Molecular Simulation of Multistate Peptide Dynamics: A Comparison Between Microsecond Timescale Sampling and Multiple Shorter Trajectories

LUCA MONTICELLI,[1,2*] ERIC J. SORIN,[3¶] D. PETER TIELEMAN,[1] VIJAY S. PANDE,[3,4] GIORGIO COLOMBO[2,5]
[1]*Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada*
[2]*Centre for Biomolecular Interdisciplinary Studies and Industrial Applications, University of Milan, 20131 Milan, Italy*
[3]*Department of Chemistry, Stanford University, Stanford, California 94305*
[4]*Department of Structural Biology, Stanford University, Stanford, California 94305*
[5]*Istituto di Chimica del Riconoscimento Molecolare, CNR via Mario Bianco, 9, 20131 Milan, Italy*

Received 23 January 2006; Revised 15 November 2007; Accepted 2 January 2008
DOI 10.1002/jcc.20935
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: Molecular dynamics simulations of the RN24 peptide, which includes a diverse set of structurally heterogeneous states, are carried out in explicit solvent. Two approaches are employed and compared directly under identical simulation conditions. Specifically, we examine sampling by two individual long trajectories (microsecond timescale) and many shorter (MS) uncoupled trajectories. Statistical analysis of the structural properties indicates a qualitative agreement between these approaches. Microsecond timescale sampling gives large uncertainties on most structural metrics, while the shorter timescale of MS simulations results in slight structural memory for beta-structure starting states. Additionally, MS sampling detects numerous transitions on a relatively short timescale that are not observed in microsecond sampling, while long simulations allow for detection of a few transitions on significantly longer timescales. A correlation between the complex free energy landscape and the kinetics of the equilibrium is highlighted by principal component analysis on both simulation sets. This report highlights the increased precision of the MS approach when studying the kinetics of complex conformational change, while revealing the complementary insight and qualitative agreement offered by far fewer individual simulations on significantly longer timescales.

 J Comput Chem 00: 000–000, 2008

Key words: distributed computing; protein folding; RN24; helix-coil transition; GROMOS

Introduction

Until recently, the difference between the timescales of events observed experimentally or theoretically prevented the results of computer simulations from being directly comparable with experiments. The advances in simulation algorithms and computer power on the one hand, and the increase in time-resolution of many experimental approaches on the other hand, now allow for simulations of peptide folding on a timescale comparable to that of experiments.[1–5] The folding of secondary structure elements in real proteins takes place on the nanosecond to microsecond timescale.[6,7] Although the simulation of the folding of secondary structure elements is now within reach for molecular simulations with atomistic detail, the formation of protein tertiary structure in longer polypeptide chains remains a challenge. Because of the stochastic nature of folding, even when protein folding takes place on a timescale compatible with molecular simulations, one should simulate hundreds of folding events in order to understand which folding paths (mechanisms) are statistically most probable.

Several techniques have been proposed to bridge the timescale gap employing all-atom simulations including loosely coupled parallel simulations, such as replica exchange molecular dynamics,[8] Markovian model approaches,[9–15] independent parallel trajectories, and ensemble dynamics.[5] In the latter approach, based on the realization that overcoming an energy barrier is a stochastic process characterized by a distribution of crossing times, many short (MS) simulations are performed.[16,17] This approach has been successfully applied to study the folding of peptides,[3,5,16] small proteins,[18,19] and small RNAs[20,21] using a variety of solvent models.

*Present address: Laboratory of Physics, Helsinki University of Technology, Helsinki, Finland.
¶Present address: Department of Chemistry and Biochemistry, California State University, Long Beach, California.
Correspondence to: L. Monticelli; e-mail: luca.monticelli@ucalgary.ca

A complication arises if one considers that escaping from a local minimum in a multidimensional space may involve different transition states, corresponding to different transition times. Moreover, the spontaneous search of the energy minimum may involve escaping a succession of local energy minima that necessarily lead to lag phases in the folding kinetics. It has been detailed by Fersht that using many short independent simulations is perfectly valid once the timescale of each individual simulation is longer than the lag phases of the folding process.[22] Paci et al. studied the folding of a small protein using a simple model implicit solvent model and found for that system, the folding rate is estimated correctly using the MS approach when trajectories longer that 1/100 of the average folding time are considered, and the sequences of events is not necessarily representative of the folding pathways when short trajectories are considered.[23] Moreover, using simulations that are too short compared with the equilibration time of the system would not only yield incorrect kinetics, but also incorrect thermodynamics for the system, as the sampled structures in a short simulation will depend on the starting conformation more strongly than in a long simulation, if a simple two state assumption is made (this is especially valid for systems that have several states with comparable stability). However, a Markov State Model analysis for this data was not made and would likely have resolved these issues.

How short can independent MS simulations be while still providing correct thermodynamic and kinetic results? Does this depend on the length of the peptide sequence only, or also on other factors? And how do results from longer simulations compare with results from many shorter ones? These questions prompted us to investigate the folding of a small natural peptide using both the MS approach as well as by sampling few long (FL) trajectories, in both cases using an explicit solvent model. We note that distributed computing networks can provide long trajectories as well as short ones (indeed, the long trajectories presented here were computed with desktop computers), and emphasize that the comparison being made here is not one of "local sampling versus distributed computing," but rather one of few long trajectories versus many shorter simulations, both of which can be sampled on various architectures including distributed computing networks, localized supercomputers, and large processing clusters. We note that in our analysis of MS trajectories here, we are doing a naive analysis (direct analysis of the MS trajectories and employing a two-state model), rather than some more sophisticated methods which have recently been used, such as Markov State Models (MSM); an MSM approach's goal is to explicit yield converged results with rather short trajectories[9,14,15], but requires a more complex analysis than what is performed here.

The peptide chosen for this study was RN24, an analogue of the C-peptide, constituted by the 13 amino-terminal residues of ribonuclease A (sequence Ac-AETAAAKFLRAHA-NH_2). The peptide sequence is of considerable interest because, despite being very short, it appears to have complex dynamics with multiple states and nonexponential folding kinetics. Indeed, RN24 was initially found to form a helix in water solution at low temperatures,[24] but more recent NMR experiments in similar conditions revealed that at least three kinds of conformations are adopted: α-helical, partially extended, and bent conformations.[25] Because of the very short size, this peptide can be simulated for a relatively long time and is therefore an ideal test case to probe the questions formulated above.

Methods

FL Sampling

An ideal α-helical structure and a random coil structure were generated using the Macromodel software package and used as starting structures for our simulations. The elongated random coil structure with minimal self-interactions was chosen to avoid the introduction of any biases toward any folded structure. The histidine ring was protonated only at one nitrogen atom, so that the total charge of the peptide was +1 (due to the presence of a glutamate, a lysine, and an arginine residue). The peptide was placed in a dodecahedral box large enough to contain 1.2 nm of solvent around the peptide in helical conformation (2782 water molecules) and 0.6 nm of solvent for the random coil conformation (2828 water molecules). XX, YY, and ZZ components of the dodecahedral box are on average about 4.9, 4.9, and 3.5 nm, yielding a volume of ~84 nm^3. The simple point charge (SPC) water model was used and one chloride ion was added to obtain electroneutrality. All simulations were carried out in version 3.1.4 of the GROMACS package[26] using the GROMOS96 43A1 force field.[27] All bond lengths were constrained to their equilibrium values using the SETTLE algorithm[28] for water and the LINCS algorithm[29] for the peptide. Dummy atom constructions and heavy hydrogen atoms were used (4 a.u.) according to a published procedure,[30] allowing for an integration time step of 5 fs. The neighbor list for the calculation of nonbonded interactions was updated every four timesteps. In all simulations, periodic boundary conditions were used and the calculation of electrostatic forces utilized the Particle Mesh Ewald (PME) method[31,32]; 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 B-spline interpolation. The relative error for the Ewald sum in the direct and reciprocal space, controlled in GROMACS by the parameter ewald_rtol, was set to 10^{-5}. A twin-range cutoff of 0.9–1.4 nm was used for the calculation of Lennard-Jones interactions. In each simulation the peptide and solvent were coupled separately to a temperature bath at 300 K, using the Berendsen algorithm with $\tau_T = 0.1$ ps,[33] and the pressure was kept at 1 bar using weak pressure coupling with $\tau_P = 4.0$ ps.[33]

The systems were energy minimized with a steepest descent method for 5000 steps. The solvent was then equilibrated in a 50 ps MD run with position restraints on the peptide. 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 of 300 K. Two simulations were then carried out for 1.6 μs each and will be referred to herein as the few long (FL) trajectories, with FL-ran indicating the random coil elongated starting structure and FL-hel the helical starting structure. Peptide structures were saved for analysis every 1 ps.

MS Sampling

Four sets of many shorter (MS) simulations were carried out using the Folding@Home distributed computing network (http://folding.stanford.edu) with four different starting structures; the same random coil and helical starting structures from the FL simulations were used, together with β-hairpin and parallel β-sheet structures sampled during the FL simulations. These simulations are referred to herein as MS-ran, MS-hel, MS-hp, and MS-sh. A summary of the simulations performed is reported in Table 1. The same force field and identical simulation parameters were used in all MS and FL simulations (see earlier). Each of the four sets of MS simulations consisted of ~800 trajectories of ~250 ns each, for a total simulation time of about 800 μs, with structures saved every 100 ps.

Analysis

For most structural metrics, uncertainties were estimated using the procedure described by Hess[34] as implemented in GROMACS. Uncertainties obtained with this procedure are substantially equivalent to the standard error. For cluster populations, the standard error is given as an estimate for uncertainties. Differences between data sets were estimated using the "Analysis of variance" (ANOVA) technique. ANOVA is a statistical method that allows to compare the average values of one variable in three or more data sets, and evaluate the significance of the differences. One-way ANOVA is useful when only one independent variable is present in the data set. The "null hypothesis" is that the average (of the dependent variable) is the same for all groups, i.e., the sets sample from the same population. ANOVA is used to calculate the probability (p-value) of the null hypothesis being true. If the p-value resulting from the ANOVA procedure is low (usually less than 0.05, by convention), then the difference between the averages is considered statistically significant, and the conclusion is that the data sets are not equivalent. When using block averages as data points for ANOVA, the analysis is sensitive to the size of the block. To simplify the analysis, standard deviations for all structural metrics were calculated using block averaging over blocks of the same size. The block size was chosen to be 1/4 of the length of each individual FL trajectory; this is much larger than the correlation time for all structural metrics. We used a MacOS X version of the statistical package R (www.r-project.org) for ANOVA.

Interproton distances were calculated from all simulations as $\langle d(t)^{-6}\rangle^{-1/6}$ averages, and violations were derived as $d_{viol} = \langle d(t)^{-6}\rangle^{-1/6} - d_{exp}$, where d_{exp} is the distance measured experimentally. Since aliphatic hydrogen atoms in the GROMOS96 43A1 force field are treated using a united-atom model, their position at each time in the trajectory was back-calculated assuming an ideal geometry. Secondary structure assignments were based on the DSSP algorithm.[35] The number of helical (N_h), beta (N_β), and turn (N_t) residues was calculated based on DSSP assignments. Unless specified otherwise, "beta" structure described herein refers to both β-hairpin and β-sheet conformations.

Table 1. Summary of the Structural Features and the Secondary Structure Elements of the Peptide Conformations Sampled in Each Simulation Set.

	FL-ran	FL-hel	MS-ran	MS-hel	MS-hp	MS-sh	FL (total)	MS (total)
Starting conformation	Elongated random coil	α-helical	Elongated random coil	α-helical	β-hairpin	Parallel β-sheet	–	–
Number of simulations	1	1	~800	~800	~800	~800	2	~3200
Total simulation time (μs)	1.6	1.6	~200	~200	~200	~200	3.2	~800
Gathering simulation time (μs)	1.375	1.375	5.5	5.5	5.5	5.5	2.75	22
<Rg> (nm)	0.672 ± 0.007	0.670 ± 0.07	0.659 ± 0.001	0.666 ± 0.001	0.664 ± 0.001	0.652 ± 0.001	0.672 ± 0.007	0.660 ± 0.001
Conformations with Rg > 0.8 nm	0.051%	0.034%	0.030%	0.028%	0.023%	0.018%	0.042%	0.025%
Average number of helical residues	1.15 ± 0.74	0.82 ± 0.37	0.92 ± 0.03	1.11 ± 0.03	0.80 ± 0.03	0.70 ± 0.03	0.99 ± 0.42	0.89 ± 0.15
Average number of beta residues	1.29 ± 0.28	1.61 ± 0.32	1.51 ± 0.05	1.46 ± 0.05	2.06 ± 0.07	1.78 ± 0.05	1.45 ± 0.30	1.70 ± 0.03
Average number of turn residues	1.41 ± 0.11	1.29 ± 0.21	1.39 ± 0.04	1.33 ± 0.03	1.41 ± 0.04	1.22 ± 0.02	1.34 ± 0.15	1.33 ± 0.03

Results are given as average ± estimated error (see Methods). The radius of gyration was calculated on all atoms. Both the radius of gyration and the secondary structure content were calculated excluding the initial 225 ns from each simulation. Error estimates are calculated using a block averaging procedure described by Hess.[29]

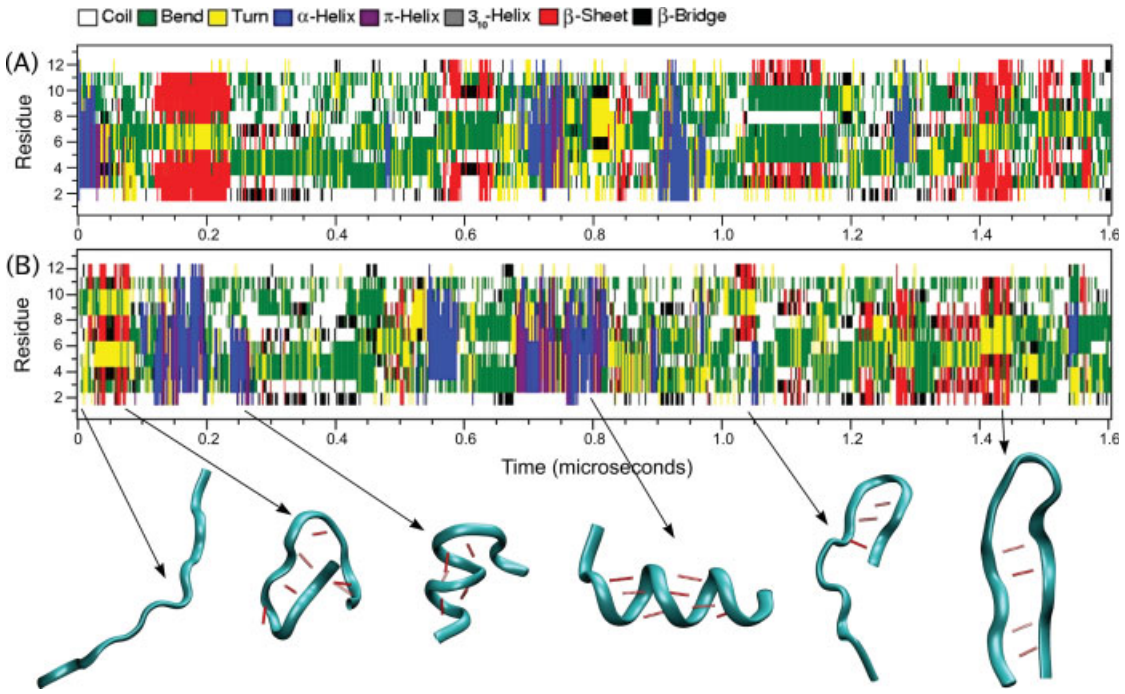

Figure 1. Secondary structure as a function of simulation time in the long trajectory starting from an ideal α-helical (A) and from a random coil structure (B). Some examples of the conformations sampled are shown as ribbons, with red dotted lines representing hydrogen bonds.

Results

Comparison of MS and FL sampling

In all simulations the peptide explores many kinds of secondary structures, ranging from ideal α-helix to *β*-hairpin. Figure 1 shows the secondary structure as a function of simulation time in the FL simulations. In Figure 2, the ensemble average number of helical, beta, and turn residues is plotted as a function of simulation time in each of the four MS sets. The metrics do not converge completely, yet the four ensembles show similar secondary structure content after ∼225 ns of simulation. For this reason, all the analyses described below were carried out only on the ensembles of structures collected after 225 ns, both in the MS and in the FL simulations. This subset contains 22 *μ*s for the MS set and 2.75 *μ*s for the FL set. The average population of helical, turn, and beta secondary structures is qualitatively similar in the FL and MS sets (Table 1).

Estimates of the error on the number of residues with a certain secondary structure were calculated by block averaging (using a procedure described by Hess[34]). The uncertainties on secondary structure metrics are much larger for the FL than for the MS sets. It is interesting to notice that FL trajectories starting from the helical conformation have the lowest average helical content; in contrast, the MS simulations starting from the helical conformation have the highest mean helical content and MS simulations starting from the *β*-hairpin conformation have the highest beta content. Although the number of turn residues shows good agreement both in FL and MS sets, FL simulations appear to predict lower number of beta residues and higher number of helical residues. To determine the extent to which the differences in secondary structure in the FL and MS sets were statistically meaningful, we used the analysis of variance (ANOVA) technique. MS data sets were split into four subsets, so that each subset had the same size as each of the FL sets. Then each MS subset and each FL set was divided into four blocks; the block size (343.75 ns) was larger than the portions of the simulations discarded as "equilibration" and larger than the correlation time for all structural metrics. When we consider the number of helical residues, one-way ANOVA shows that the two FL sets are statistically equivalent (*p*-value: 0.66) and that the differences between the FL and the combined MS data sets are not statistically significant (*p*-value: 0.72). The MS subsets generated from the same starting structure are also always statistically equivalent (*p*-values $>$ 0.66 in all four cases), but when all MS sets with different starting structures are considered the *p*-value is $\sim 10^{-6}$, indicating that the sets are not equivalent. Interestingly, differences are found to be not statistically significant if we consider the combined MS-ran and MS-hel sets, as well as the combined MS-hp and MS-sh sets; both combinations are also found to be equivalent to the combined FL sets. These apparently conflicting results are explained by the large uncertainties in the FL sets. Because the variance in FL sets is always larger than in MS subsets of the same size, we conclude that this is due to the sampling technique. Analysis of variance on the number of beta residues gives very similar results, whereas analysis on the number of turn residues fails to show significant differences, suggest-

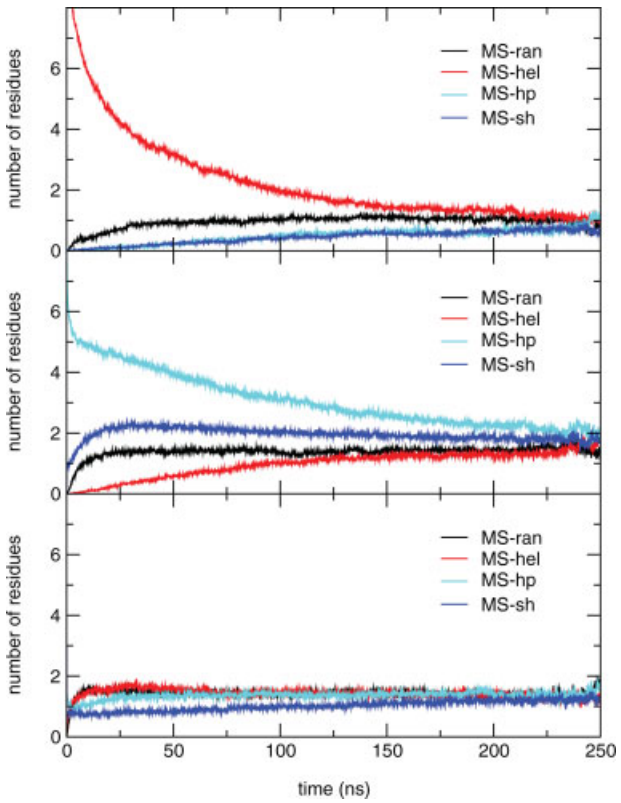

Figure 2. Ensemble average of the population of helix (top panel), beta (central panel), and turn (bottom panel) secondary structure as a function of time in the four MS simulation sets. Secondary structure was calculated using the DSSP algorithm. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ing that this metric is not capable of discriminating among different sampling methods.

Figure 3 shows the secondary structure preference as a function of residue number in the FL simulations and in the four MS sets. Convergence at the residue level was assessed within each FL simulation by splitting them in 2 halves (corresponding to 0.6875 μs each), and within the MS set by comparing results of each set (corresponding to 5.5 μs each). Significant differences in the distributions are evident in FL sets, whereas distributions are very similar in the MS simulations. In both FL and MS sets, residues Ala4 to Lys7 appear to have the highest preference for the helical conformation, β-turns are usually centered at residues Ala5-Ala6 or Ala6-Lys7; residues 2–4 and 9–11 have highest preference for the beta conformation.

The average radius of gyration (R_g) is 0.67 nm and 0.66 nm in the FL and MS simulations, respectively, much lower than the value for the completely elongated conformation (1.25 nm) and slightly lower than the value for the ideal α-helix (0.70 nm). Conformations with a radius of gyration larger than 0.8 nm are less than 0.1% in both the FL and the MS simulations, with slightly lower values for the MS ensemble (see Table 1), indicating that elongated conformations have very low population in the unfolded ensemble, which is mostly composed of very compact coil structures. Figure 4A shows that distributions of the radius of gyration in the six simulation sets are reasonably similar. Only the MS simulation starting from the β-sheet structure shows a significant difference compared to all others: a peak around 0.61 nm is present in the distribution, corresponding to the radius of gyration of the parallel β-sheet conformation used as a starting structure. This suggests that the population of β-sheet-like conformations is higher than in the other simulation sets.

The root-mean-squared deviation (RMSD) relative to the ideal α-helix was calculated for each structure sampled. Figure 4B shows the distribution of RMSD in the six simulation sets. Results for the MS simulation sets starting from a hairpin or a β-sheet conformations show average RMSD relative to the ideal α-helix higher than the other sets, and therefore appear to sample different conformations. In particular, these distributions show peaks around 0.56 and 0.50 nm, respectively, corresponding to the RMSD of the starting structures. The secondary structure was calculated for all the conformations accounting for the peaks, and a high content in beta structures was found, confirming that a memory of the starting structures is present in the MS sets having β-hairpin and β-sheet starting structures.

One-way ANOVA on both the radius of gyration and the RMSD confirmed that the differences in the two FL sets are not statistically meaningful, whereas differences among MS sets with different starting structures are significant (p-value $< 10^{-6}$). Again, FL data sets are always statistically equivalent to MS data sets, both when the starting structure is the same and when it is different.

Assignment of states and kinetics of the equilibrium

Because the conformational equilibrium sampled in both the FL and MS simulation sets includes a broad range of conformations, potentially the best way to analyze the kinetics of the conformational equilibrium is to assign each of the structures sampled to a specific conformational state (which corresponds to a related ensemble of thermodynamic microstates), and then calculate the probability of the transitions between all possible states. Assignment of conformations to states was performed by clustering; the numbers of residues in α-helical, 3_{10}-helical, π-helical, and beta conformations were used as criterion for clustering. A total of 220,016 structures were included in the clustering for the MS simulation sets and 27,502 for the FL sets, representing 22 μs and 2.75 μs of equilibrium sampling with 100-ps time resolution. The MS data set was clustered as described previously using a modified K_{means} algorithm[10] with 25 random starting cluster centers and requiring consistent cluster assignments of all data in the set for 10 consecutive iterations prior to convergence. Because the clustering here is used to compare sampling methods rather than determining absolute character of states in the data, artifacts of the clustering will be unimportant.

Eleven clusters were identified, and the mean secondary structure content of the each one of them is shown in Table 2, alongside the populations of the clusters in the MS and FL data sets. Apparently, the agreement between FL and MS is only qualitative: the maximum absolute difference in the population of the states is $\sim$3%, while relative differences show much larger values, up to 123%, corresponding to differences in free

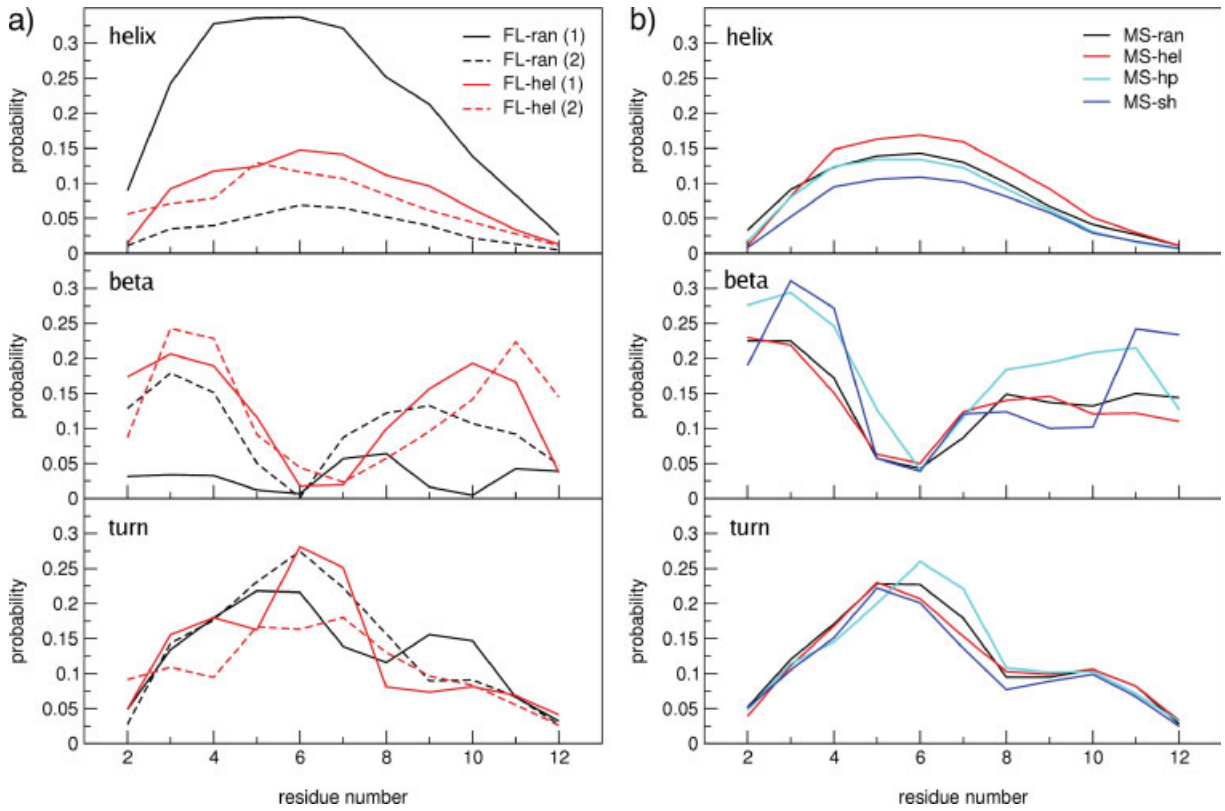

Figure 3. Population of helical, extended, and turn conformations as a function of residue number in the FL simulations (A) and MS simulations (B). For the FL set, each simulation was split in 2 halves, and the distributions for each half are compared; numbers in parentheses indicate the first half (1) and the second half (2) of the simulations, respectively. Only the portions of trajectories after 225 ns were taken into account. Each FL subset corresponds to 0.6875 μs, whereas each MS subset corresponds to 5.5 μs. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

energy between the two sets as large as ~5 kJ/mol. To evaluate the agreement between the two sampling methods from a statistical point of view, we split both the FL and the MS data sets into subsets of the same size, and determined the probability of the subsets being statistically equivalent using one-way ANOVA. Analysis of variance shows that differences between the two FL sets are not statistically significant, whereas differences between MS and FL sets are significant for clusters 1 and 3. Differences among the four MS sets are also significant for the same clusters (p-value < 0.001 in both cases). To find the source of the differences in the MS set, we performed ANOVA excluding one MS subset and then two MS subsets from the analysis. Results show that MS-ran and MS-hel sets are always equivalent to each other and equivalent to FL-ran and FL-hel sets, whereas the MS-hp and MS-sh sets always show statistically significant differences with all other sets.

Based on the clusters (states) listed in Table 2, the MS and FL data sets were parsed for transitions and the relative rates were calculated, based on 100 ps intervals. Figure 5 shows representative structures for each of the 11 conformational states found, as well as the transition rates between them. If we consider the whole MS data set (which is 8 times larger than the FL data set), the transition rates observed in the MS and FL simulation sets are qualitatively similar only for transitions to and from some of the most populated clusters (gray cells in Fig. 5); for less populated clusters, FL sampling shows lower rates more often than higher rates, indicating that the number of transitions missed by FL but observed in MS sampling is greater.

Out of 121 possible transitions, 18 are not observed in both sets, 28 are not observed in the FL set, and only 2 are not observed in the MS set. Only 42 transitions show higher rates in FL simulations, whereas the others show higher rates using the MS sampling. The largest deviations between the two methods are found for transition to and from clusters 8 to 11, which have population lower than 1%. Focusing on the seven most populated clusters, large differences in the average transition times are registered for the transition from cluster 6 (long α-helix) to clusters 1 and 2 (β-bridged and random coil), from cluster 3 (β-hairpin) to cluster 5 (short π-helix), from clusters 1 and 2 to cluster 6; in 4 out of 5 of these instances, the difference is due to insufficient sampling in FL simulations, while only in the case of the transition from cluster 6 to 1 (from long α-helix to β-bridged) sampling appears to be better in the FL set. We attribute this to the slow rate of transition here, relative to the short MS simulation timescale, which highlights the complimentary nature of FL-type simulations compared to MS simulations.

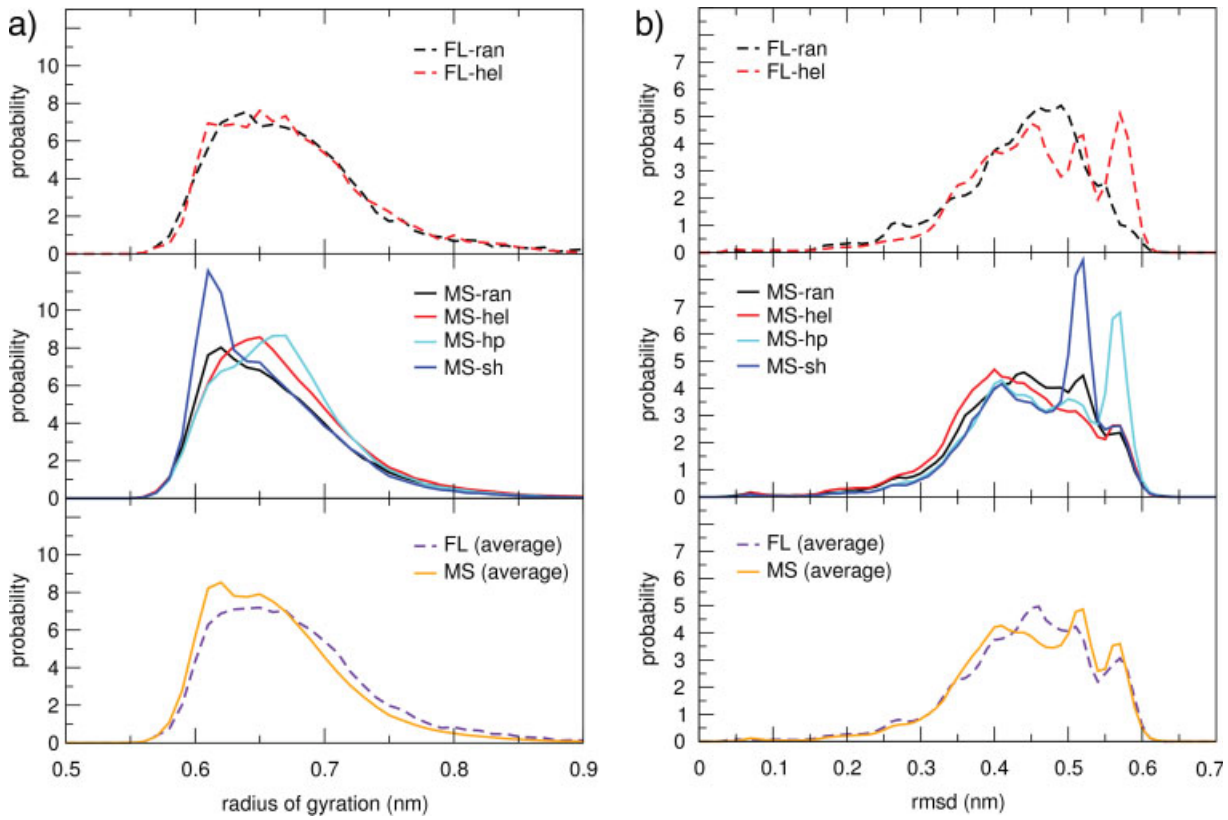

Figure 4. (A) Normalized distribution of the radius of gyration in the six simulation sets. Only the portions of trajectories after 225 ns were taken into account. (B) Normalized distribution of the root-mean-squared deviation (RMSD) of the structures from the ideal α-helical conformation in the six simulation sets. Only the portions of trajectories after 225 ns were taken into account. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

These results also highlight the significance of using Markov State Model approaches, especially in which adaptive sample methods[13] are employed.

Transition rates were also calculated for subsets of the MS and FL data sets of the same size (16 MS and 2 FL subsets, each one totaling 1.375 μs). Again, we compared the number of

Table 2. Clustering of Conformational States in MS and FL Simulations.

#	Cluster name	$\langle N_\alpha \rangle$	$\langle N_{310} \rangle$	$\langle N_\pi \rangle$	$\langle N_\beta \rangle$	%MS	%FL	Δ% (relative)	ΔG (kJ/mol)
1	Beta bridged	0.00	0.01	0.00	2.67	38.48 ± 2.13	36.63 ± 4.87	5.27	−0.135
2	Random coil	0.00	0.00	0.00	0.00	36.37 ± 2.11	39.31 ± 4.36	−7.47	+0.195
3	Beta-hairpin	0.02	0.01	0.00	6.68	10.04 ± 1.56	7.40 ± 4.19	35.67	−0.761
4	Short Alpha	4.92	0.00	0.00	0.09	8.36 ± 1.20	7.97 ± 2.83	4.59	−0.116
5	Short Pi	0.00	0.03	5.75	0.00	2.27 ± 0.52	2.56 ± 2.16	−13.71	+0.321
6	Long Alpha	7.79	0.00	0.00	0.00	2.06 ± 0.51	2.98 ± 1.54	−37.88	+0.802
7	310 Nucleus	0.00	3.04	0.00	0.36	1.15 ± 0.13	1.20 ± 0.16	−2.77	+0.068
8	Long Pi	0.00	0.00	8.83	0.00	0.99 ± 0.32	1.37 ± 1.21	−47.63	+0.972
9	Alpha/Pi	4.30	0.00	5.21	0.00	0.23 ± 0.09	0.49 ± 0.36	−123.05	+2.003
10	Alpha/310	4.84	3.00	0.00	0.03	0.06 ± 0.02	0.06 ± 0.04	−8.42	+0.201
11	Pi/Beta	0.00	0.00	5.03	2.32	0.04 ± 0.02	0.01 ± 0.01	86.72	−5.041

$\langle N_\alpha \rangle$ indicates the average number of α-helical residues in the cluster, $\langle N_{310} \rangle$ the average number of 3_{10}-helical residues, $\langle N_\pi \rangle$ the average number of π-helical residues, and $\langle N_\beta \rangle$ the average number of beta residues. % MS and % FL indicate the population of the cluster in the MS and FL simulations, respectively, given as averages ± standard error (calculated by block averaging on blocks of the same size for FL and MS sets).

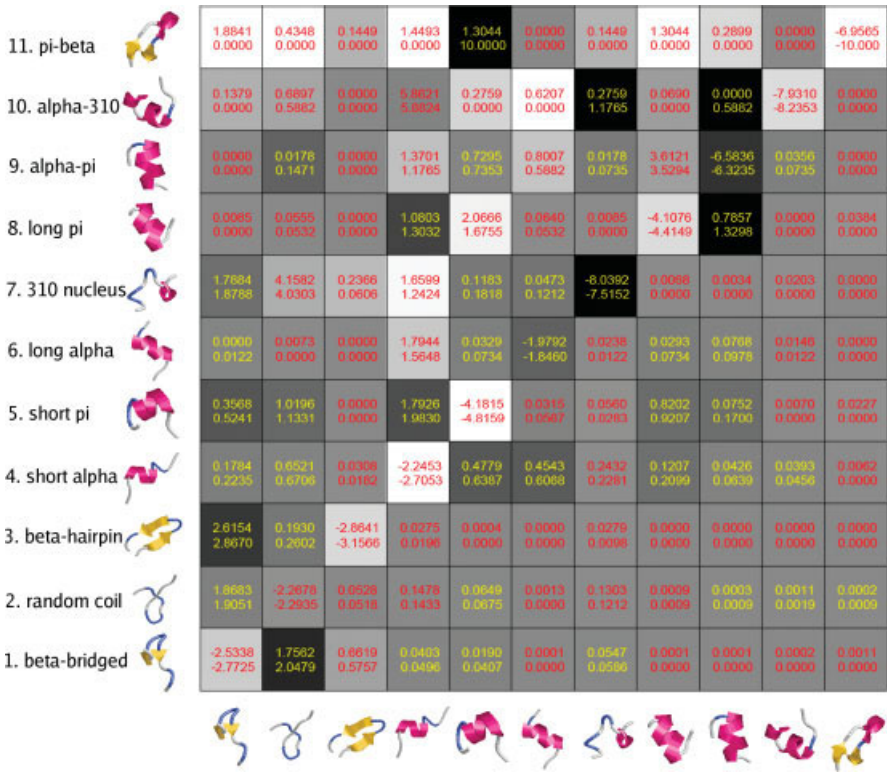

Figure 5. Microstate transition rate matrix for FL sampling (top numbers) versus MS sampling (bottom numbers) in inverse nanoseconds. Background colors scale from black (FL sampling overestimating the rate) to gray (similar rates) to white (transitions unobserved in FL sampling); rates are in yellow when FL overestimates the rate (darker gray squares), in red when FL underestimates the rate (lighter gray squares); the relative difference between FL and MS rates is used for the shading scale. A representative structure for each cluster is shown, alongside the ranking of the cluster (based on population). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

transitions missed in each subset and the differences in the transition rates. We found that, on average, ~12 transitions are missed in MS subsets but observed in FL subsets, and ~10 transitions are missed in FL subsets but observed in MS subsets. As for transition rates, we calculated the relative deviation of each rate from the average rate for the same transition (calculated for each transition taking into account all MS and FL data sets), and summed up all the relative deviations for each subset considered; this sum provides a quantitative measure of the deviation of the rates from the average for each subset.

Results show that the relative deviations from the average rate are comparable for MS and FL simulations. We then split each rate matrix into 4 matrices, containing transition rates for the conversions between clusters (a) 1–6 and 1–6, (b) 1–6 and 7–11, (c) 7–11 and 1–6, (d) 7–11 and 7– 11. Relative deviations from the average rate were similar for FL and MS data sets when we consider transitions involving the most populated clusters, whereas much larger deviations were found in the FL set when considering transitions among the least populated clusters (cluster 7 to 11). We also divided the transitions in 3 groups based on their rate: fast, intermediate, and slow transitions, and checked the relative deviations of the rates in each group. Consistently with the findings above, we found that larger deviations are observed with MS sampling for slow transitions and with FL sampling for fast transitions. These results show that, considering data sets of the same size, the MS methodology provides more reliable estimates of the transition rates between conformations with lower populations and with faster transition rates, while FL sampling has some advantage when transition rates are slow.

Free energy landscapes

The potential of mean force (PMF) was calculated for each simulation set as:

$$\Delta G_x = -RT \ln(N_x/N_{tot})$$

where T is the simulation temperature, N_X indicates the population of the particular state X and N_{tot} the total population sampled. Principal components, backbone dihedral angles, and secondary structure were used to define the state in different PMF plots. PMF surfaces can be considered as approximations to the equilibrium free energy landscapes projected onto these metrics.

The plot of the PMF as a function of the first two principal components (Fig. 6) shows a relatively smooth energy land-

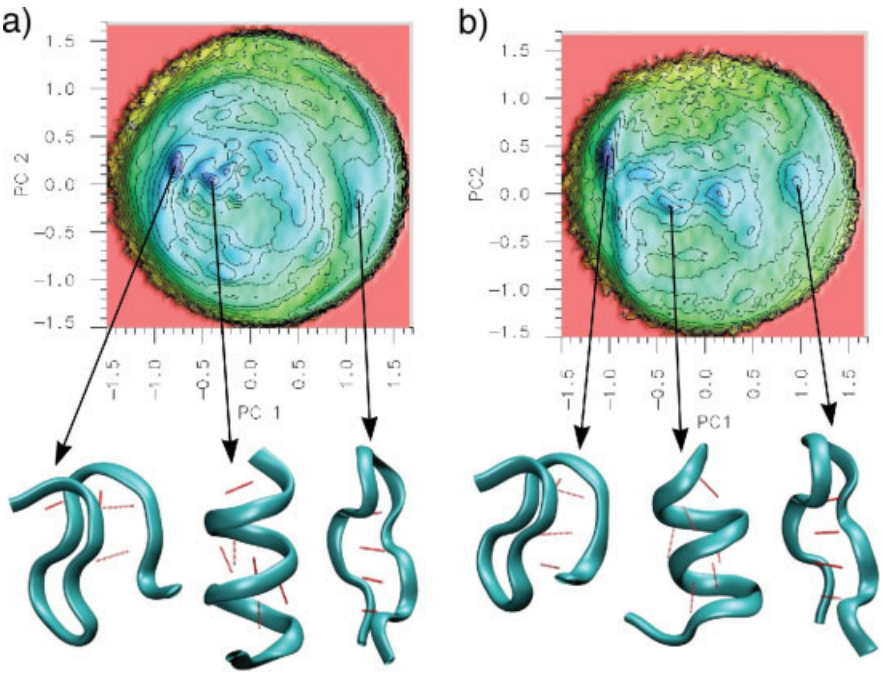

Figure 6. Potential of mean force for the FL and MS simulations as a function of the first two principal components. Only the portions of trajectories after 225 ns were taken into account. Contour lines are drawn with 0.5 kT spacing. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

scape, with numerous minima separated by low energy barriers, as has recently been observed for larger helix-coil systems[10]. PMFs for both data sets were constructed using the same principal components, from a data set including all FL and MS simulations. Both for the FL sets and for the MS sets, the lowest 25 minima are within 2 kT from the absolute minimum. Structures found within 1 kJ/mol from each minimum were clustered using the backbone RMSD as a similarity criterion, with a cutoff of 0.15 nm. Each minimum contains only one main cluster (with a population of 75% or higher), and the central structure of the most populated cluster was chosen as a representative structure. Location of the lowest energy minima is shown in Figure 6, together with the central structure of the cluster. Both the location and the structures found in the PMF plots for the FL and MS sets are similar. Cluster centroids show well-defined secondary structure: a parallel β-sheet and an α-helix are found in the two lowest energy minima, whereas an anti-parallel β-hairpin structure is found in a more shallow but much broader minimum. The barrier separating the β-hairpin conformation from all other minima is the only one larger than 1.5 kT in both PMF plots. The lowest energy minimum, represented by the parallel β-sheet conformation, also shows a barrier around 1.5 kT. Because the conversion between β-hairpin (or β-sheet) and other structures takes place on a longer timescale compared to most other conformational transitions in the system, we conclude that the PMF as a function of the first two principal components highlights a correlation between free energy landscape and kinetics.

The plot of the PMF as a function of φ and ψ backbone dihedral angles is shown in Figure 7. The most populated regions in the φ/ψ space are virtually identical for the FL and MS simulations. The presence of a relative minimum around $\varphi = +60$ and $\psi = -60$ highlights the presence of tight turns (γ-turns), which are not evidenced by DSSP analysis. The difference plot (Fig. 7C) shows that differences between the FL and MS sets are mainly localized in the less populated regions (high free energy in Figs. 7A and 7B). A PMF was also calculated as a function of the number of helical and beta residues; results (not shown) confirm that the absolute minimum in both FL and MS simulation sets corresponds to the absence of secondary structure elements, and relative minima are found for conformations with both beta and helical residues.

Comparison between simulations and experiments

Because most of the experimental data available on the RN24 peptide were collected in conditions different from those of the simulations (low temperature, low pH, and high ionic strength), we do not know if the simulation parameters (force field and methodology) would provide a reasonable match with experimental values. Moreover, the lack of a well-defined helical secondary structure found in our simulations suggests that the force field employed here may not be predictive of experimental results. Therefore, a comparison between simulations and experimental results should be considered meaningful only at a qualitative level. However, even if the results are not predictive of experiment, one can still ask how these different methods predict the behavior of this complex dynamical system—as we compare simulation methods with each other, at some level, the

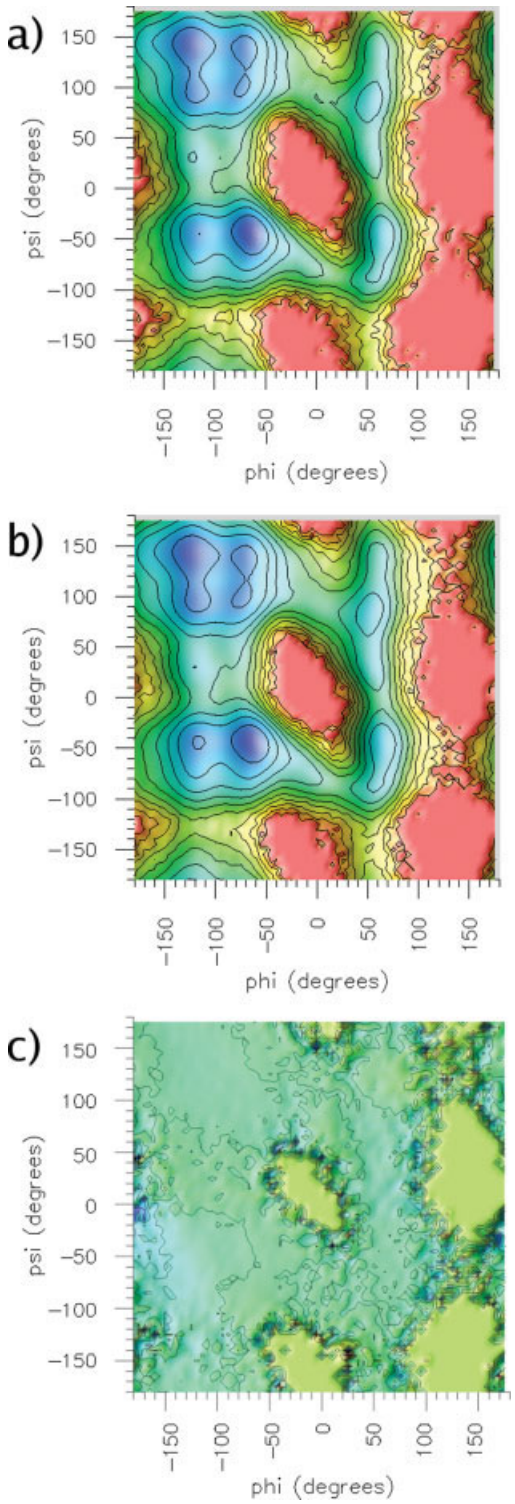

Figure 7. Potential of mean force as a function of the backbone φ and ψ dihedral angles for (A) the FL simulations, (B) the MS simulation, and (C) difference between the two. Only the portions of trajectories after 225 ns were taken into account. Contour lines are drawn with 1 kT spacing in all plots.

agreement with experiment is not necessary to allow a comparison of computational methods. Finally, we see fairly complex dynamics for a peptide, including both helical and hairpin structures, which yields a natural challenge for any computational method to describe.

A helical content of about 50% was measured by circular dichroism at 3°C and pH 5 for the RN24 peptide[24] and it was shown that this helical content decreases sharply as temperature increases.[36] Related peptides, like the C-peptide and the S-peptide, also show a significant helical content at low temperature, but the helix is not observed above 30°C.[37] NMR experiments were also carried out on the RN24 peptide at the temperature of 275 K, at low pH and high ionic strength,[25] and showed nonhelical long-range NOE peaks, suggesting multiple conformations in aqueous solution that include both helical and bent structures. As shown in Table 1, the helical content in both the FL and the MS simulations reaches ~10%, which is compatible with the experiments at higher temperature. Besides the difference in temperature, other factors might contribute to the low helical content in our simulations, such as the difference in pH and ionic strength, inadequacies inherent to the force field employed, and inadequate simulation methodology. Force field inadequacies probably play a role in determining the high population of π-helical conformations observed, which accounts for ~1/3 of the total helical content in all simulations. The high content of π-helix was previously shown by Brooks et al. to be a force field artifact,[38] probably due to poor parameterization of backbone dihedral angles.

Average distances calculated from the trajectories are compared with published NOE-derived inter-proton distances. All *(i, i+2)* and longer-range connectivities reported by Osterhout et al.[25] were compared to distances calculated from the simulations, for a total of 20 inter-proton distances. Table 3 shows the violations for each of these distances. Because of the lower helical content and to higher flexibility of the peptide at higher temperature, one could expect large discrepancies between calculated and measured distances. In fact, differences are not very large, due to the highly non-linear weighing of distances in the NOE experiment. Despite the relatively low sensitivity of the NOE experiment to minor changes in a conformational equilibrium, the comparison between calculated and measured distances shows differences between the simulation sets. Even though 3 and 5 violations are observed in simulation FL-ran and FL-hel respectively, no distance is violated for more than 0.05 nm. In the case of the MS simulations, no violations larger than 0.05 nm are found in the sets starting from a helical and a random coil conformation, whereas 3 and 2 violations are found when the starting structure was a β-hairpin and a β-sheet, respectively. NOE violations in 2 MS sets are due to the persistence of hairpin-like or β-sheet-like structures, as evidenced by secondary structure analysis. Averaging over all four MS simulation sets gives no violations larger than 0.05 nm, in agreement with NMR data; this is consistent with the highly non-linear weighing of distances in NOE experiments.

Discussion

Several theoretical studies have been carried out on RN24 and similar peptides.[39–42] In the present work, we investigate the re-

Table 3. Calculation of NOE-Derived Distances in the FL Simulations Separately (FL-hel, FL- ran) and Together (FL-avg), and in Each Set of MS Simulations Separately (MS-hel, MS-ran, MS- hp and MS-sh) and Together (MS-avg).

#	Atom1-Atom2	FL-hel	FL-ran	FL-avg	MS-hel	MS-ran	MS-hp	MS-sh	MS-avg
1	E2(Hα) - A5(HN)	0.031	0.013	0.022	0.018	0.050	0.054	0.078	0.048
2	E2(Hα) - A5(Hβ*)	−0.056	−0.057	−0.056	−0.054	−0.014	−0.018	0.018	−0.021
3	T3(Hα) - A6(Hβ*)	−0.065	−0.101	−0.086	−0.090	−0.093	−0.071	−0.057	−0.079
4	A4(Hα) - K7(HN)	0.000	−0.030	−0.017	−0.023	−0.041	−0.014	−0.009	−0.023
5	A4(Hα) - K7(Hβ*)	−0.081	−0.115	−0.101	−0.100	−0.131	−0.090	−0.088	−0.106
6	A5(Hα) - K7(HN)	−0.023	−0.031	−0.027	−0.052	−0.032	−0.035	−0.025	−0.036
7	A5(Hα) - F8(HN)	−0.016	−0.017	−0.016	−0.017	−0.033	−0.011	0.005	−0.016
8	A5(Hα) - F8(Hβ*)	−0.080	−0.083	−0.081	−0.074	−0.097	−0.058	−0.062	−0.075
9	K7(Hα) - L9(HN)	0.005	−0.009	−0.002	−0.007	−0.009	0.000	0.006	−0.002
10	K7(Hα) - R10(HN)	0.044	0.004	0.021	0.038	0.012	0.050	0.062	0.038
11	K7(Hα) - R10(Hβ*)	−0.022	−0.046	−0.035	−0.019	−0.035	−0.001	0.016	−0.013
12	F8(Hα) - R10(HN)	−0.032	−0.015	−0.024	−0.026	−0.026	−0.020	−0.048	−0.031
13	F8(Hα) - A11(HN)	0.034	0.032	0.033	0.047	0.033	0.065	0.039	0.044
14	F8(Hα) - A11(Hβ*)	−0.018	−0.020	−0.019	−0.008	−0.020	0.016	0.008	−0.003
15	L9(Hα) - H12(Hβ*)	0.021	−0.012	0.05	−0.007	−0.014	0.011	0.025	0.002
16	T3(HN) - A6(Hβ*)	−0.028	−0.068	−0.051	−0.108	−0.118	−0.091	−0.088	−0.103
17	T3(HB) - A5(HN)	−0.110	−0.096	−0.104	−0.119	−0.134	−0.120	−0.118	−0.123
18	T3(Hβ) - A6(HN)	−0.035	−0.032	−0.034	−0.057	−0.071	−0.051	−0.056	−0.060
19	E2(Hγ*) - K7(HN)	−0.012	−0.106	−0.075	−0.115	−0.091	−0.080	−0.076	−0.092
20	T3(HN) - A6(HN)	−0.015	0.000	−0.008	−0.024	−0.038	−0.023	−0.018	−0.026
Number of violations		5	3	4	3	3	6	9	4
Number of violations > 0.05 nm		0	0	0	0	0	3	2	0
Average violation (nm)		0.006	0.002	0.007	0.005	0.005	0.010	0.013	0.006

In all cases, only the portions of the trajectories beyond 225 ns were taken into account (corresponding to 1.375 μs for each FL simulation and 5.5 μs for each MS simulation).

versible folding of the RN24 peptide using molecular dynamics simulations, and we focus on the comparison between few long trajectories and thousands of shorter ones obtained using the MS approach. The MS technique has been recently used to describe the folding of peptides and small proteins in atomistic detail; it allows for a sampling 3–4 orders of magnitude longer than traditional calculations, and therefore allows for a direct comparison between simulated and experimentally determined bulk protein folding dynamics.[3] More recent work involving distributed computing has involved the construction of Markov State Models,[9,13,14] which allow for a more sophisticated use of a MS data set, but in this article, we examine the most simple style of MS analysis. In the present work, when referring to the MS sampling methodology, we mean: performing a large number of independent simulations, each one longer than the lag phases of the folding process (equilibration time), and using only the portion of the trajectories beyond the equilibration time for the calculation of the properties of the system.

Discrepancies between values of different structural metrics on the FL sets are always larger than those among MS sets. At the same time, uncertainties on the structural metrics (calculated by block averaging on each MS and FL set separately) are always significantly larger for the FL sets, and sometimes they are comparable to the average value of the metric. As a consequence, differences between the two FL simulations are not statistically significant whereas differences among MS simulations are statistically significant in some cases. The FL sets are found to be always equivalent to the MS-ran and MS-hel sets, which have the same starting structures, and also to the combined MS data sets, but not to the combined MS-hp and MS-sh sets. The latter sets show the largest discrepancies with all the others, due to the presence of a slight "memory" of β-hairpin and β-sheet starting structures, as highlighted by the analysis of RMSD and radius of gyration.

As for the kinetics of the equilibrium, the rates of interconversion between different species calculated from each of the four MS complete simulation sets are substantially equivalent, while the FL results show large variations. Also, the number of transitions that are unobserved in the FL simulations is much higher than the number of transitions missed in the MS simulations. If we break the MS sets into smaller parts of the same size as the FL sets, the deviations from the average transition rates are comparable for the MS and FL subsets when we consider transitions between highly populated clusters, whereas they are much larger in the FL subsets when we consider transitions between clusters with lower population. The FL approach is statistically less precise in the prediction of fast transition rates, but captures a few very slow transitions that are missed by the MS approach. A closer look at the average transition times reveals that transitions involving long β-hairpins and long α-helices are slower, whereas transitions involving short α- helices, 3_{10}-helices, and random coils are faster. Conversions of random coils to short helices (both α and 3_{10}) and β-bridged structures take place, on average, in 7–8 ns and 0.5 ns, respectively. The slow-

est transitions are found between long helices and long hairpins. This is consistent with the observation that the most likely mechanism of conversion between those structures involves conformations with very little or no secondary structure.[43] These features of the complex equilibrium studied highlight the complementarity of the FL and MS approaches, although it is worth noting that worse statistics is a major shortcoming of the FL approach (indeed, the FL approach misses 28 transition seen in the MS approach, while the MS approach only misses 2 seen in FL).

The differences in the results obtained with the two approaches are related both to the difference in the size of the data sets and to the sampling methodology. Our analysis of the kinetics clearly illustrates this point. When we consider the complete FL and MS data sets, we observe that MS sampling misses only 2 transitions whereas FL simulations miss 28 transitions. This large difference is mainly due to the size of the data sets: indeed, when we consider data sets of the same size, the number of missed transitions is similar. At the same time, when considering data sets of the same size, the MS approach gives more reproducible transition rates for fast transitions while the FL approach is more reliable for slow transitions. The MS approach also gives much lower uncertainties on most structural metrics analyzed. These differences are therefore due to the different sampling methodology.

On the basis our results, we find some limitations for both the MS and the FL approach. First, the timescale of each individual simulation should be longer than the lag phases of the folding process.[22] This observation is relevant for both the MS and the FL methodology. Paci et al. also highlighted the importance of sufficiently long individual trajectories in order to obtain reliable estimates of the folding time.[18] In the present work, we confirm that the length of each individual trajectory is crucial in the determination of thermodynamic properties of the system using the MS approach (if one does not use more sophisticated methods, such as Markov State Models). We also observe that, in the FL approach, the simulation length should be multiple times longer than the lag phases of the folding process in order to provide reasonable average properties for the system. Unfortunately, the minimum length required for the convergence of the simulations is not known *a priori*. In studies by Sorin et al. on the helix-coil transition in alanine-based peptides, a simulation time of ~50 ns was shown to be sufficient for complete convergence to ensemble equilibrium, allowing prediction of the thermodynamic properties of the peptides in several AMBER force fields under a variety of conditions with high precision.[10,44]

As complete ensemble convergence was not reached in our trajectories, we can conclude that the time required would be longer than 250 ns. This large difference in equilibration time in comparison to the polyalanine-based peptides studied by Sorin et al. may be due to several factors, including the force field, the simulation methodology, and the peptide sequence. Unlike the studies on polyalanine-based peptides, the RN24 peptide samples β-hairpin and β-sheet structures, which interconvert with helices on longer timescales than other transitions, consistent with longer folding times observed experimentally for β-hairpin and β-sheet structures.[6,7] Still, as the two systems were studied using different force fields and methodologies, it is non-trivial to assess the relative contributions of each of these factors to the difference in equilibration times.

One important limitation of the FL sampling approach is the high uncertainty on most structural metrics and on the prediction of numerous transition rates. Although the predictions obtained with the FL method are qualitatively in agreement with those obtained using the MS approach, the uncertainties obtained with the block averaging procedure are sometimes comparable to the average values; this is the case, for example, for the average number of residues with a helical secondary structure.

The main limitation of the MS approach appears to be the computational power required to observe near-convergence. Considering that each MS simulation set in the present work consists of 800 trajectories, an individual simulation length of 50 ns corresponds to a total simulation time of 160 μs. If we discard the initial 25 ns and calculate the average secondary structure over the next 25 ns, we get probabilities ranging between 1% and 32% for helical conformations (with higher population for simulations starting from the helical structure), and between 4% and 39% for beta structures (with higher population for the simulations starting from beta structures); the discrepancies are very large, despite the total sampling time of 160 μs. On the contrary, the two FL simulations, totaling 3.2 μs, give probabilities reasonably close to near-equilibrium values for most structural metrics, although the uncertainties in this case are large. However, adaptive sampling methods[13] suggest that MS methods can be done in a way that is more efficient than FL methods, within an MSM framework.

In our work, the FL approach required much less data storage and peptide conformations could be stored with higher temporal resolution, allowing for a detailed analysis of molecular motions on very short time scales. Indeed, long trajectories have been used to describe the mechanisms of conformational changes of the RN24 peptide, including the nucleation and propagation of helical structures.[43] In particular, a detailed analysis of the hydrogen bonding dynamics could only be performed considering structures stored with a temporal resolution of 1 ps. We point out here that data storage requirements are not related to the different techniques but only to the amount of sampling. Simulations run with the FL approach reaching the same amount of sampling as the MS data set would have exactly the same storage requirements. This leads us to a final consideration about the sampling methodologies. Because of the difficulties in the parallelization of MD algorithms, it is practically very difficult to obtain with few long simulations the same amount of sampling obtained here with many shorter simulations, using the same computational power and the same kind of hardware (PC), in a reasonable amount of time. In our case, each FL simulation required ~10 months, using one dual processor PC, while ~2 months were sufficient to collect 800 μs with the MS approach, using hundreds of PCs at the same time.

Conclusions

Molecular dynamics simulations were carried out on the RN24 peptide using few long microsecond timescale trajectories and many shorter independent simulations. Both data sets compare

favorably with NMR experiments and circular dichroism. Near-convergence of structural metrics was achieved with the MS approach when each individual simulation was longer than 225 ns, although a "memory" of the starting structure was observed at shorter timescales in systems starting from β-hairpin and β-sheet conformations. Secondary structure, radius of gyration, and free energy landscapes features indicate a qualitative agreement between FL and the MS approach. Uncertainties in all structural metrics are consistently larger in the FL sets than in MS subsets, even when subsets of the same size are considered. Large uncertainties are therefore related to the sampling methodology. Because of the large uncertainties in the FL data set, differences in structural metrics between the MS and the FL approach are not statistically significant according to the analysis of variance. Instead, significant differences are observed between the MS simulation sets starting from random or helical structures and those starting from β-hairpin or β-sheet structures, due to the presence of a slight memory of the starting structure in the latter sets. We hypothesize that the long equilibration times observed for RN24 derive from the complexity of the equilibrium, which involves numerous species with greatly varying secondary structural content. In particular, we notice that transitions between beta structures and all other conformations are the slowest. Beta structures are separated from other conformations by a higher energy barrier in the free energy landscape of the RN24 peptide, which highlights a correlation between free energy landscape and kinetics.

Analysis of the kinetics of the complex conformational equilibrium shows that the MS method provides more precise transition rates for fast transitions and for those transitions involving less populated conformational states. FL shows some advantage in the detection of very slow transitions. This difference appears to be related to the sampling methodology, not to the size of the data sets. The differences in the predictions obtained with the two sampling methodologies highlight the complementarity of the FL and MS approach for the study of conformational equilibria of peptides.

Acknowledgments

This work would not have been possible without the worldwide Folding@Home volunteers who contributed invaluable processor time (http://folding.stanford.edu). E.J.S. was supported by Veatch and Krell/DOE CGSF predoctoral fellowships. The computation was supported by ACS-PRF (36028-AC4), NSF Molecular Biophysics, NSF MRSEC CPIMA (DMR-9808677), and a gift from Intel. D.P.T. is an Alberta Heritage Foundation for Medical Research (AHFMR) Senior Scholar, CIHR New Investigator and Sloan Foundation Fellow. L.M. was supported by post-doctoral fellowships by the CISI-University of Milan and the AHFMR.

References

1. Duan, Y.; Kollman, P. A. Science 1998, 282, 740.
2. Garcia, A. E.; Onuchic, J. N. Proc Natl Acad Sci USA 2003, 100, 13898.
3. Snow, C. D.; Nguyen, N.; Pande, V. S.; Gruebele, M. Nature 2002, 420, 102.
4. Zagrovic, B.; Sorin, E.; Pande, V. Biophys J 2002, 82, 474A.
5. Zagrovic, B.; Sorin, E. J.; Pande, V. J Mol Biol 2001, 313, 151.
6. Eaton, W. A.; Munoz, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J. Annu Rev Biophys Biomol Struct 2000, 29, 327.
7. Eaton, W. A.; Munoz, V.; Thompson, P. A.; Henry, E. R.; Hofrichter, J. Acc Chem Res 1998, 31, 745.
8. Sugita, Y.; Okamoto, Y. Chem Phys Lett 1999, 314, 141.
9. Singhal, N.; Snow, C. D.; Pande, V. S. J Chem Phys 2004, 121, 415.
10. Sorin, E. J.; Pande, V. S. Biophys J 2005, 88, 2472.
11. Swope, W. C.; Pitera, J. W.; Suits, F. J Phys Chem B 2004, 108, 6571.
12. Park, S.; Pande, V. S. J Chem Phys 2006, 124, 054118.
13. Singhal-Hinrichs, N.; Pande, V. S. J Chem Phys 2007, 126, 244101.
14. Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. J Chem Phys 2007, 126, 155101.
15. Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. J Chem Phys 2007, 126, 155102.
16. Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B. Biopolymers 2003, 68, 91–109.
17. Shirts, M.; Pande, V. S. Science 2000, 290, 1903.
18. Rhee, Y. M.; Sorin, E. J.; Jayachandran, G.; Lindahl, E.; Pande, V. S. Proc Natl Acad Sci USA. 2004, 101, 6456.
19. Zagrovic, B.; Snow, C. D.; Shirts, M. R.; Pande, V. S. J Mol Biol 2002, 324, 1051.
20. Sorin, E. J.; Rhee, Y. M.; Nakatani, B. J.; Pande, V. S. Biophys J 2003, 85, 790.
21. Sorin, E. J.; Rhee, Y. M.; Pande, V. S. Biophys J 2005, 88, 2516.
22. Fersht, A. R. Proc Natl Acad Sci USA. 2002, 99, 14122.
23. Paci, E.; Cavalli, A.; Vendruscolo, M.; Caflisch, A. Proc Natl Acad Sci USA. 2003, 100, 8217.
24. Brown, J. E.; Klee, W. A. Biochemistry 1971, 10, 470.
25. Osterhout, J. J.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Dyson, H. J.; Wright, P. E. Biochemistry 1989, 28, 7059.
26. Berendsen, H. J. C.; van der Spoel, D.; Vandrunen, R. Comput Phys Commun 1995, 91, 43.
27. Daura, X.; Mark, a. E.; Van Gunsteren, W. F. J Comput Chem 1998, 19, 535.
28. Miyamoto, S.; Kollman, P. A. J Comput Chem 1992, 13, 952.
29. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. J Comput Chem 1997, 18, 1463.
30. Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. J Comput Chem 1999, 20, 786.
31. Darden, T.; York, D.; Pedersen, L. J. Chem Phys 1993, 98, 10089.
32. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. J Chem Phys 1995, 103, 8577.
33. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. and Haak, J. R. J Chem Phys 1984, 81, 3684.
34. Hess, B. J. Chem Phys 2002, 116, 209.
35. Kabsch, W.; Sander, C. Biopolymers 1983, 22, 2577.
36. Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. Nature 1987, 326, 563.
37. Bierzynski, A.; Kim, P. S.; Baldwin, R. L. Proc Natl Acad Sci USA. 1982, 79, 2470.
38. Feig, M.; MacKerell, A. D.; Brooks, C. L. J Phys Chem B 2003, 107, 2831.
39. Ohkubo, Y. Z.; Brooks, C. L. Proc Natl Acad Sci USA. 2003, 100, 13916.
40. Tirado-Rives, J.; Jorgensen, W. L. Biochemistry 1991, 30, 3864.
41. Schaefer, M.; Bartels, C.; Karplus, M. J Mol Biol 1998, 284, 835.
42. Schaefer, M.; Bartels, C.; Karplus, M. Theor Chem Acc 1999, 101, 194.
43. Monticelli, L.; Tieleman, D. P.; Colombo, G. J Phys Chem B 2005, 109, 20064.
44. Sorin, E. J.; Pande, V. S. J Comput Chem 2005, 26, 682.