hairpin of tendamistat. J Mol Biol 2000;296:255–268.
5. Daura X, Gademann K, Schafer H, Jaun B, Seebach D, van Gunsteren WF. The β-peptide hairpin in solution: conformational study of a β-hexapeptide in methanol by NMR spectroscopy and MD simulation. J Am Chem Soc 2001;123:2393–2404.
6. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744.
7. Bursulaya BD, Brooks CL. Comparative study of the folding free energy landscape of a three-stranded β-sheet protein with explicit and implicit solvent models. J Phys Chem B 2000;104:12378–12383.
8. Smith LJ, Dobson CM, van Gunsteren WF. Side-chain conformational disorder in a molten globule: molecular dynamics simulations of the A-state of human α-lactalbumin. J Mol Biol 1999;286:1567–1580.
9. Wong KB, Clarke J, Bond CJ, Neira JL, Freund SMV, Fersht AR, Daggett V. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. J Mol Biol 2000;296:1257–1282.
10. Paci E, Smith LJ, Dobson CM, Karplus M. Exploration of partially unfolded states of human α-lactalbumin by molecular dynamics simulation. J Mol Biol 2001;306:329–347.
11. Seebach D, Ciceri PE, Overhand M, Jaun B, Rigo D, Oberer L, Hommel U, Amstutz R, Widmer H. Probing the helical secondary structure of short-chain β-peptides. Helv Chim Acta 1996;79:2043–2066.
12. Daura X, van Gunsteren WF, Mark AE. Folding-unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. Proteins Struct Funct Genet 1999;34:269–280.
13. Blake CCF, Johnson LN, Mair GA, North ACT, Phillips DC, Sarma VR. Structure of hen-egg white lysozyme: a three dimensional Fourier synthesis at 2.0Å resolution. Nature 1965;206:757–761.
14. Yang JJ, van den Berg B, Pitkeathly M, Smith LJ, Bolin KA, Keiderling TA, Redfield C, Dobson CM, Radford SE. Native-like secondary structure in a peptide from the alpha-domain of hen lysozyme. Fold Design 1996;1:473–484.
15. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag AG an der ETH Zürich; 1996. p 1–1024.
16. Ramanadham M, Sieker LC, Jensen LH. SRLSQ refinement of triclinic lysozyme. Acta Crystallogr A 1987;43(Suppl):13.
17. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular forces. Dordrecht: Reidel; 1981. p 331–342.
18. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
19. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. J Comput Phys 1977;23:327–341.
20. van Gunsteren WF, Berendsen HJC. Computer simulation of molecular dynamics—methodology, applications, and perspectives in chemistry. Angew Chem Int Ed Eng 1990;29:992–1023.
21. Smith PE, van Gunsteren WF. Consistent dielectric properties of the simple point charge and extended point charge water models at 277 and 300K. J Chem Phys 1994;100:3169–3174.