



## A comparison of methods for calculating NMR cross-relaxation rates (NOESY and ROESY intensities) in small peptides

K. Anton Feenstra<sup>a</sup>, Christine Peter<sup>b</sup>, Ruud M. Scheek<sup>a</sup>, Wilfred F. van Gunsteren<sup>b</sup> & Alan E. Mark<sup>a,\*</sup>

<sup>a</sup>Bioson Research Institute and Laboratory of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands; <sup>b</sup>Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich, Switzerland

Received 28 January 2002; Accepted 20 May 2002

**Key words:** conformational dynamics, large time-scale dynamics, molecular dynamics, peptide nmr

### Abstract

Three methods for calculating nuclear magnetic resonance cross-relaxation rates from molecular dynamics simulations of small flexible molecules have been compared in terms of their ability to reproduce relaxation data obtained experimentally and to produce consistent descriptions of the system. The importance of the accuracy of the simulation versus the amount of sampling of phase space has also been assessed by comparing different length simulations performed with different time step schemes. A nine-residue peptide from the protein HPr of *E. coli* was used as a test system. The work shows that, in this case, single conformations or a limited ensemble of configurations are insufficient to properly describe the behavior of the peptide and that different approaches to incorporate molecular motions lead to significant differences in the cross-relaxation rates calculated. The correlation between the cross-relaxation rates calculated from simulations performed with different time step schemes was high and increased with increasing simulation length indicating that the extent of sampling is more important than the details of the atomic motion.

**Abbreviations:** ED – essential dynamics; FFT – fast Fourier transform; MD – molecular dynamics; MSF – mean square fluctuation; MSD – mean square displacement; NMR – nuclear magnetic resonance; NOE – nuclear Overhauser effect; NOESY – NOE spectroscopy; PDB – protein data bank; RMSD – root mean square difference; ROESY – rotating frame NOESY; SPC – simple point charge; TOCSY – total correlation spectroscopy.

### Introduction

Nuclear magnetic resonance (NMR) spectroscopy is currently the only experimental method that can yield high resolution structural information on peptides and proteins in solution. In particular, interproton dipolar cross relaxation rates derived from nuclear Overhauser effect (NOE or ROE) experiments can provide through space information on the proximity of different groups. In general the observed cross relaxation rates or NOE intensities are directly related to interatomic distances. This assumes the molecule (peptide

or protein) can be approximated as an isotropically tumbling rigid structure or in some cases an ensemble of isotropically tumbling structures. While this is in practice a reasonable assumption in the case of well structured larger molecules (Abseher et al., 1995; Schneider et al., 1999), evidence suggests that for small flexible molecules (such as peptides) ignoring internal dynamics and its correlation with the overall motion of the molecule may give rise to severe artifacts in the interpretation of NMR data (Jardetzki, 1980; Tropp, 1980; Lipari and Szabo, 1982; Brüschweiler et al., 1992; Post, 1992; Abseher et al., 1994; Daura et al., 1999a; Bürgi et al., 2001).

\*To whom correspondence should be addressed. E-mail: a.e.mark@chem.rug.nl

The recent availability of molecular dynamics (MD) simulations of small peptide systems that are much longer than the overall tumbling time of the molecules means that it is now possible to use approaches that involve less severe assumptions. For example, from the dynamical correlations in the simulation the cross relaxation rates can be computed directly. First, spectral densities are obtained by a Fourier transform of the time-correlation functions of the spherical harmonics pertaining to each of the proton pairs (see Equation 3 of Peter et al., 2001). Then, using the spectral densities, a set of coupled differential equations that describe the relaxation of the system of spins can be constructed. This relaxation matrix can subsequently be solved by diagonalization (Boelens et al., 1988; Bonvin et al., 1994; Peter et al., 2001). The incorporation of such dynamical correlations in the calculated cross-relaxation rates means that the results will be sensitive to the precise details of the simulation. This will include the extent of sampling of conformational space and the details of the dynamics as determined by the time step scheme used.

In the current paper we consider these issues by investigating the case of a nine-residue peptide from the histidine containing phosphocarrier protein HPr from *E. coli* (van Nuland et al., 1994). This particular peptide has been proposed as a possible folding nucleation site and thus was expected to adopt a stable spatial fold. Preliminary MD simulations supported this notion. Here we investigate whether the occurrence of a particular fold can be verified by comparing cross-relaxation rates calculated using different approaches to incorporate atomic motion with cross-relaxation rates observed experimentally.

## Methods

### Simulation set-up

#### Systems

Simulations were performed on a nine-residue peptide fragment corresponding to residues 35–43 from the protein HPr, an 85 residue  $\alpha$ /anti-parallel- $\beta$  sandwich protein (van Nuland et al., 1994). HPr is the histidine-containing phosphocarrier protein from the phosphoenolpyruvate-dependent phosphotransferase system of *E. coli*. The amino-acid sequence of the fragment is V T S N G K S A S. The N-terminus was acetylated and the C-terminus amidated,  $-\text{COCH}_3$  and  $-\text{CONH}_2$  respectively, in order to minimize the difference between the isolated peptide and the fragment in the protein.

*Table 1.* Overview of the five simulations of the nonapeptide in water. From left to right: starting conformation; time step ( $\Delta t$ ); treatment of hydrogen atoms and planar groups (H/plane): normal or **d/m** dummified and modified masses (Feenstra et al., 1999); identifying label; simulation length; number of water molecules; average box length; average radius of gyration  $R_{\text{gyr}}$  of the peptide in the simulation,  $R_{\text{gyr}}$  of folded starting structure is 0.55 nm, that of the extended starting structure is 0.99 nm

Starting structure	$\Delta t$ (fs)	H/ plane	Label	Length (ns)	No. H <sub>2</sub> O	Box (nm)	$R_{\text{gyr}}$ (nm)
Folded	2	n	F <sub>2</sub>	573	744	2.91	0.58
Folded	4	d/m	F <sub>4</sub>	250	852	3.00	0.57
Folded	7	d/m	F <sub>7</sub>	600	852	3.00	0.58
Extended	4	d/m	E <sub>4</sub>	196	1538	3.63	0.59
Extended	7	d/m	E <sub>7</sub>	691	1538	3.63	0.58
					881 <sup>a</sup>	2.93 <sup>a</sup>	

<sup>a</sup>Changed from a cubic box to a truncated octahedron at 303 ns.

### Parameters

All MD simulations were performed using the GROMOS-96 force field (Daura et al., 1998; van Gunsteren et al., 1996) and the GROMACS molecular dynamics package (Berendsen et al., 1995; van der Spoel et al., 1999). The LINCS algorithm (Hess et al., 1997) was used to constrain the length of all covalent bonds. Time steps used were 2, 4 and 7 fs. For the simulations at 4 and 7 fs, high frequency degrees of freedom involving hydrogen atoms were eliminated from the system where possible, or the high-frequency motions were reduced by modifying the mass distribution, as described previously (Feenstra et al., 1999; Feenstra, 2002). This allows a time step up to 7 fs to be used. A twin-range cut-off was used to evaluate the non-bonded interactions. Within the short-range cut-off of 1.0 nm the Van der Waals and Coulomb interactions were calculated every step. Within the long-range cut-off of 1.4 nm Coulomb interactions were only re-calculated during the neighbor-list update, about every 20 fs (10, 5 or 3 steps, depending on the time step). The temperature was maintained by weak coupling to a temperature bath (Berendsen et al., 1984) at 300 K with a time constant of 0.1 ps. The protein and water were coupled independently. The pressure was maintained by weak coupling with a time constant of 1.0 ps to an external pressure bath of 1 bar. A relative dielectric constant ( $\epsilon_r$ ) of 1.0 was used.

### Equilibration

The ‘folded’ structure of the nonapeptide was taken from the NMR solution structure of HPr (PDB

(Berman et al., 2000) entry 1hdn (van Nuland et al., 1994)). This structure and a fully extended structure were first energy minimized using a steepest descent algorithm. The resulting structures were each solvated in a cubic or truncated octahedral periodic box of simple point charge (SPC) water (Berendsen et al., 1981), with a minimum distance of 0.6 nm between the peptide and the box wall. The water box was constructed by replicating a cubic box containing 216 equilibrated SPC water molecules. All water molecules with the oxygen atom closer to any protein atom than the sum of their respective Van der Waals radii were removed. Simulation E<sub>7</sub> (Table 1) was changed from a cubic to a smaller truncated octahedral box after 303 ns by removing the water falling outside the octahedral box after centering the peptide. After energy minimization 100 steps of unrestrained MD using a 2 fs time step was performed to relax the systems.

#### *Production simulations*

Five separate simulations were performed for lengths varying from 200 to 700 ns using two different starting structures, folded and extended, and three different MD integration time steps, 2 fs, 4 fs and 7 fs. An overview of the simulations is given in Table 1.

#### *Analysis of conformational space*

##### *Covariance analysis*

To analyze the sampling of conformational space, Principle Component Analysis (PCA) of the fluctuations in the atomic coordinates of the backbone (N, C and C<sub>α</sub>) atoms, also referred to as essential dynamics (ED) analysis (Amadei et al., 1993), was performed. This was then used to estimate the degree of convergence between the trajectories by determining the overlap of the subspaces sampled. The subspace sampled in a trajectory is given by the positional covariance matrix  $M$  determined during the PCA analysis. As a measure of the overlap  $\mathcal{S}$  between two trajectories, the following expression was used (Hess, 2002):

$$\mathcal{S}(M_1, M_2) = 1 - \sqrt{\frac{\text{tr}([\sqrt{M_1} - \sqrt{M_2}]^2)}{\text{tr}(M_1) + \text{tr}(M_2)}}, \quad (1)$$

where  $M_1$  and  $M_2$  correspond to the positional covariance matrices of the peptide in the two trajectories and  $\text{tr}(X)$  denotes the trace of the matrix  $X$ . Extreme

values of  $\mathcal{S}$  are 1 for identical matrices (i.e. completely overlapping subspaces) and 0 for orthogonal subspaces (i.e. no overlap).  $\mathcal{S}$  is proportional to the overlap of the square root of the fluctuations. A more detailed description of this measure and a summary of its properties is given by Hess (2002).

##### *Cluster analysis*

Clustering of structures from the simulations into conformations was performed based on the atomic positional root-mean-square difference (RMSD) of the backbone atoms of the middle five residues (residues 3–7). A RMSD cut-off of 0.1 nm was used to determine neighboring structures belonging to one conformation. The procedure was as follows. For each structure, the number of neighbor structures was determined. The structure with the largest number of neighbors was then assigned to be the central structure of the first cluster or conformation and all its neighbors considered to be members of that cluster. These structures were then removed from the pool of structures and the procedure repeated until no structures remained in the pool. A more complete description of this clustering algorithm is given by Daura et al. (1999b).

#### *Analysis of physical properties*

##### *Diffusion and rotational properties*

Diffusion constants ( $D$ ) were calculated from the mean square displacement (MSD) of the center of mass of the peptide in each separate simulation using the Einstein relation for diffusion in three dimensions and a linear fit of a plot of the MSD vs. time. Rotational correlation times ( $\tau_{\text{rot}}$ ) for the peptide were calculated from a single-exponential fit to the averaged autocorrelation function for each of the components of the normalized normal vector of the plane through the C<sub>α</sub> atoms of residues 3, 5 and 7.

##### *Hydrogen bond analysis*

A hydrogen bond was considered to exist if the distance between the hydrogen and the acceptor atom was less than 0.25 nm and the angle donor-hydrogen-acceptor was larger than 120 degrees (Kabsch and Sander, 1983). Only hydrogen bonds present for more than 4% of the trajectory were considered significant. Hydrogen bond lifetimes were estimated from the integral of the autocorrelation of a function describing the existence of a hydrogen bond. This function is

assigned a value of one when the hydrogen bond exists and zero otherwise. To determine the average hydrogen bond lifetime the correlation functions for all hydrogen bonds were first summed. The average lifetime for each simulation was then calculated from the combined correlation function after subtracting the asymptotic value at infinite time and normalization.

### *NMR experiments*

#### *Sample*

The nonapeptide NAc-V-T-S-N-G-K-S-A-S-CONH<sub>2</sub>, was purchased from Eurosequence, Groningen as a freeze-dried powder. It was synthesized using solid-phase Fmoc chemistry from the C-terminus to the N-terminus and purified by HPLC to better than 95%. The N-terminus was acetylated and the C-terminus amidated. The sample composition was as follows: 50 mM potassium phosphate buffer at pH 6.5, with 5% (v/v) D<sub>2</sub>O, approximately 1 mg ml<sup>-1</sup> azide and 8 mg ml<sup>-1</sup> or 9 mM of peptide, in a total volume of 300 μl.

#### *Spectra*

All spectra were recorded on a Varian Unity Inova 600 MHz NMR spectrometer. <sup>1</sup>H TOCSY (Bax and Davis, 1985a; Braunschweiler and Ernst, 1983), <sup>1</sup>H NOESY (Macura and Ernst, 1980) and <sup>1</sup>H ROESY (Bax and Davis, 1985b; Bothner-By et al., 1984) experiments were performed. Watergate detection (Piotto et al., 1992) was used in the NOESY and ROESY experiments to suppress the water resonance. In addition, for the roesy experiments the water was pre-saturated. NOESY experiments were performed with a mixing time of 200 ms. Roesy experiments were performed with mixing times of 50, 100, 150, 200 and 300 ms. A spectral width of 8 kHz was used in both dimensions and free-induction decays were acquired with 1 024 data points. The maximum evolution time was 64 ms and 81 ms for the NOESY and ROESY experiments respectively. Positive and negative frequencies were discriminated as described by States et al. (1982). The sample temperature was 5 °C for all experiments. Data processing was performed using the program package SNARF (van Hoesel, 2000). Data points were weighted by a Lorentzian-to-Gaussian transformation in both frequency domains. Polynomial baseline corrections were applied in both domains. Cross-peak intensities were measured from peak heights, which is estimated to be accurate to within about 10%.

### *Calculation of NMR cross relaxation rates*

Theoretical ROESY spectra were calculated using three different approaches. The cross relaxation rates obtained are directly related to intensities and can thus be compared to intensities in the corresponding experimental 2D-NMR spectra.

The first approach (called 'flexible') accounts for aspects of internal dynamics by computing the spectral density functions for all proton pairs from the time correlation functions of the interproton vectors and by solving the system of coupled differential equations for the intensities by diagonalizing the relaxation matrix as described by Peter et al. (2001). Compared to the conventional method of analysis in which interproton distances inferred from experimental peak intensities are compared directly to average distances in simulations this approach has the advantage that effects due to internal dynamics as well as of spin diffusion on the spectral intensities are included. Internal dynamics come in via fluctuations in the interproton distance, individual correlation times for the motion of each interproton vector and possible effects of the coupling of internal dynamics with overall tumbling of the molecule. This influences the functional form of the time correlation function of the affected interproton vectors and thus the spectral density functions. Compared to the procedure described in the reference (Peter et al., 2001) small modifications were made. Specifically, in this study the spectral density functions are computed for *all* proton pairs and not only for a selected set. This was made possible by the use of fast Fourier transforms (FFT) to compute the time correlation functions.

The second approach (called 'rigid') computes  $r^{-6}$ -averages,  $\langle r^{-6} \rangle$ , for all interproton vectors over all structures of a trajectory or a subset of such structures, e.g. the member structures of a cluster or conformation. Spectral densities were computed from the  $r^{-6}$ -averages using a Lorentz function and an effective correlation time ( $\tau_{\text{eff}}$ ) which was the same for all proton pairs. From these spectral densities, the relaxation matrix was constructed and diagonalized to obtain the intensities. This approach takes into account effects due to spin diffusion, but not all effects of internal dynamics. Specifically the effects of internal motion on the interproton distances are included but not effects due to the angular motion of the interproton vector.

The third approach (called 'naive') computes  $r^{-6}$ -averages for all interproton vectors. Intensities are simply assumed to be proportional to these average

distances. Fluctuations in the molecule only effect the average distances. A relaxation matrix approach is not used and spin-diffusion is neglected. This is close to the conventional way of analyzing NMR spectra in terms of a single averaged structure and analyzing MD simulations by comparing experimentally derived distances with simulated distance averages. For a mixing time approaching zero the intensities obtained using the ‘rigid’ and the ‘naive’ approaches are identical as spin-diffusion effects are negligible for short mixing times. For this reason the ‘naive’ approach is also referred to as the initial rate approximation.

## Results

### *Sampling and convergence*

Five separate simulations of the nonapeptide were performed using two different starting conformations and three different time step schemes (see Table 1). To estimate the degree of overlap between the regions of configurational space sampled in the different trajectories, their respective covariance matrices were used to obtain an  $\mathcal{S}$  value based on equation 1. The values obtained ranged between 0.8 and 0.9 (data not shown) indicating a high level of overlap (Hess, 2002). A graphical representation of this overlap is presented in Figures 1A–E which show projections of the separate trajectories onto the first two eigenvectors of the combined trajectories. The similarity in the configurations sampled was also assessed by calculating the root-mean-square difference (RMSD) of the backbone atom positions for the middle five residues of the peptide for all possible pairs of structures of the combined trajectories. The variation of the RMSD values within a given trajectory was similar to the variation between trajectories. There are frequent occurrences of high and low RMSD values between different trajectories demonstrating the regions of conformational space sampled in the different trajectories are very similar.

The dynamical properties of the nonapeptide such as the diffusion constant ( $D$ ), the rotational correlation time of the peptide ( $\tau_{\text{rot}}$ ), and the hydrogen bond lifetime ( $\tau_{\text{HB}}$ ) for each of the five simulations are summarized in Table 2. The differences between the trajectories for  $D$  and  $\tau_{\text{rot}}$  are small and within the uncertainty of the calculations.

*Table 2.* Dynamical properties of the nonapeptide from the simulations. From left to right: simulation label (Sim.); diffusion constant for the peptide ( $D$ , with a standard deviation of  $0.1 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ ); overall rotation correlation time for the peptide ( $\tau_{\text{rot}}$ ); average lifetime of internal hydrogen bonds ( $\tau_{\text{HB}}$ ); and the effective correlation time ( $\tau_{\text{eff}}$ ) used in the ‘rigid’ and the ‘naive’ calculation methods of ROESY spectra as described in the text

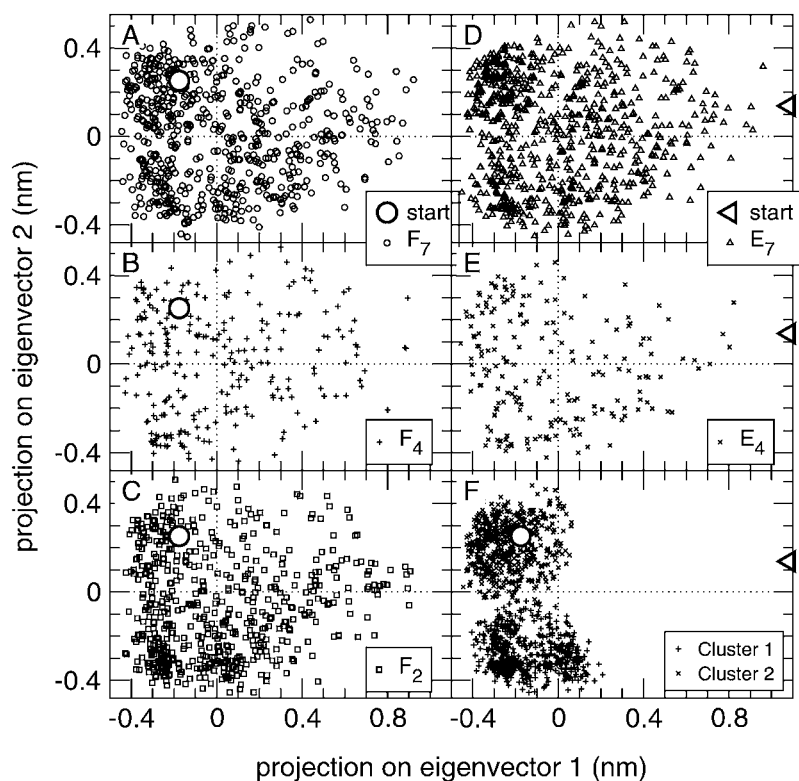
Sim.	$D$ ( $10^{-9} \text{ m}^2 \text{ s}^{-1}$ )	$\tau_{\text{rot}}$ (ps)	$\tau_{\text{HB}}$ (ns)	$\tau_{\text{eff}}$ (ps)
F7	0.55	213	0.35	65
F4	0.50	261	0.37	80
F2	0.45	252	0.61	80
E7	0.60	215	0.38	65
E4	0.59	236	0.53	75
avg.	0.54	235	0.45	73

### *Conformational analysis*

Based on the backbone RMSD values described above, cluster analysis was performed on the combined trajectories, using a cut-off for the RMSD of 0.1 nm. The trajectories are dominated by two clusters with roughly equal populations. The first cluster (23% occurrence) corresponds to a  $\beta$ -turn conformation with a one-residue register shift in hydrogen bonding with respect to the conformation found in the protein. The second cluster (22% occurrence) corresponds to the  $\beta$ -turn conformation found in the protein. No other cluster is populated for more than 8% of the time. These minor clusters cover a wide range of conformations. For example, the third cluster (8% occurrence) has a curled conformation with a single turn of  $\alpha$ -helix at one end. Figure 2 shows the middle structures of the two largest clusters, as well as backbone traces for a selection of structures belonging to each cluster.

Figure 1F shows a projection of the conformations from the two largest clusters onto the two largest eigenvectors from the ED analysis of the combined trajectories. It is clear from this projection that the clusters are well-separated in conformational space. Approximately 500 transitions in and out of each of the two largest clusters were observed during the trajectory suggesting that the relative populations are statistically reliable.

Table 3 lists the occurrence of backbone-backbone hydrogen bonds in the combined trajectories of more than 4%. The maximum occurrence of any hydrogen bond in the combined trajectory is only 11%, indicating that the peptide is very dynamic. Many hy-



*Figure 1.* Comparison between sampled space of each of the five trajectories (A–E) and comparison between the two largest clusters of the trajectories combined (panel F). Shown are the projection of structures taken from each of the trajectories onto the first two eigenvectors of the five trajectories combined (small symbols), a projection of structures from the two largest clusters from the combined trajectories (small symbols in panel F), and a projection of the two starting structures: Large circle for the folded structure and large triangle for the extended structure. One structure in every 1 ns of the trajectory was used.

drogen bonds, however, were observed with a higher occurrence in specific clusters. The hydrogen bonding pattern in cluster 2 corresponds closely with that observed in the ensemble of NMR solution structures of the whole protein. No significant difference in hydrogen bonding pattern was found between the five separate simulations (data not shown).

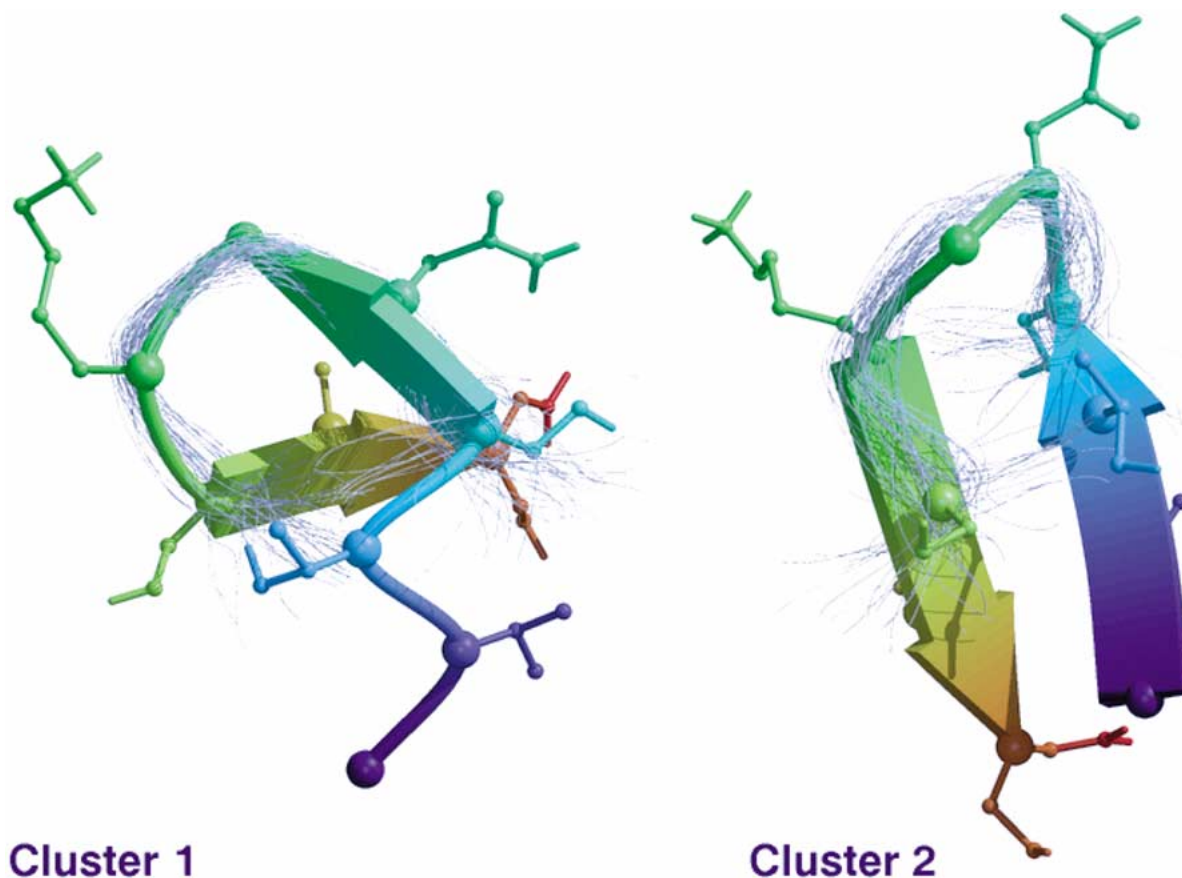
#### *NMR experiments*

##### *Assignments*

Chemical shift assignments derived from TOCSY and NOESY experiments are listed in Table 4. Nuclei with an identical chemical environment or for which no distinction in chemical shift could be made, are grouped together. These are indicated by an asterisk ('\*') in place of a number in the atom names. The majority of resonances could be unambiguously assigned. No stereospecific assignments were made.

##### *ROESY intensities*

ROESY spectra were recorded using a series of 5 mixing times. Intensities for all 5 mixing times could be measured for 113 cross peaks. Of these, 10 correspond to non-overlapping cross-peaks between non-adjacent residues, i.e., are structurally most relevant. In addition, 85 non-ROE locations were identified. These are defined as positions in the spectra which correspond to the chemical shifts of a pair of protons where the experimental intensities are below the noise cut-off value, i.e., no peak is observed. A comparison between the experimental and the theoretical intensities was made for the 113 ROE peaks and the 85 non-ROEs. Approximately half of the peaks involve backbone hydrogens. Results are presented for a mixing time of 200 ms.



*Figure 2.* Structures from the two largest clusters from the combined trajectories. Cartoon models represent the middle structure of each cluster. Thin gray lines represent backbone traces of every tenth structure from the corresponding cluster. Plots were generated using a modified version of MolScript (Kraulis, 1991; Esnouf, 1997) and Raster3D (Bacon and Anderson, 1988; Merritt and Murphy, 1994).

### *Comparisons using theoretical intensities*

#### *Calculation*

ROE intensities were calculated for all possible proton pairs for each of the simulations using the three different approaches described in the Methods section, ‘flexible’ (accounts for all aspects of internal dynamics and spin-diffusion effects), ‘rigid’ (includes spin-diffusion effects but accounts for conformational fluctuations only through distance averaging) and ‘naive’ (no spin-diffusion effects, only distance averaging). The pairs of protons within CH<sub>2</sub>, NH<sub>3</sub> and NH<sub>2</sub> groups were excluded from the comparison as the distances between these protons depend solely on the bonded interaction parameters of the force field. There were 63 protons in the peptide of which 56 were observed (see Table 4). This corresponds to 53 relaxation sites (using a single site for methyl protons, and one site for all other protons). In total intensities

for 1378 pairs and 53 diagonal elements were calculated. These are referred to as ‘proton pair intensities’. The effective correlation time  $\tau_{\text{eff}}$  used in calculating the ‘rigid’ intensities was empirically adjusted such that the calculated intensities of the backbone proton pairs match (as closely as possible) those from the ‘flexible’ approach. The effective correlation time is much smaller than the overall rotational correlation time of the peptide ( $\tau_{\text{rot}}$ ) because it includes contributions from internal dynamics. The values of  $\tau_{\text{eff}}$  for each of the five simulations are listed in Table 2.

#### *Sampling effects*

To illustrate the effects of the simulation protocol and of sampling on the calculated intensities, examples of the correlation between the calculated proton pair intensities from different simulations using a ‘rigid’ and a ‘flexible’ treatment and a mixing time of 200 ms are shown in Figures 3 and 4. Figure 3 illustrates the in-

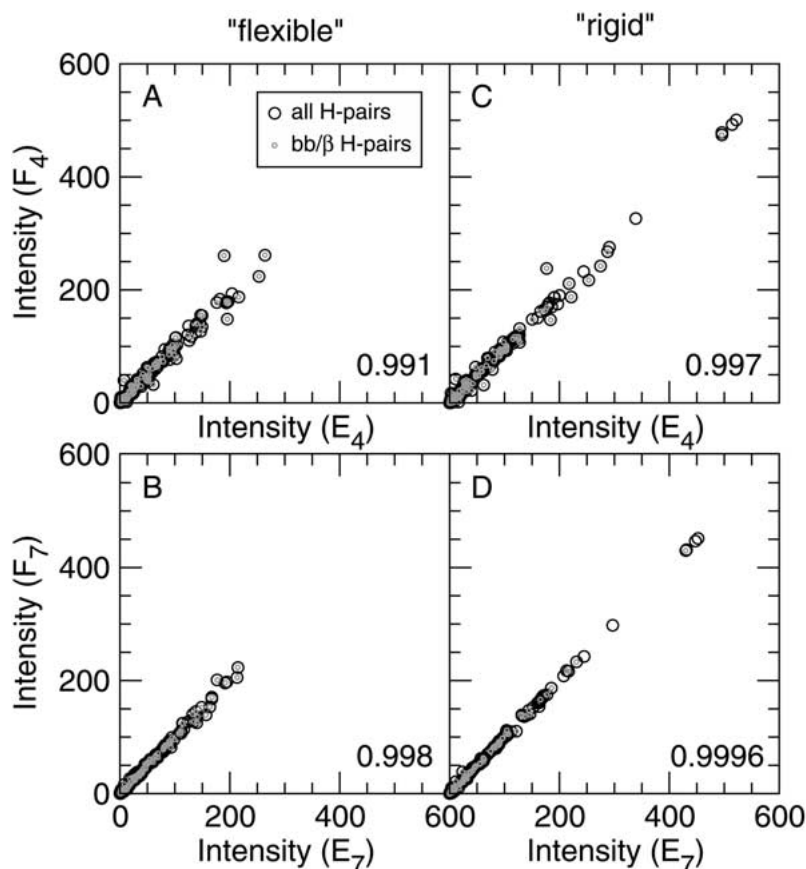


Figure 3. Comparison of theoretical proton pair intensities using the 'flexible' (A and B) and 'rigid' (C and D) methods between simulations with different starting structures (E: Extended; F: Folded) and of different lengths ( $E_4$ ,  $F_4$ :  $\sim 200$  ns;  $E_7$ ,  $F_7$ :  $\sim 600$  ns). Distinction is made between all proton pairs and those involving backbone or  $\beta$  protons. Numbers refer to the correlation coefficients for a linear fit to the plot of all proton pairs.

fluence of the starting structure (E versus F) and of the simulation length (200 ns for  $F_4$  with  $E_4$  versus 600 ns for  $F_7$  with  $E_7$ ). In all cases a high degree of correlation was found ( $> 0.99$ ). This indicates that the spectra are either insensitive to the differences between the simulations or that the simulations are indeed very similar. The correlations are in general higher for the longer simulations ( $F_7$  vs.  $E_7$ : 0.998 'flexible' and 0.9996 'rigid') than for the shorter simulations ( $F_4$  vs.  $E_4$ : 0.991 'flexible' and 0.999 'rigid'). The calculated peak intensities converge with increased sampling independent of the starting structure. The data can also be weighted by using  $I_{\text{ROE}}^{1/6}$  instead of  $I_{\text{ROE}}$ . This subjectively weights the data according to its structural relevance (AB et al., 2001). Since  $I_{\text{ROE}}^{-1/6}$  is related to interatomic distance,  $I_{\text{ROE}}^{1/6}$  can be thought of as a reciprocal apparent distance. Using this weighting gives a slightly clearer separation of the correlation:  $F_7$  vs.  $E_7$

yields 0.99, whereas  $F_4$  vs.  $E_4$  yields 0.97. The correlations for the 'flexible' spectra are lower than for the 'rigid' ones. In the figure it is also indicated which intensities correspond to backbone-proton pairs. No significant difference between all pairs as opposed to backbone pairs is evident.

Figure 4 shows examples of the correlation between simulations performed using different time steps and different treatments of hydrogens and planar groups starting from the folded structure. The simulation protocol has apparently little effect on the properties of the ensemble simulated in regard to the calculation of the NMR proton pair intensities.

#### Theoretical models

In this section the intensities computed using the different theoretical models described above are compared in order to assess the importance of various approximations within the models. Figure 5 shows the



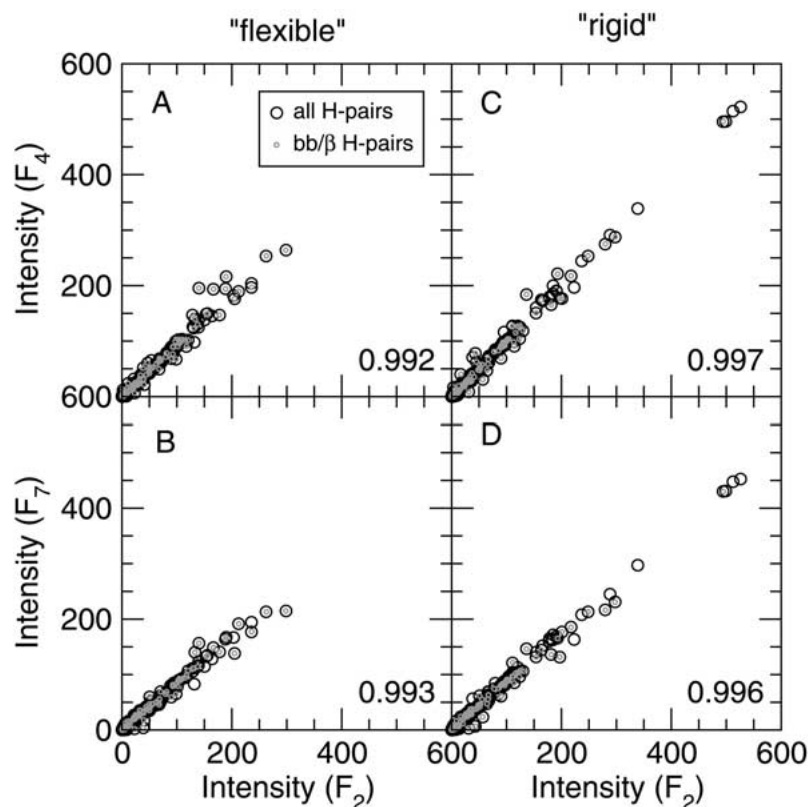
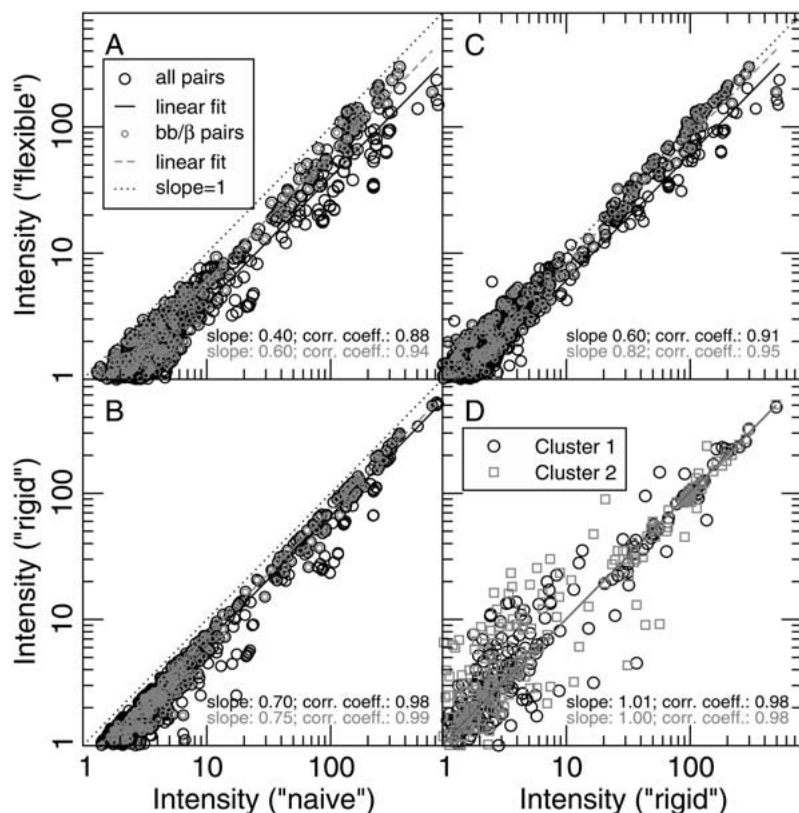


Figure 4. Comparison of theoretical proton pair intensities using the ‘flexible’ (A and B) and ‘rigid’ (C and D) methods between simulations with different time steps/hydrogen treatments ( $F_2$ : Normal dynamics with  $\Delta t = 2$  fs;  $F_4$  and  $F_7$ : Dummified hydrogens and modified masses with  $\Delta t = 4$  fs and 7 fs, respectively; see Table 1). Distinction is made between all proton pairs and those involving backbone or  $\beta$  protons. Numbers refer to the correlation coefficients for a linear fit to the plot of all proton pairs.

correlation of the theoretical intensities computed with the ‘flexible’ approach with those computed with the ‘naive’ approach (Figure 5, panel A), the correlation between the ‘rigid’ and the ‘naive’ approach (panel B), and the correlation between the ‘flexible’ and the ‘rigid’ approach (panel C). The values are plotted using a log-log scale, which roughly weights the peaks according to their importance. Less intense peaks which often correspond to structurally relevant distant interactions are emphasized. The comparison between the ‘rigidly’ and the ‘naively’ computed spectra (panel B) highlights the importance of spin diffusion. Especially for low intensity peaks, spin diffusion can change the relative intensity of two peaks by a factor of two or more. In the figure, the important aspect is not the deviation of the slope of the correlation curve from 1. Spin diffusion, in general, lowers cross peak intensities by providing additional relaxation pathways. What is relevant, is that this effect is different for different proton pairs changing the relative intensities of

the cross peaks. This results in the spread of the points and the deviation of the correlation coefficient from 1. The effect is greater for longer mixing times and is significant for mixing times longer than 100 ms for this system.

The comparison of the ‘flexible’ and the ‘rigid’ approach (Figure 5C) highlights those effects of internal dynamics on the spectral intensities that go beyond simple distance averaging. In small highly flexible molecules, separate regions of the molecule may show individual correlation times significantly different from the molecule as a whole. For example, whereas the rotational correlation time for the molecule overall is approximately 230 ps the effective average rotational correlation time is only around 70 ps. Thus, some parts of the molecules show much greater motion than the whole which will in turn affect the intensities of proton pairs located in those regions. Again, it is not the slope of the correlation curve that is relevant. Additional motions will in gen-



*Figure 5.* Comparison between theoretical proton pair intensities, calculated using three different approaches at a mixing time of 200 ms and for the  $F_2$  trajectory and for the two largest clusters of the combined trajectories. The intensities obtained with the three methods, 'naive', 'rigid' and 'flexible', are plotted against each other in A–C. Distinction is made between all proton pairs and those involving backbone or  $\beta$  protons. D shows intensities calculated using the 'rigid' method for the two largest clusters vs. the whole trajectory (for backbone protons only). Linear fits are also shown and the corresponding slopes and correlation coefficients are reported.

eral shorten the 'effective' correlation time and induce further relaxation processes. Instead, one should focus on the spread of the points. The spread shows that internal dynamics have a different effect on different intensities and thus on the relative lengths of distances inferred from experimental NMR spectra. The comparison between the 'flexible' and the 'naive' approach (Figure 5A) illustrates these influences. In Figure 5 it can be seen that the effect of neglecting spin diffusion and local motion is more significant for pairs involving a proton on a sidechain than for pairs involving two backbone protons.

#### *Complete ensembles and single conformations*

The sensitivity of the ROESY spectra to the backbone conformation was investigated by calculating theoretical ('rigid') spectra for groups of structures representing a particular backbone conformation (i.e., members of one cluster) and comparing these spectra with that calculated from the complete trajectory. Of interest

was whether a single conformation or a small number of conformations, for example the two most important clusters, were sufficient to approximate the complete ensemble of states simulated. Or, whether a particular cluster represented the experimental NMR data better than the other clusters. In Figure 5D the correlation of the 'rigid' spectra of the first two clusters with the 'rigid' spectrum of the complete simulation is displayed (for backbone protons only) on a logarithmic scale. The spread of the points shows that the spectrum is sensitive to the conformation. The correlation coefficient is nevertheless very high and in fact higher than the correlation for the comparison between the 'flexible' and the 'rigid' approach; 0.95 for 'flexible' vs. 'rigid' and 0.98 for both clusters 1 and 2 vs. complete trajectory. This high correlation coefficient is in part due to the fact that it was computed with a linear fit, which strongly emphasizes the peaks with high intensities. Such proton pairs are very close in

**Table 3.** Selected hydrogen bonds in the backbone which occur in the combined trajectories for more than 4% (21 in total). From left to right occurrence in the ensemble of protein NMR solution structures, in all five trajectories and in the two largest clusters. An occurrence larger than 10% is indicated in bold. An occurrence smaller than 4% is considered insignificant and is indicated by ‘-’

Hydrogen-Bond		Occurrence in			
Donor	Acceptor	Protein	Combined trajectories	Cluster No.	
N-H	O			1	2
Thr 2 – Ace 1		-	4	4	5
Thr 2 – Ser 9		-	4	<b>14</b>	-
Ser 3 – Val 1		<b>17</b>	5	5	9
Ser 3 – Lys 6		-	4	-	<b>13</b>
Ser 3 – Ser 7		-	5	6	8
Asn 4 – Val 1		-	6	5	-
Asn 4 – Thr 2		<b>13</b>	6	<b>11</b>	5
Asn 4 – Ser 7		-	8	<b>33</b>	-
Gly 5 – Ser 3		-	5	5	-
Gly 5 – Ala 8		-	6	-	<b>10</b>
Lys 6 – Ser 3		<b>93</b>	6	-	<b>18</b>
Lys 6 – Asn 4		<b>70</b>	9	-	<b>18</b>
Ser 7 – Ser 3		-	5	-	9
Ser 7 – Asn 4		-	<b>11</b>	<b>37</b>	-
Ser 7 – Gly 5		-	5	6	5
Ala 8 – Val 1		<b>93</b>	5	-	<b>12</b>
Ala 8 – Ser 3		-	5	-	<b>12</b>
Ala 8 – Asn 4		-	7	<b>21</b>	-
Ala 8 – Gly 5		-	5	5	-
Ala 8 – Lys 6		-	4	5	6
Ser 9 – Thr 2		-	5	<b>17</b>	-

space and usually also close in sequence. This means that they are of little relevance structurally.

Both cluster 1 and cluster 2 represent the spectrum of the complete simulation equally badly. Using  $I_{\text{ROE}}^{1/6}$  instead of a  $I_{\text{ROE}}$  (AB et al., 2001) weighting the correlation coefficients obtained correspond better to the ‘visual’ impression gained from Figure 5D: ‘flexible’ vs. ‘rigid’ gives 0.99, cluster 1 vs. complete trajectory gives 0.87 and cluster 2 vs. complete trajectory gives 0.84.

In summary, neither of the predominant conformations is sufficient to reproduce the spectrum calculated from the complete trajectory. In addition, neither of the two conformations represents the ensemble significantly better than the other. It is probably not possible to represent a spectrum of such a flexible compound by just one conformation. The spectra are sensitive to the conformations sampled and any comparison between

**Table 4.**  $^1\text{H}$  chemical shift assignments ( $\sigma_{\text{H}}$  in ppm) for all observable  $^1\text{H}$  nuclei, 56 in total. Multiple  $^1\text{H}$  nuclei which are bonded to the same heavy atom and have identical chemical shifts are grouped together, which is indicated using a ‘\*’ instead of a number in the atom name. Non-overlapping methylene resonances are indicated by the subscripts 1 or 2 referring to low and high field resonances respectively

Residue	Atom	$\sigma_{\text{H}}$	Residue	Atom	$\sigma_{\text{H}}$			
	Ace	H $_{\alpha^*}$	2.048	6	Lys	H	8.255	
1	Val	H	8.371			H $_{\alpha}$	4.381	
		H $_{\alpha}$	4.155			H $_{\beta 1}$	1.853	
		H $_{\beta}$	2.074			H $_{\beta 2}$	1.757	
		H $_{\gamma^{**}}$	0.951			H $_{\gamma^*}$	1.405	
2	Thr	H	8.450			H $_{\delta^*}$	1.656	
		H $_{\alpha}$	4.263			H $_{\epsilon^*}$	3.026	
		H $_{\beta}$	4.444	7	Ser	H	8.542	
		H $_{\gamma 2^*}$	1.206			H $_{\alpha}$	4.450	
3	Ser	H	8.550			H $_{\beta^*}$	3.882	
		H $_{\alpha}$	4.447	8	Ala	H	8.644	
		H $_{\beta 1}$	3.891			H $_{\alpha}$	4.359	
		H $_{\beta 2}$	3.851			H $_{\beta^*}$	1.417	
4	Asn	H	8.658	9	Ser	H	8.362	
		H $_{\alpha}$	4.734			H $_{\alpha}$	4.361	
		H $_{\beta^*}$	2.852			H $_{\beta^*}$	3.872	
		H $_{\delta 21}$	7.730			NH $_2$	H $_{\text{N}1}$	7.616
		H $_{\delta 22}$	7.029				H $_{\text{N}2}$	7.289
5	Gly	H	8.502					
		H $_{\alpha^*}$	3.946					

experimental intensities and intensities computed from MD simulations is only meaningful if the ensemble simulated contains all ‘relevant’ conformations.

#### Comparison between theoretical and experimental intensities

Intensities calculated from the simulation data were also compared to intensities obtained experimentally. The theoretical intensities were generated by summing the individual proton pair intensities for all pairs possibly associated with a location in the spectra obtained experimentally, based on the chemical shift assignments listed in Table 4 and the observed peak widths. Experimental intensities for which the theoretical prediction would depend primarily on bonded interaction parameters in the force field were excluded from the comparison. In total there were 98 locations for which peak intensities could be determined experimentally. In addition, 85 locations were considered for which peak intensities could be calculated but for which no peak was evident in the experimental spectrum (non-ROEs).

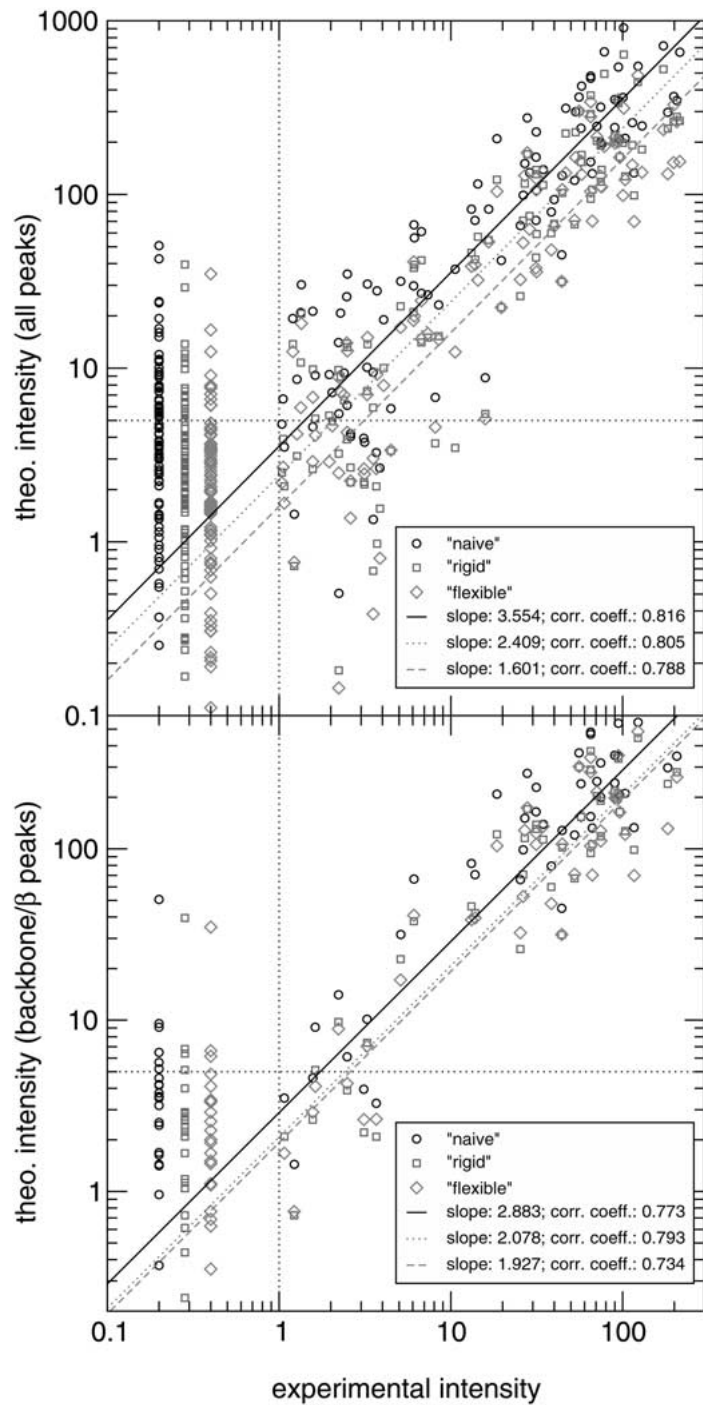


Figure 6. Comparison between theoretical and experimental ROE peak and non-ROE intensities. Experimental non-peak ‘intensities’ are set to arbitrary values of between 0.2 and 0.4 (on the left-hand side in the plot) with an offset added for clarity between the three data sets in each plot. Dotted lines indicate the noise level (1) as determined from the experimental spectra. Theoretical intensities are shown for three models for the molecular motion. The top panel shows all proton pairs, the bottom panel only those involving backbone or  $\beta$  protons.

Figure 6 shows the correlation between the calculated and experimental intensities using the three methods. The correlation between the theoretical and experimental intensities is virtually identical for the three methods (vis.  $\sim 0.8$ ). The noise level in the experimental spectra is about 1 in the units used in Figure 6. The absolute values of the experimentally derived intensities and those calculated using the various methods cannot be compared directly. Therefore, it was required that the predicted non-intensities were below 5. The average calculated intensity for the non-ROE's was 6.7 for the 'naive' method, 4.1 for 'rigid' and 3.2 for 'flexible'. This however is not a significant difference. The theoretically computed intensities from the 'naive' approach are systematically higher than those from the 'rigid' approach which are again higher than those from the 'flexible' approach. This is discussed in the section 'Theoretical models' and is also reflected in the different slopes in Figure 6.

Based on the correlation between the experimental and the theoretical intensities no distinction can be made between the three calculation methods used. None of the methods reproduces the experimental data better than the others. The deviations from experiment may be due to (i) the force field and/or simulation methods used, (ii) the time scale that was simulated which is still much shorter than the experimental timescale, (iii) limitations in the models used to calculate the theoretical intensities or (iv) limitations within the experimental data. For example, cross-relaxation rates were estimated from peak heights as opposed to peak volumes. In weighting these various factors it should be noted that the comparison in Figure 6 was performed using the  $F_2$  simulation which does not have the special treatment of hydrogens and planar groups. Results obtained for the other trajectories (data not shown) are very similar to those for  $F_2$ .

Although the spectra computed from the complete simulation and from single clusters are very different, no single conformation nor group of conformations gave a significantly better agreement with the experimental data than the whole trajectory. In all the cases the correlation between intensities computed theoretically and intensities obtained experimentally were similar ( $\sim 0.8$ , data not shown).

## Discussion

In the simulations in aqueous solution the nonapeptide adopts a small number of comparatively stable con-

formations. The dominant cluster or conformation, a  $\beta$ -turn, is populated for 23% of the time. The second most populated conformation, also a  $\beta$ -turn, has a hydrogen bonding pattern similar to that of the nonapeptide in the protein HPr and is only slightly less common than the first being populated for 22%. The question we have tried to address is whether the conformations sampled in the simulations are compatible with the measured data.

Clearly, if the simulations accurately reproduce the behavior of the peptide in solution both in terms of the conformations sampled and in terms of the dynamic properties of the system, the proton pair intensities calculated using the 'flexible' approach should correspond most closely to the the experimentally measured intensities. What was found was that given the accuracy of the calculations and the correspondence between the experimental and theoretical peak intensities it is not possible to make a distinction between the different methods used to derive the theoretical intensities. A similar correspondence with the experimental data was found for each of the methods even when intensities were naively assumed to be directly proportional to  $\langle r^{-6} \rangle$  averaged distances. All three methods show a large spread in intensities around the experimental values. This spread which was in the order of a factor of 10 would correspond to a spread in the interproton distances of about a factor of  $10^{1/6} \approx 1.5$  or a deviation of around 50%. This does not mean that an appropriate treatment of local motion is unimportant. It does mean, however, that if local dynamics is included in the analysis the requirement that the simulations correctly reproduce small differences in the relative populations of different conformers will be more stringent.

By comparing the properties calculated from the different trajectories the effects of using different time steps and using different methods to treat hydrogen atoms and planar groups was investigated. The sampling of conformational space was analyzed by calculating the overlap between different trajectories. This was achieved by means of an  $\mathcal{S}$  parameter as described in the Methods and Results sections and by comparing the projection of the trajectories onto the two largest eigenvectors of the combined trajectory (Figure 1). In both cases no effect of the size of the time step was apparent. A direct comparison of all conformations in terms of the RMSD of the atomic positions of the backbone of the middle five residues also showed that the trajectories were effectively indistinguishable. Marginal differences were, however,

observed in the occurrence of two out of the 21 hydrogen bonds analyzed between simulations performed using different time steps. Likewise, dynamical parameters such as the diffusion, rotational correlation time, average hydrogen bond lifetime, average cluster lifetime all showed marginal differences as might be expected.

Our results demonstrate that for the case of a small flexible polypeptide in solution the use of different approaches to treat the effects of molecular motion can lead to significant variations in calculated NOE/ROE intensities and hence in the interpretation of experimental spectra. This shows again that in the case of small flexible peptides in solution 2D-NMR spectra alone provide insufficient information to determine what are the dominant conformations in solution. Detailed information on the nature of the molecular motion is needed to correctly back calculate the experimental data. Only when we have simulations of sufficient length and accuracy to be able to predict the experimental spectra in detail will we be able to have a true picture of the behavior of these systems at an atomic level.

## Acknowledgements

K.A.F. acknowledges support from the Netherlands Foundation for Life Sciences (SLW) with financial aid from the Netherlands Organization for Scientific Research (NWO).

## References

- AB, E., Schuurman-Wolters, G.K., Nijlant, D., Dijkstra, K., Saier, M.H., Robillard, G.T. and Scheek, R.M. (2001) *J. Mol. Biol.*, **308**, 993–1009.
- Abseher, R., Ludemann, S., Schreiber, H. and Steinhauser, O. (1994) *J. Am. Chem. Soc.*, **116**, 4006–4018.
- Abseher, R., Ludemann, S., Schreiber, H. and Steinhauser, O. (1995) *J. Mol. Biol.*, **249**, 604–624.
- Amadei, A., Linssen, A.B.M. and Berendsen, H.J.C. (1993) *Proteins: Struct. Funct. Gen.*, **17**, 412–425.
- Bacon, D.J. and Anderson, W.F. (1988) *J. Mol. Graphics*, **6**, 219–220.
- Bax, A. and Davis, D.G. (1985a) *J. Magn. Reson.*, **63**, 203–206.
- Bax, A. and Davis, D.G. (1985b) *J. Magn. Reson.*, **63**, 207–213.
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermans, J. (1981) In *Intermolecular Forces*, Pullman, B. (Ed.), D. Reidel Publ. Co. Dordrecht, pp. 331–342.
- Berendsen, H.J.C., van der Spoel, D. and van Drunen, R. (1995) *Comp. Phys. Comm.*, **91**, 43–56.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) *Nucl. Acids Res.*, **28**, 235–242.
- Boelens, R., Koning, T.M.G. and Kaptein, R. (1988) *J. Mol. Struct.*, **173**, 299–311.
- Bonvin, A.M.J.J., Vis, H., Breg, J.N., Burgering, M.J.M., Boelens, R. and Kaptein, R. (1994) *J. Mol. Biol.*, **236**, 328–341.
- Bothner-By, A.A., Stephons, R.L., Lee, J., Warren, C.D. and Jeanloz, R.W. (1984) *J. Am. Chem. Soc.*, **106**, 811–813.
- Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.*, **53**, 521–528.
- Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M. and Ernst, R.R. (1992) *J. Am. Chem. Soc.*, **114**, 2289–2302.
- Bürgi, R., Pitera, J. and van Gunsteren, W.F. (2001) *J. Biomol. NMR*, **19**, 305–320.
- Daura, X., Antes, I., van Gunsteren, W.F., Thiel, W. and Mark, A.E. (1999a) *Proteins*, **36**, 542–555.
- Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W.F. and Mark, A.E. (1999b) *Angew. Chem. Intl. Ed.*, **38**, 236–240.
- Daura, X., Mark, A.E. and van Gunsteren, W.F. (1998) *J. Comp. Chem.*, **19**, 535–547.
- Esnouf, R.M. (1997) *J. Mol. Graphics*, **15**, 132–134.
- Feenstra, K.A. (2002) *Long Term Dynamics of Proteins and Peptides*, PhD thesis, Nijenborgh 4, 9747 AG, Groningen.
- Feenstra, K.A., Hess, B. and Berendsen, H.J.C. (1999) *J. Comp. Chem.*, **20**, 786–798.
- Hess, B. (2002) *Phys. Rev.*, **E65**, 031910.
- Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) *J. Comp. Chem.*, **18**, 1463–1472.
- Jardetzki, O. (1980) *Biochim. Biophys. Acta*, **621**, 227–232.
- Kabsch, W. and Sander, C. (1983) *Biophys. J.*, **22**, 2577–2637.
- Kraulis, P.J. (1991) *J. Appl. Cryst.*, **24**, 946–950.
- Lipari, G. and Szabo, A. (1982) *J. Am. Chem. Soc.*, **104**, 4546–4559.
- Macura, S. and Ernst, R.R. (1980) *Mol. Phys.*, **41**, 95–117.
- Merritt, E.A. and Murphy, M.E.P. (1994) *Act. Cryst.*, **D50**, 869–873.
- Peter, C., Daura, X. and van Gunsteren, W.F. (2001) *J. Biomol. NMR*, **20**, 297–310.
- Piotto, M., Sandek, V. and Sklenár, V. (1992) *J. Biomol. NMR*, **2**, 661–665.
- Post, C.B. (1992) *J. Mol. Biol.*, **224**, 1087–1101.
- Schneider, T.R., Brünger, A.T. and Nilges, M. (1999) *J. Mol. Biol.*, **285**, 727–740.
- States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Reson.*, **48**, 286–292.
- Tropp, J. (1980) *J. Chem. Phys.*, **72**, 6035–6043.
- van der Spoel, D., Hess, B., Feenstra, K.A., Lindahl, E. and Berendsen, H.J.C. (1999) *Gromacs User Manual Version 2.0*, Nijenborgh 4, 9747 AG Groningen, The Netherlands. Internet: <http://md.chem.rug.nl/~gmx>
- van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. and Tironi, I.G. (1996) *Biomolecular Simulation: GROMOS96 Manual and User Guide*, BIOMOS B.V., Zürich, Groningen.
- van Hoesel, F.H.J. (2000) *SNARF v. 0.8.9*. University of Groningen, Groningen.
- van Nuland, N., Hangyi, I.W., van Schaik, R.C., Berendsen, H.J.C., van Gunsteren, W.F., Scheek, R.M. and Robillard, G.T. (1994) *J. Mol. Biol.*, **237**, 544–559.