PROTEINS: Structure, Function, and Genetics 36:542–555 (1999)

The Effect of Motional Averaging on the Calculation of NMR-Derived Structural Properties

Xavier Daura,[1] Iris Antes,[1,2] Wilfred F. van Gunsteren,[1] Walter Thiel,[2] and Alan E. Mark[1]*
[1]*Laboratory for Physical Chemistry, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland*
[2]*Institute of Organic Chemistry, University of Zürich, Zürich, Switzerland*

ABSTRACT **The effect of motional averaging when relating structural properties inferred from nuclear magnetic resonance (NMR) experiments to molecular dynamics simulations of peptides is considered. In particular, the effect of changing populations of conformations, the extent of sampling, and the sampling frequency on the estimation of nuclear Overhauser effect (NOE) inter-proton distances, vicinal ^3J-coupling constants, and chemical shifts are investigated. The analysis is based on 50-ns simulations of a β-heptapeptide in methanol at 298 K, 340 K, 350 K, and 360 K. This peptide undergoes reversible folding and samples a significant proportion of the available conformational space during the simulations, with at 298 K being predominantly folded and at 360 K being predominantly unfolded. The work highlights the fact that when motional averaging is included, NMR data has only limited capacity to distinguish between a single fully folded peptide conformation and various mixtures of folded and unfolded conformations. Proteins 1999;36:542–555.**

Key words: molecular dynamics; computer simulation; NMR; NOE distance; ^3J-coupling constant; chemical shift; peptides

INTRODUCTION

A long standing problem in the analysis of structure data extracted from nuclear magnetic resonance (NMR) experiments is incorporation of motional averaging.[1–10] The NMR signal reflects primarily short range through-bond interactions (J-coupling constants), short range through-space interactions via the nuclear Overhauser effect (NOE's) or local perturbations to electronic shielding (chemical shifts). This makes the NMR signal sensitive to local conformation, and NMR remains the only experimental technique that can be used to determine the structure of a peptide or protein in solution at atomic resolution.[11] The difficulty when using NMR as a structural tool is that the NMR signal depends in a highly non-linear fashion on the local spatial arrangement of atoms. In addition, in solution, molecules are not constrained by a regular environment but are free to sample a range of conformations due to thermal motion. This means that although the NMR signal is dependent on the relative positions of individual atoms and the fluctuations in these positions, it is far from trivial to invert the observed data and deduce the interactions which gave rise to the original signal.[12,13] This is further complicated by the fact that the NMR effect itself is weak. That is, even at high field there is only a slight excess of aligned spins. Experiments depend on the combined signal of a large collection of individual molecules accumulated on a millisecond time scale. In short, NMR data represents a complex time and ensemble average of a non-linear property. Based on the NMR signal alone it is not possible to solve for the spatial arrangements of atoms directly. To make the problem tractable a motional model must be introduced. The simplest and most common approach is to assume rapid, low amplitude internal motion, the frequency of which is widely separated from the overall tumbling of the molecule. Essentially, the molecule is treated as a quasi-rigid rotor. Within this assumption inter-proton distances can be inferred from NOE buildup rates and ^3J-coupling constants can be used to infer torsional dihedral angles from empirical relationships such as the Karplus curve.[11,14] Based on this data the structures of proteins and many peptides can be solved. It is clear, however, that such a rigid rotor model, especially for peptides but perhaps also for some proteins, is unrealistic.[15] There have, therefore, been repeated attempts to incorporate more detailed models of the atomic motion in the analysis of NMR data, in order to better understand how the measured quantity can be related to actual atomic motions in solution. Such investigations have been based either on analytical models of the atomic motion,[8] or on atomic level molecular dynamics (MD) simulations.[2–7,9,10] Karplus and co-workers, in particular, studied the effect of motions in the picosecond time scale on the calculation of vicinal coupling constants, ring current shifts, and NOE's, concluding such motions were of limited significance.[2–4] More recent studies suggest that motion on a nanosecond timescale can result in errors in interatomic distances of up to 50%.[7,9] Motions which influence the NMR occur, however, on a timescale extending to milliseconds. To truly investigate how motional averaging affects the calculation of NMR properties it is

Grant sponsor: Schweizerischer Nationalfonds; Grant number: Project 21–41875.94.
Alan E. Mark's present address is Department of Biophysical Chemistry, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands.
*Correspondence to: Dr Alan E. Mark, Laboratory for Physical Chemistry, Swiss Federal Institute of Technology (ETH), ETH-Zentrum, CH-8092 Zürich, Switzerland. E-mail: mark@igc.phys.chem.ethz.ch
Received 22 February 1999; Accepted 30 April 1999

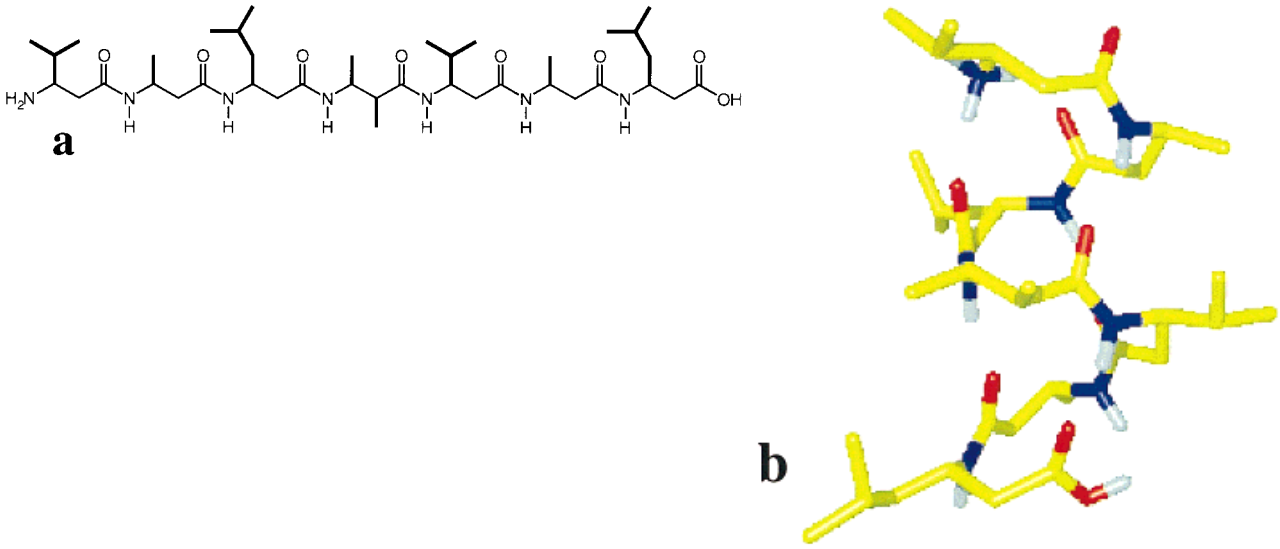

Fig. 1. **a)** Structural formula of the β-heptapeptide studied (H-β-HVal-β-HAla-β-HLeu-(*S*,*S*)-β-HAla(αMe)-β-HVal-β-HAla-β-HLeu-OH). In the simulations both end groups were protonated in line with experimental data. **b)** Model structure derived from NMR data at 298 K.[31]

necessary that the simulations not only accurately reproduce the appropriate NMR data but that either the simulations extend to a relevant timescale or a significant fraction of the phase space available to the molecule is sampled. Until recently such extensive sampling in simulations has not been possible and in most studies the degree of atomic motion has been severely underestimated.[16]

In this study we have used four 50-ns MD simulations of a β-heptapeptide in methanol.[17] These simulations, which cover a range of temperatures (298 K, 340 K, 350 K, and 360 K), were the first in which it was possible to produce a temperature-dependent equilibrium between folded and unfolded conformations of a peptide in solution. For the current investigation the system has a number of advantages. The β-heptapeptide simulated (H-β-HVal-β-HAla-β-HLeu-(*S*,*S*)- β- HAla(αMe)- β-HVal-β-HAla-β-HLeu-OH), which is shown in Figure 1, forms a stable (*M* or left-handed)-3_1-helix in methanol at 298 K. The secondary structure propensity of this peptide in methanol is greater than that for similar sized α-peptides in water. It has also been previously demonstrated that, using the GROMOS96 43A1 force field,[18] simulations of this β-heptapeptide in methanol satisfy all available experimental NMR data without the inclusion of artificial restraints.[19] Finally, although methanol is a strongly hydrogen bonding solvent and the range of conformations sampled is extensive, methanol has a lower density and lower diffusion rate than water, making it not only computationally less expensive to simulate but also facilitating conformational transitions. The simulations have been used to examine how the extent of sampling and frequency of sampling affects the estimation of NOE intensities, vicinal ^3J-coupling constants, and chemical shifts. In particular we have attempted to address the question of to what extent can mixtures of folded and unfolded species be distinguished based on commonly available NMR data.

MATERIALS AND METHODS

Molecular Dynamics Simulations

The molecular dynamics simulations on which this work is based have been described previously.[17] Only a summary is presented here. Four 50-ns MD simulations were performed using the GROMOS96 package of programs in conjunction with the GROMOS96 43A1 force field.[18] The dynamics of the β-heptapeptide (Fig.1) in methanol were studied at a series of temperatures, 298 K, 340 K, 350 K and 360 K, at 1 atm pressure and with periodic boundary conditions. The temperature and pressure were maintained by weak coupling to an external bath. The initial structure of the peptide for the simulations at 298 K, 340 K, and 350 K was the 3_1- helical fold shown in Figure 1b. The system contained the β-heptapeptide and 962 methanol molecules in a rectangular box.[17] For the simulation at 360 K the peptide was initially fully extended (all backbone dihedral angles set to 180°), and the system contained the β-heptapeptide and 1,778 methanol molecules in a truncated octahedron.[17] A twin-range cut-off of 0.8 nm/1.4 nm was used for all non-bonded interactions. The shortest distance peptide-wall was initially 1.4 nm. In all simulations the (periodic) box was sufficiently large that a totally extended conformation of the β-heptapeptide would not span its shortest axis. Configurations of the system were stored every 0.5 ps.

Inter-Proton Distance and ^3J-Coupling Constant Calculations

In the GROMOS96 43A1 force field aliphatic hydrogen atoms are treated as united atoms together with the

carbon atom to which they are attached.[18] Inter-proton distances and ^3J-coupling constants involving aliphatic hydrogens were thus calculated by defining virtual (for CH_1 and pro-chiral CH_2) and pseudo (for CH_3) atomic positions for these hydrogens at the time of analysis.[18]

Average inter-proton distances were calculated from the simulation as $<r^{-6}>^{-1/6}$, where r is the inter-proton distance for a given structure from the simulation. $<r^{-3}>^{-1/3}$ averages are given, in some cases, for comparison (see Discussion). Average inter-proton distance violations were calculated as $<r^{-6}>^{-1/6} - r_{exp}$ for the 42 r_{exp} average inter-proton distances inferred from the NOE intensities observed in the ROESY NMR spectrum measured at 298 K (Table I). Negative values of $<r^{-6}>^{-1/6} - r_{exp}$ cannot be considered violations, given the upper-bound nature of the experimental average inter-proton distances. The average over all 42 average inter-proton distance violations was also calculated, with only positive values of $<r^{-6}>^{-1/6} - r_{exp}$ being included in the sum. ^3J-coupling constants were calculated from the simulation using the Karplus relation,[14] $^3J(\mathrm{H,H}) = A\cos^2\theta + B\cos\theta + C$, where A, B, and C were chosen equal to 6.4 Hz, −1.4 Hz, and 1.9 Hz, respectively, for the calculation of ^3J(HN,HC),[20] and equal to 9.5 Hz, −1.6 Hz, and 1.8 Hz, respectively, for the calculation of ^3J(HC,HC).[21] Average ^3J-coupling constant violations were calculated as $<^3J> - {}^3J_{exp}$ for the 21 $^3J_{exp}$ average ^3J-coupling constants measured experimentally at 298 K (Table II). The average over all 21 average ^3J-coupling constant violations was also calculated, taking the absolute value of the individual average violations $<^3J> - {}^3J_{exp}$.

The averages were calculated for different sets of structures from each of the simulations (see Results). Some of these sets are related to a clustering of structures from the simulation. Results from this cluster analysis have been reported previously.[22] Only a brief description of the clustering algorithm and of the results is presented here. To find clusters of similar structures in the trajectory, the atom-positional root-mean-square deviation (RMSD) between all pairs of structures was determined for a pool of 5,000 structures taken at 0.01 ns intervals from the trajectory. For each structure the number of other structures for which the backbone atom-positional RMSD was less than or equal 0.1 nm (residues 2–6) (neighbor structures) was determined. The structure with the highest number of neighbors was taken as the center of a cluster, and formed together with all its neighbors a (first) cluster of similar structures. The structures forming this cluster were thereafter eliminated from the pool of structures. The process was repeated until the pool of structures was empty. In this way, a series of non-overlapping clusters of similar structures was obtained. At each of the four temperatures studied, 298 K, 340 K, 350 K, and 360 K, the cluster number 1 corresponded to the folded (left-handed 3_1-helical) conformation of the β-heptapeptide (Figure 1b), and incorporated 97%, 50%, 39%, and 25%, respectively, of the ensembles of 5,000 structures.[22] In total, nine clusters were found at 298 K, 158 clusters were found at 340 K, 137 clusters were found at 350 K, and 219 clusters were found at 360 K.

The effect of changing the relative weights of the folded (cluster 1) and unfolded (clusters 2 to M, M being the total number of clusters) conformations on the averages over all average inter-proton distance violations and over all average ^3J-coupling constant violations, has been studied. The weights of the M clusters have been changed by modifying the number of structures (members) of each cluster according to: $N_1' = N_1\, w_f$, $N_2' = N_2\, w_u$, $N_3' = N_3\, w_u$, ... , $N_M' = N_M\, w_u$, where $0 \le w_f \le N/N_1$ and $w_u = (N - N_1\, w_f)/(N - N_1)$, and with N being the total number of structures, N_m the number of members of cluster m in the simulation, N_m' the modified number of members of cluster m, w_f the weight factor for cluster 1, and w_u the weight factor for clusters 2 to M. N is a constant for any value of w_f, i.e., $N = N_1 + N_2 + N_3 + ... + N_M = N_1' + N_2' + N_3' + ... + N_M'$. For $w_f = 0$ the ensemble of N structures is completely constituted by unfolded conformations (0% weight of cluster 1); for $w_f = w_u = 1$ the simulation weights of each cluster are obtained; for $w_f = N/N_1$ the ensemble of N structures consists only of the folded conformation (100% weight of cluster 1).

Chemical Shift Calculations

Chemical shifts for the amide nitrogen, α-carbon, and carbonyl carbon atoms of the peptide backbone were calculated using a method recently proposed by Patchkovskii and Thiel.[23,24] In this approach chemical shifts are calculated using a semi-empirical MNDO[25] (modified neglect of diatomic overlap) model specifically parameterized to reproduce experimental NMR data of organic molecules, thereby overcoming deficiencies in earlier approaches based on semi-empirical theory.[23,24] In the present study the chemical shifts were calculated in vacuo for structures of the β-heptapeptide selected from the trajectory. Both of the approaches proposed[23,24] were applied in conjunction with this system. In the first approach (A) nine orbital exponents and resonance parameters have been adjusted to reproduce experimental liquid-phase chemical-shift data for 299 small organic and inorganic molecules.[23,24] In the second approach (B) an expanded set of 16 parameters has been fitted to reproduce experimental reference data for 97 small molecules in the liquid and in the gas phase.[23] Using method B the root-mean-square (RMS) error for the chemical shifts of the C (13.6 ppm) and N (39.6 ppm) atoms is less than 5% of the experimental chemical shift range for the reference molecules[23] (346 ppm for C and 933 ppm for N); the corresponding RMS errors are somewhat higher for method A.[23] We note, however, that the observed chemical shift range of C^{13} and N^{15} in amino acids in proteins is smaller (∼200 ppm for C and ∼150 ppm for N) and the dependence of the chemical shift on conformation or environment is much smaller again.[26] Proton chemical shifts are reproduced less reliably and have thus not been considered in this study.

Test calculations for a set of trial structures consisting of the central members of all clusters with a minimum of 20 members at 298 K, 340 K, 350 K, and 360 K showed no significant difference between methods A and B. As ex-

TABLE I. Average Inter-Proton Distances Inferred From Experimental NOE Intensities Observed in the ROESY NMR Spectrum Measured at 298 K, and Average Inter-Proton Distance Violations from the MD Simulations at 298 K, 340 K, 350 K, and 360 K†

		298 K		340 K			350 K	360 K
NOE number	H atoms	r_{exp} (nm)	$\langle v \rangle_{trj}$ (nm)	$\langle v \rangle_{trj}$ (nm)	$\langle v \rangle_{cls}$ (nm)	$\langle v \rangle_{cnt}$ (nm)	$\langle v \rangle_{trj}$ (nm)	$\langle v \rangle_{trj}$ (nm)
1 NH(1)	H-C$^{\beta}$(1)	0.28	−0.05	−0.04	−0.04	−0.02	−0.04	−0.05
2 H-C$^{\beta}$(1)	H^{ax}-C$^{\alpha}$(1)	0.29	−0.02	−0.02	−0.02	−0.01	−0.02	−0.04
3 H-C$^{\beta}$(1)	H^{eq}-C$^{\alpha}$(1)	0.30	−0.05	−0.05	−0.05	−0.05	−0.05	−0.08
4 NH(2)	H^{ax}-C$^{\alpha}$(1)	0.24	0.00	0.01	0.01	0.04	0.01	0.00
5 NH(2)	H^{eq}-C$^{\alpha}$(1)	0.29	−0.05	−0.05	−0.05	−0.07	−0.05	−0.05
6 NH(2)	H-C$^{\beta}$(2)	0.33	−0.05	−0.05	−0.05	−0.05	−0.05	−0.10
7 NH(2)	H-C$^{\beta}$(4)	0.35	0.02	0.06	0.05	0.09	0.07	0.10
8 NH(2)	H-C$^{\beta}$(5)	0.33	−0.02	0.01	0.01	0.01	0.02	0.04
9 H-C$^{\beta}$(2)	H-C$^{\gamma}$(2)	0.22	−0.08	−0.08	−0.08	−0.08	−0.08	−0.08
10 H-C$^{\beta}$(2)	H^{eq}-C$^{\alpha}$(2)	0.23	0.01	0.01	0.01	0.01	0.01	0.00
11 NH(3)	H^{ax}-C$^{\alpha}$(2)	0.22	0.00	0.01	0.01	−0.01	0.00	0.01
12 NH(3)	H^{eq}-C$^{\alpha}$(2)	0.31	−0.03	−0.05	−0.05	−0.06	−0.04	−0.07
13 NH(3)	H-C$^{\beta}$(3)	0.31	−0.03	−0.03	−0.03	−0.02	−0.03	−0.06
14 NH(3)	H^{ax}-C$^{\alpha}$(3)	0.26	−0.01	−0.01	−0.01	−0.01	−0.01	−0.02
15 NH(3)	NH(4)	0.38	−0.01	−0.03	−0.03	−0.02	0.00	−0.02
16 NH(3)	H-C$^{\beta}$(5)	0.34	0.01	0.04	0.04	0.05	0.05	0.08
17 NH(3)	H-C$^{\beta}$(6)	0.32	−0.05	−0.04	−0.04	−0.02	−0.03	−0.02
18 H-C$^{\beta}$(3)	H-C$^{\delta}$(3)	0.30	−0.04	−0.04	−0.04	−0.06	−0.05	−0.06
19 NH(4)	H^{ax}-C$^{\alpha}$(3)	0.33	−0.11	−0.10	−0.10	−0.10	−0.09	−0.10
20 NH(4)	H^{eq}-C$^{\alpha}$(3)	0.28	0.01	−0.01	−0.01	−0.01	−0.04	−0.04
21 NH(4)	H-C$^{\beta}$(4)	0.29	0.00	−0.01	−0.01	0.00	0.00	−0.05
22 NH(4)	H-C$^{\gamma}$(4)	0.30	−0.11	−0.11	−0.11	−0.12	−0.11	−0.11
23 NH(4)	H-C$^{\beta}$(6)	0.32	0.02	0.06	0.06	0.12	0.07	0.05
24 NH(4)	H-C$^{\beta}$(7)	0.37	−0.07	−0.06	−0.06	0.05	−0.04	−0.02
25 H-C$^{\beta}$(4)	H^{ax}-C$^{\alpha}$(1)	0.26	−0.02	0.01	0.01	0.11	0.02	0.04
26 NH(5)	NH(4)	0.37	0.00	0.00	0.00	0.03	−0.03	−0.01
27 NH(5)	H^{ax}-C$^{\alpha}$(4)	0.22	0.00	0.00	0.00	0.00	0.00	−0.01
28 NH(5)	Me-C$^{\alpha}$(4)	0.35	−0.10	−0.10	−0.10	−0.10	−0.09	−0.10
29 NH(5)	H-C$^{\beta}$(5)	0.35	−0.07	−0.07	−0.07	−0.06	−0.07	−0.11
30 NH(5)	H^{ax}-C$^{\alpha}$(5)	0.25	−0.01	−0.01	−0.01	−0.02	−0.01	−0.01
31 NH(5)	NH(6)	0.35	0.05	0.03	0.04	0.10	0.06	0.03
32 H-C$^{\beta}$(5)	H^{ax}-C$^{\alpha}$(2)	0.23	0.03	0.05	0.05	0.03	0.06	0.09
33 H-C$^{\beta}$(5)	H-C$^{\gamma}$(5)	0.26	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
34 H-C$^{\beta}$(5)	H^{eq}-C$^{\alpha}$(5)	0.25	−0.01	−0.01	−0.01	0.00	0.00	−0.02
35 NH(6)	H-C$^{\beta}$(6)	0.29	−0.01	−0.01	−0.01	−0.01	−0.01	−0.05
36 NH(6)	H^{ax}-C$^{\alpha}$(6)	0.25	0.00	0.00	0.00	0.01	0.00	0.00
37 NH(6)	H^{ax}-C$^{\alpha}$(5)	0.22	0.00	0.01	0.01	0.04	0.02	0.01
38 H-C$^{\beta}$(6)	H^{ax}-C$^{\alpha}$(3)	0.25	−0.03	0.01	0.01	0.08	0.02	0.02
39 H-C$^{\beta}$(6)	H^{eq}-C$^{\alpha}$(6)	0.26	−0.02	−0.02	−0.02	−0.02	−0.01	−0.03
40 NH(7)	H^{ax}-C$^{\alpha}$(6)	0.24	−0.01	0.01	0.01	−0.01	0.00	−0.01
41 NH(7)	H-C$^{\beta}$(7)	0.30	−0.02	−0.02	−0.02	−0.02	−0.02	−0.06
42 NH(7)	H^{ax}-C$^{\alpha}$(7)	0.27	−0.02	−0.01	−0.01	0.00	−0.01	−0.02
$\langle v \rangle_{42}$			0.004	0.008	0.008	0.018	0.010	0.012

†Axial (pro-R) and equatorial (pro-S) C$^{\alpha}$ hydrogens are indicated by the superscripts ax and eq, respectively. Residue numbers are indicated in parentheses. r_{exp} refers to the average inter-proton distance inferred from the NMR experiment at 298 K.[19,31] The inter-proton distance of 0.23 nm from NOE number 19 was misprinted in Daura et al.[19] $\langle v \rangle$ refers to the difference between the average inter-proton distance for a given set of structures from the simulation (weighted as $\langle r^{-6} \rangle^{-1/6}$) and the distance inferred from the NOE intensity (r_{exp}). Negative values of this difference cannot be considered violations, as the experimental data is interpreted in terms of upper bounds only. The subscript trj refers to an average over all 10^5 structures stored (1 per 0.5 ps). The subscript cls refers to an average over all 5,000 structures used for the cluster analysis (1 per 10 ps) (see Materials and Methods). The subscript cnt refers to an average over the central member structures of all clusters, weighted by the number of members of each cluster. The subscript 42 refers to an average over all 42 average violations, with only positive values contributing to the sum.

pected[23] the three-center contributions to the carbon and nitrogen chemical shifts were negligibly small. For these reasons, we only report chemical shifts obtained from method A without any three-center contributions (A2 in the notation of Patchkovskii and Thiel[23]). The geometries were taken from the MD runs (see below) after checking

TABLE II. Average ^{3}J-Coupling Constants Measured at 298 K by NMR, and Average ^{3}J-Coupling Constants Violations from the MD Simulations at 298 K, 340 K, 350 K, and 360 K†

		298 K		340 K			350 K	360 K
^{3}J number	H atoms	$^3J_{exp}$ (Hz)	$\langle v\rangle_{trj}$ (Hz)	$\langle v\rangle_{trj}$ (Hz)	$\langle v\rangle_{cls}$ (Hz)	$\langle v\rangle_{cnt}$ (Hz)	$\langle v\rangle_{trj}$ (Hz)	$\langle v\rangle_{trj}$ (Hz)
1 H-C$^\beta$(1)	H-C$^\gamma$(1)	4.7	0.1	0.3	0.3	4.3	0.3	1.4
2 H-C$^\beta$(1)	H^{ax}-C$^\alpha$(1)	11.5	−0.7	−1.0	−1.0	−0.4	−1.0	−2.8
3 H-C$^\beta$(1)	H^{eq}-C$^\alpha$(1)	2.8	0.6	1.0	1.0	−0.1	0.9	1.2
4 NH(2)	H-C$^\beta$(2)	9.2	0.0	−0.2	−0.2	0.1	−0.2	−0.4
5 H-C$^\beta$(2)	H^{ax}-C$^\alpha$(2)	12.0	0.5	−0.1	−0.1	−0.1	0.0	−0.5
6 H-C$^\beta$(2)	H^{eq}-C$^\alpha$(2)	4.5	−0.7	−0.8	−0.8	−2.0	−1.0	−0.9
7 NH(3)	H-C$^\beta$(3)	9.6	−0.4	−0.6	−0.5	−1.3	−0.6	−0.9
8 H-C$^\beta$(3)	H^{ax}-C$^\alpha$(3)	12.3	0.3	−0.8	−0.8	−0.5	−0.8	−1.1
9 H-C$^\beta$(3)	H^{eq}-C$^\alpha$(3)	4.5	−0.6	−0.6	−0.7	−0.6	−1.0	−0.7
10 NH(4)	H-C$^\beta$(4)	9.3	0.1	−0.1	−0.1	0.0	−0.1	−0.2
11 H-C$^\beta$(4)	H^{ax}-C$^\alpha$(4)	10.8	1.1	−2.0	−2.0	−1.9	−0.4	−3.1
12 NH(5)	H-C$^\beta$(5)	9.6	−0.1	−0.4	−0.4	−0.1	−0.4	−0.5
13 H-C$^\beta$(5)	H-C$^\gamma$(5)	7.0	−1.7	−0.4	−0.4	−2.4	−1.0	−1.1
14 H-C$^\beta$(5)	H^{ax}-C$^\alpha$(5)	12.3	0.2	−1.1	−1.1	−0.9	−1.3	−1.5
15 H-C$^\beta$(5)	H^{eq}-C$^\alpha$(5)	3.9	−0.5	−0.6	−0.6	−1.2	−0.5	−0.1
16 NH(6)	H-C$^\beta$(6)	8.7	0.5	0.3	0.2	0.6	0.2	0.1
17 H-C$^\beta$(6)	H^{ax}-C$^\alpha$(6)	11.6	0.7	−0.8	−0.8	0.3	−1.9	−1.2
18 H-C$^\beta$(6)	H^{eq}-C$^\alpha$(6)	3.8	−0.4	0.1	0.1	−0.8	1.0	0.4
19 NH(7)	H-C$^\beta$(7)	9.5	−0.4	−0.6	−0.6	−2.1	−0.6	−0.7
20 H-C$^\beta$(7)	H^{ax}-C$^\alpha$(7)	10.0	1.4	−0.8	−0.9	−2.7	−1.1	−0.5
21 H-C$^\beta$(7)	H^{eq}-C$^\alpha$(7)	4.5	−1.2	0.2	0.2	3.9	0.4	−0.2
$\langle v\rangle_{21}$			0.58	0.61	0.62	1.25	0.70	0.94

†Axial (pro-R) and equatorial (pro-S) C$^\alpha$ hydrogens are indicated by the superscripts ax and eq, respectively. Residue numbers are indicated in parentheses. $^3J_{exp}$ refers to the average ^{3}J-coupling constant measured experimentally at 298 K.[19] $\langle v\rangle$ refers to the average violation of the experimental ^{3}J-coupling constant, that is, the difference between the average ^{3}J-coupling constant, estimated from the corresponding torsional dihedral angle using the Karplus relation,[14,20,21] for a given set of structures from the simulation and the ^{3}J-coupling constant measured experimentally by NMR at 298 K. The subscript trj refers to an average over all 10^5 structures stored (1 per 0.5 ps). The subscript cls refers to an average over all 5,000 structures used for the cluster analysis (1 per 10 ps) (see Materials and Methods). The subscript cnt refers to an average over the central member structures of all clusters, weighted by the number of members of each cluster. The subscript 21 refers to an average over all 21 average violations, taken in absolute values.

that geometry optimization at the MNDO level of the same set of structures prior to the calculation of chemical shifts did not systematically improve the results.

Even though the accuracy of the special MNDO-NMR parameterizations for carbon and nitrogen chemical shifts is generally similar to that of low-level ab initio or density functional methods,[23] several single-point ab initio calculations were carried out for comparison, at the RHF/6–311G* level (restricted Hartree-Fock/6–311G* basis[27]). Each such chemical shift calculation with the GAUSSIAN94 program[28,29] was about 10^4 times more expensive than an MNDO-A2 calculation with the MNDO97 program.[30]

RESULTS

Violations of Inter-Proton Distances and ^{3}J-Coupling Constants

Figure 2 illustrates the variation in the violations of the inter-proton distances inferred from the experimental NOE intensities observed in the ROESY NMR spectrum at 298 K, together with a comparison between the calculated ^3J-coupling constants for four structures selected from the trajectory at 340 K (Fig. 3) and the ^3J-coupling constants measured by NMR at 298 K. The structures selected include the structure with the minimum backbone atom-positional RMSD from the model 3_1-helix, a hairpin-like structure, the structure with the minimum radius of gyration, and the structure with the maximum radius of gyration. The distance violations and ^3J- coupling constants averaged over the entire trajectory (1 structure per 0.5 ps) at 340 K are also shown in Figure 2 for comparison. From the 42 NOE's observed experimentally, nine are especially indicative of an M-3_1-helical structure: the NH/HC$^\beta$ with NOE sequence numbers 7, 8, 16, 17, 23, and 24, and the HC$^\beta$/H^{ax}C$^\alpha$ with NOE sequence numbers 25, 32, and 38 (Table I). Obviously, the inter-proton distances and the ^3J-coupling constants depend on the conformation. The NOE-distance violations for the structure with the minimum backbone atom-positional RMSD from the model 3_1-helix are, except for NOE number 42, below 0.06 nm. For each of the other structures selected, the biggest violations range between 0.5 and 1.0 nm. When averaging over the entire trajectory at 340 K (using $<r^{-6}>^{-1/6}$) the maximum violation is 0.06 nm. The ^3J-coupling constants calculated for each of the selected structures show comparable deviations from those measured experimentally at

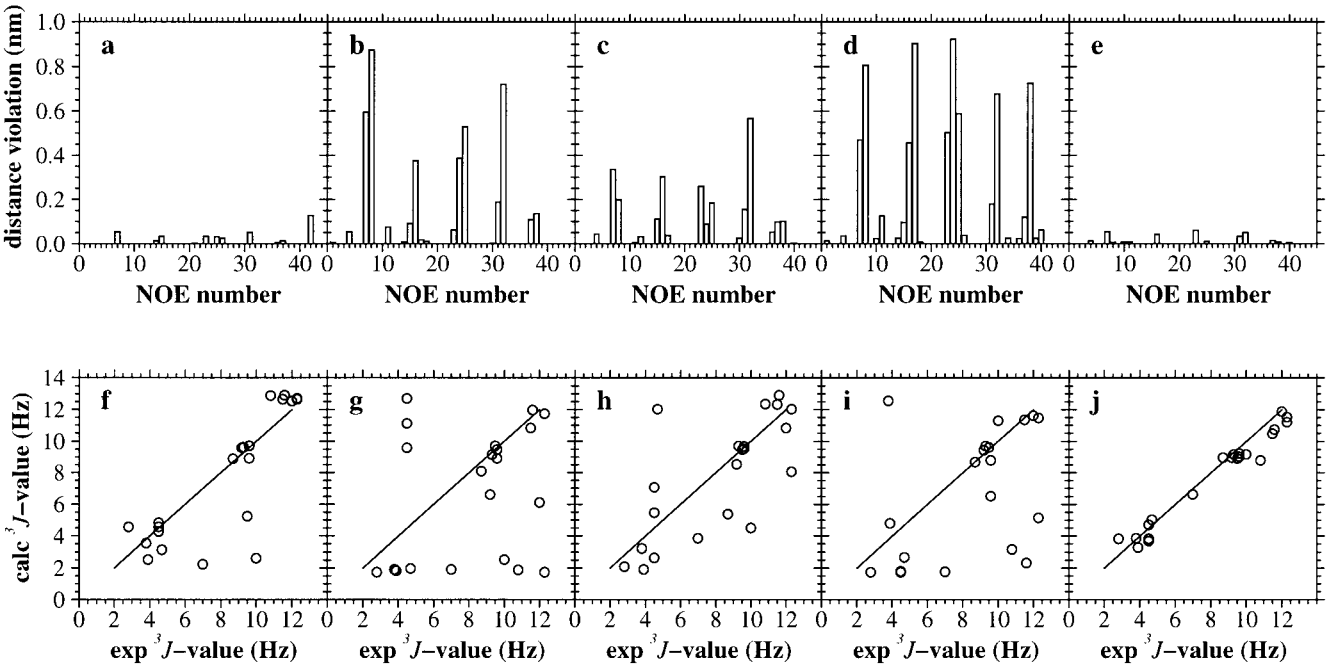

Fig. 2. **Panels a–e**: violations of average inter-proton distances inferred from 42 experimental nuclear Overhauser effect (NOE) intensities observed in the ROESY NMR spectrum measured at 298 K[19,31] by the corresponding inter-proton distances in four selected structures from the simulation at 340 K (panels a–d, for structures a–d, respectively, in Figure 3), and by the corresponding inter-proton distances averaged (using $<r^{-6}>^{-1/6}$) over all 10^5 recorded structures (1 per 0.5 ps) from the simulation at 340 K (panel e); **a)** the structure with the minimum atom-positional root-mean-square deviation (RMSD) from the modelled 3_1-helical structure for the backbone atoms of residues 1 to 7; **b)** a hairpin-like structure; **c)** the structure with the minimum radius of gyration; **d)** the structure with the maximum radius of gyration. **Panels f–j**: comparison of 21 experimental average ^3J-coupling constants measured at 298 K[19] with the corresponding calculated ^3J-coupling constants for the same four selected structures (panels f-i, structures a-d, respectively, in Figure 3), and with the corresponding calculated ^3J-coupling constants averaged over all recorded structures from the simulation at 340 K (panel j).

298 K, although there is in general a better correlation between the calculated values for the structure with the minimum backbone atom-positional RMSD from the model 3_1-helix and the experimental ones. As the structures were taken directly from the trajectory they show, as expected, a thermal distribution of dihedral angles. When averaging the ^3J-coupling constants over the entire trajectory there is, however, a clear correlation between the calculated and the experimental values. Figures 2 and 3 show that although individual structures sampled in the simulation at 340 K vary widely, their average structural properties conform closely to the data measured experimentally at 298 K. A detailed comparison between the calculated and experimental inter-proton distances and ^3J-coupling constants at the experimental temperature of 298 K was presented in earlier work as part of a validation of the simulation methodology.[19]

Figure 4, together with Tables I and II, illustrates the effect of changing the ensemble of conformations sampled, in particular the proportion of the 3_1-helix in the ensemble, on the correlation between the calculated inter-proton distances and ^3J-coupling constants and those derived from experiment at 298 K. Figure 4a–d shows the NOE-distance violations averaged (using $<r^{-6}>^{-1/6}$) over the 50-ns trajectories (with 1 structure per 0.5 ps) at 298 K, 340 K, 350 K, and 360 K, respectively (see also Table I). The scale differs by a factor of 10 with respect to Figure 2. As has been discussed previously, the ensemble of structures sampled changes significantly with temperature, with the 3_1-helix constituting approximately 97%, 50%, 39%, and 25% of the ensemble at 298 K, 340 K, 350 K, and 360 K, respectively. Nevertheless, even at 360 K the maximum NOE-distance violation is still less than 0.1 nm. In addition, a comparison between the average ^3J-coupling constants calculated at each of the temperatures investigated and the experimental ^3J-coupling constants at 298 K is shown in Figure 4e–h. In general, the calculated average ^3J-coupling constants tend to decrease with increasing temperature (see also Table II), but the correlation between the calculated and experimental (at 298 K) values is remarkable even for the simulation at 360 K (Fig. 4h), where the 3_1-helical fold represents only 25% of the ensemble. Indeed, the NH/HC$^\beta$ ^3J-coupling constants measured experimentally by NMR decrease only slightly with increasing temperature, in good agreement with the NH/HC$^\beta$ ^3J-coupling constants calculated from the simulations as shown in Table III. This small drift of the NH/HC$^\beta$ ^3J-coupling constants observed experimentally with increasing temperature is not necessarily indicative of high thermal (up to 353 K) stability. Table I lists, for detailed comparison, the 42 inter-proton distances derived from the NMR data at 298 K (r_{exp}) and the difference between the corresponding average distances in the simulations ($<r^{-6}>^{-1/6}$) and r_{exp}. For the trajectory at 340 K, $<r^{-6}>^{-1/6}$ is calculated using three different subsets of structures the

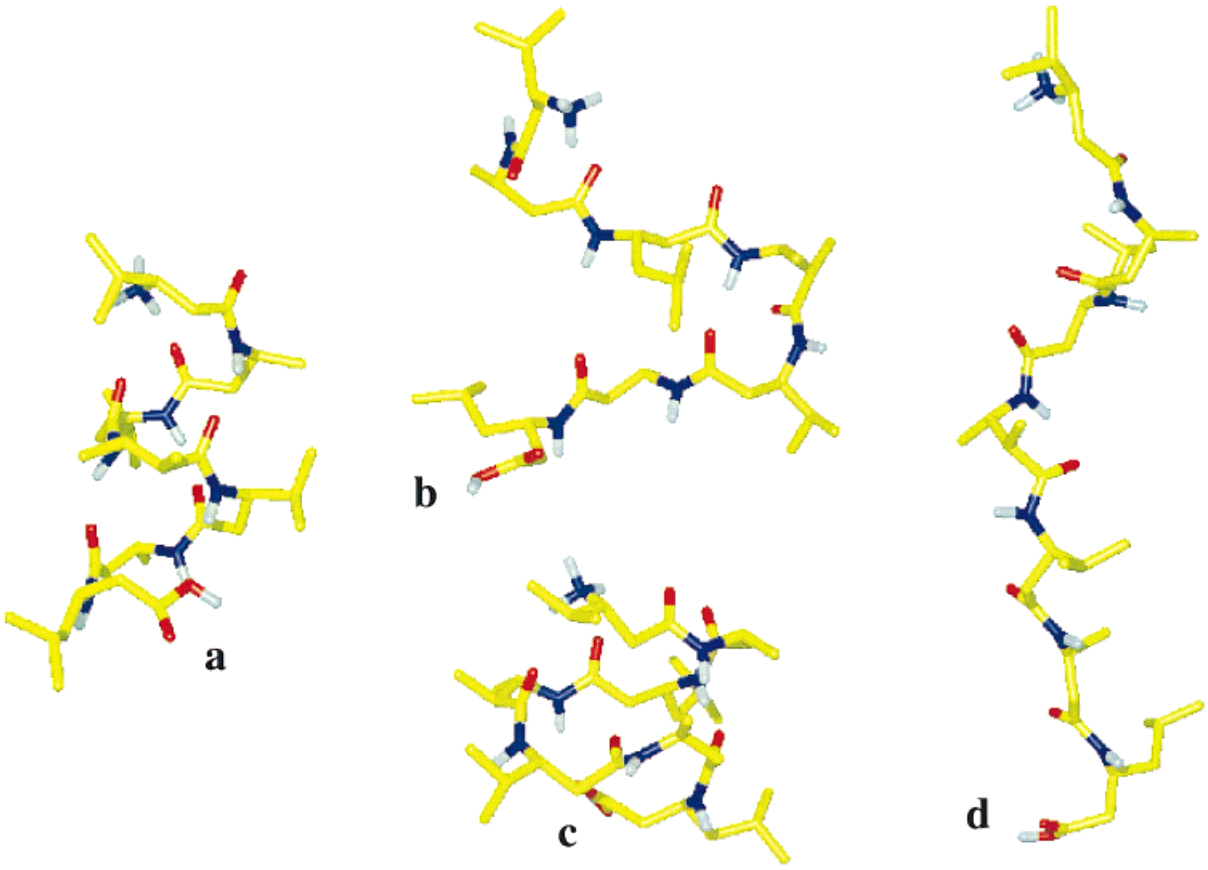

Fig. 3. Selected structures from a set of 5,000 taken at 10-ps intervals from the simulation at 340 K: **a)** the structure with the minimum atom-positional root mean-square-deviation (RMSD) from the modelled 3_1-helical structure (Fig. 1b) for the backbone atoms of residues 1 to 7 (at time = 16.36 ns); **b)** a hairpin-like structure (at time = 23.02 ns); **c)** the structure with the minimum radius of gyration (at time = 44.95 ns); **d)** the structure with the maximum radius of gyration (at time = 4.08 ns).

results of which are listed in columns 6, 7, and 8 of Table I. In column 6 ($<v>_{trj}$) the average runs over 100,000 structures sampled at 0.5 ps intervals in the simulation. In column 7 ($<v>_{cls}$) the average runs over 5,000 structures sampled at 10 ps intervals in the simulation. In column 8 ($<v>_{cnt}$) the average runs over the central member structures of the 158 clusters found at 340 K, with weights corresponding to the number of members of each cluster. The difference between column 6 and column 7 is marginal, indicating that the choice of a 10 ps time interval for the analysis might be appropriate (small loss of structural information in the ensemble). Column 8 gives an indication of the degree to which the entire ensemble can be represented by a small number of structures spanning the same volume of conformational space. Table II lists the ^3J-coupling constants observed by NMR at 298 K together with the average violations of these values in the four simulations. The results shown in each of the columns in Table II are analogous to those shown in Table I.

Effect of Simulation Time on the Average Violation

Figure 5 shows the effect of increasing the length of the simulation on the average distance violation and the average ^3J-coupling constant violation. The solid line in Figure 5a corresponds to an average over all 42 average distance violations, with the average distances calculated as $<r^{-6}>^{-1/6}$. The dashed line in Figure 5a also corresponds to an average over the individual average distance violations, but with the average distances calculated as $<r^{-3}>^{-1/3}$. Note, in this simulation the starting structure satisfies the experimental NOE distances. As the simulation progresses the average violation increases, as alternative conformations, including a range of unfolded conformations, are sampled. The average violation falls again as the trajectory progressively reflects a better approximation of the ensemble of possible conformations with appropriate Boltzmann weights. Figure 5b shows the average over all 21 average ^3J-coupling constant violations as a function of simulation time. The progression of this curve is comparable to that of the average distance violation given in Figure 5a.

Effect of Sampling Frequency on the Average Violation

Figure 6 shows the effect of the time resolution (time between successive structures) used in the analysis on the average distance violation (Fig. 6a) and the average ^3J-

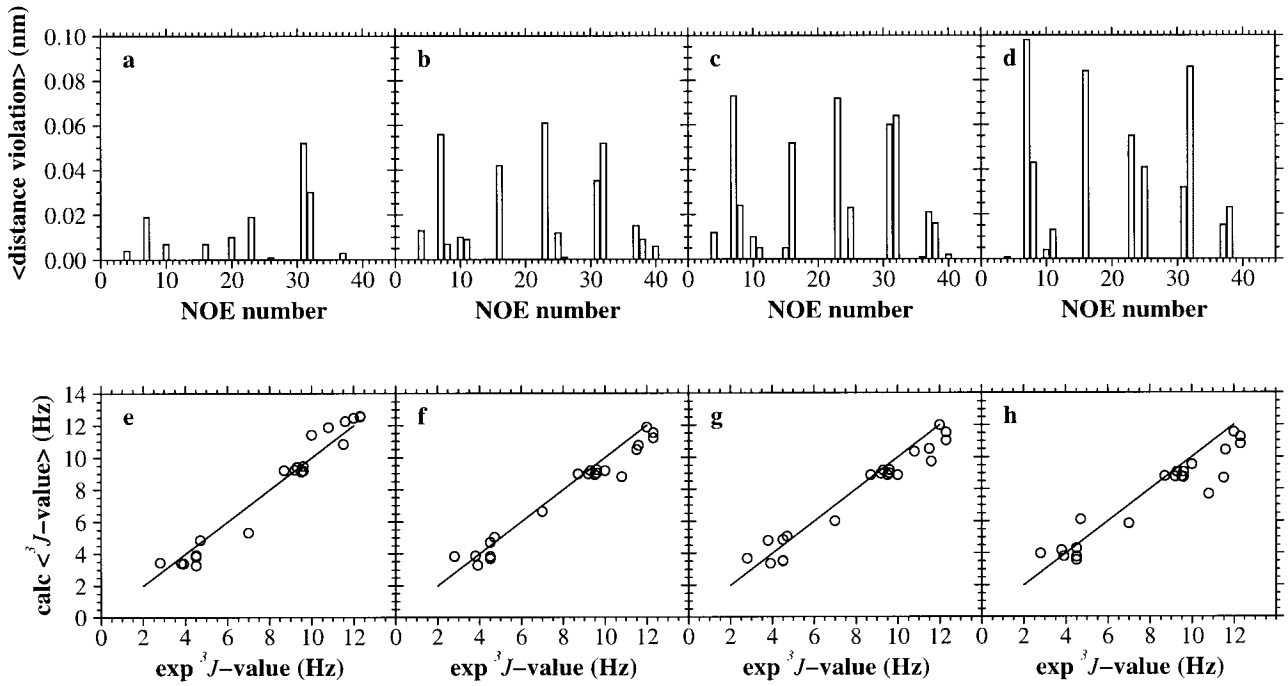

Fig. 4. Violations of the 42 average inter-proton distances inferred from the NMR data at 298 K[19,31] by the corresponding inter-proton distances averaged (using $< r^{-6}>^{-1/6}$) over all 10^5 recorded structures (1 per 0.5 ps) from the 50-ns simulations at: **a)** 298 K; **b)** 340 K; **c)** 350 K; **d)** 360 K. Comparison of the 21 experimental average ^3J-coupling constants measured at 298 K[19] with the corresponding calculated ^3J-coupling constants averaged over all recorded structures from the 50-ns simulation at: a) 298 K; b) 340 K; c) 350 K; d) 360 K.

TABLE III. Average ^{3}J-Coupling Constants Measured by NMR and Calculated From the MD Simulations, at Different Temperatures†

		298 K		343 K	340 K	353 K	350 K	360 K
^{3}J number	H atoms	$^3J_{exp}$ (Hz)	$^3J_{calc}$ (Hz)	$^3J_{exp}$ (Hz)	$^3J_{calc}$ (Hz)	$^3J_{exp}$ (Hz)	$^3J_{calc}$ (Hz)	$^3J_{calc}$ (Hz)
4 NH(2)	H-C$^\beta$(2)	9.1	9.2	8.9	9.0	8.4	9.0	8.8
7 NH(3)	H-C$^\beta$(3)	9.4	9.2	8.9	9.0	9.0	9.0	8.7
10 NH(4)	H-C$^\beta$(4)	9.0	9.4	8.6	9.2	8.5	9.2	9.0
12 NH(5)	H-C$^\beta$(5)	9.4	9.5	9.1	9.2	9.0	9.2	9.1
16 NH(6)	H-C$^\beta$(6)	8.6	9.2	8.2	9.0	7.8	8.9	8.8
19 NH(7)	H-C$^\beta$(7)	9.4	9.1	9.1	8.9	8.7	8.9	8.8

†^{3}J-sequence numbers are as in Table II. Residue numbers are indicated in parentheses. $^3J_{exp}$ refers to the average ^{3}J-coupling constant measured experimentally. $^3J_{calc}$ refers to the average ^{3}J-coupling constant from simulation, estimated from the corresponding torsional dihedral angle using the Karplus relation.[14,20,21] The experimental temperatures were 298 K, 343 K, and 353 K.[32] The simulation averages were calculated over 100,000 structures taken at 0.5 ps intervals. Note, $^3J_{exp}$ at 298 K are not equal to those shown in Table II. These correspond to independent measures of the NMR ^{3}J-coupling constants, and thus give an idea of the experimental uncertainty.

coupling constant violation (Fig. 6b). The solid and dashed lines in Figure 6a correspond to average distance violations, with average distances calculated as $<r^{-6}>^{-1/6}$ and $<r^{-3}>^{-1/3}$, respectively. It is shown that the average violation is in both Figure 6a and 6b remarkably insensitive to decreasing sampling frequency (i.e., decreasing time resolution, from left to right) for the range shown. This is due in part to the constant sampling of the 3_1-helical fold.[17] In the simulation at 360 K, for example, where the initial structure was linear and the sampling of the 3_1-helical fold was more sparse, there is a slight increase in average violation with decreasing time resolution (results not shown). Clearly, the average violation alone cannot be used to assess how representative of the whole ensemble is a subset of structures: while the average violation for the subset of 100 structures taken at 0.5-ns time intervals might be close to the average violation for the subset of 5,000 structures taken at 0.01-ns, the small subset is certainly not representative of the variety of conformations present in the larger one, even if in this particular case the weight of the 3_1-helical conformation might be similar in both subsets.

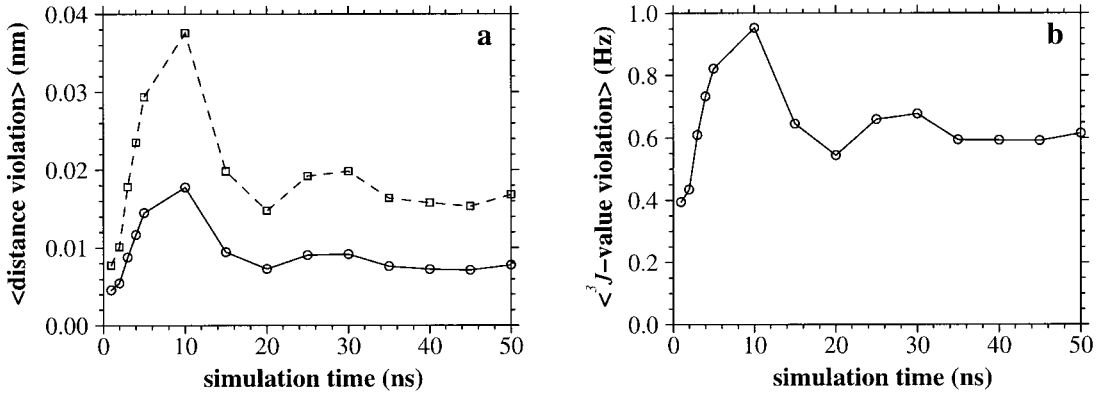

Fig. 5. For the simulation of the β-heptapeptide in methanol at 340 K: **a)** The average over all 42 average violations of the average inter-proton distances inferred from the experimental NOE intensities observed in the ROESY NMR spectrum measured at 298 K[19,31] as a function of simulation time. The solid line corresponds to $< r^{-6}>^{-1/6}$ averaging of the calculated inter-proton distances. The dashed line corresponds to $< r^{-3}>^{-1/3}$ averaging of the calculated inter-proton distances. Note that the value at time 50-ns corresponds to $<v>_{42}$ (column $<v_{trj}>$ at 340 K) in Table I. **b)** The average over all 21 average violations of the experimental average ^3J-coupling constants measured at 298 K[19] as a function of simulation time. Note that the value at time 50-ns corresponds to $<v>_{21}$ (column $<v>_{trj}$ at 340 K) in Table II.

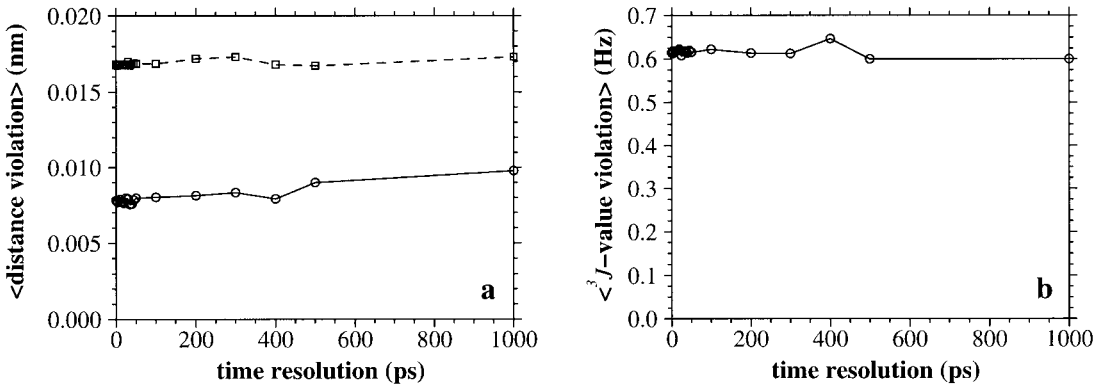

Fig. 6. For the simulation at 340 K: **a)** The average over all 42 average violations of the experimental average inter-proton distances measured at 298 K[19,31] as a function of time resolution, i.e., time between successive structures in the analysis. The solid line corresponds to $< r^{-6}>^{-1/6}$ averaging of the calculated inter-proton distances. The dashed line corresponds to $< r^{-3}>^{-1/3}$ averaging of the calculated inter-proton distances. **b)** The average over all 21 average violations of the experimental average ^3J-coupling constants measured at 298 K[19] as a function of time resolution.

Effect of the Number of Clusters Considered on the Average Violation

Figure 7 shows the effect of the number of clusters included in the averages over structures (see Materials and Methods) on the average distance violation (Fig. 7a) and the average ^3J-coupling constant violation (Fig. 7b). The solid and dashed lines in Figure 7a correspond to average distance violations, with average distances calculated as $<r^{-6}>^{-1/6}$ and $<r^{-3}>^{-1/3}$, respectively. The average distance violation has its minimum when only cluster 1 is considered, and increases as more (consecutive) clusters are added to the average, reaching a plateau approximately after the first 50 clusters have been included in the average. For the first 30 clusters, which include 79% of the ensemble used in the cluster analysis,[22] the average distance violation is already very close to the one for the entire ensemble ($<r^{-6}>^{-1/6}$ weighting). The average ^3J-coupling constant violation has its minimum when only clusters 1 and 2 are considered, and has otherwise a similar progression to the average distance violation.

Effect of the Percentage Weight Given to Cluster Number 1 on the Average Violation

Figure 8 shows the effect of changing the percentage weight of cluster 1 (see Materials and Methods) on the average distance violation (Fig. 8a) and the average ^3J-coupling constant violation (Fig. 8b). The solid and dashed lines in Figure 8a correspond to average distance violations, with average distances calculated as $<r^{-6}>^{-1/6}$ and $<r^{-3}>^{-1/3}$, respectively. In Figure 8a the maximum and minimum average distance violations correspond, respectively, to a 0% weight of cluster 1 (i.e., an ensemble formed completely of unfolded conformations) and a 100% weight of cluster 1 (i.e., an ensemble formed completely of the folded conformation). The average distance violation corresponding to the original ensemble (50% weight of cluster 1) is very close to the one for an artificial ensemble of folded conformations. In Figure 8b the maximum and minimum average ^3J-coupling violations correspond, respectively, to a 0% weight of cluster 1 and an 88% weight of cluster 1. Note, although the average ^3J-coupling constants are

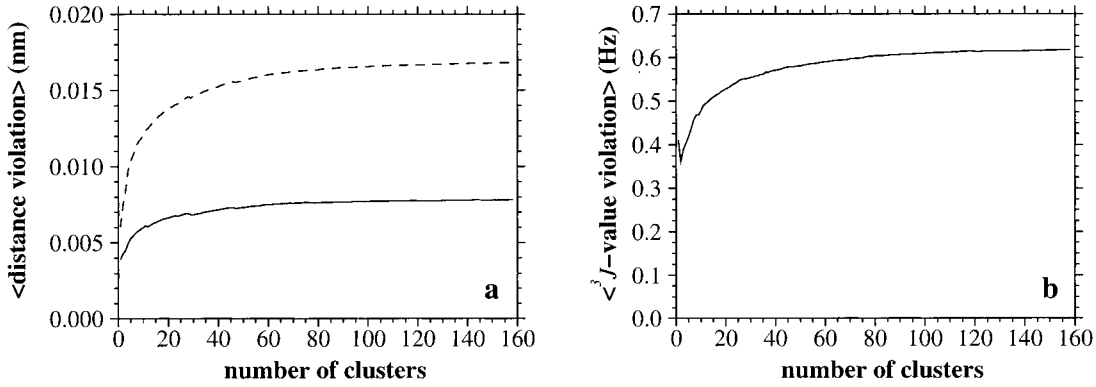

Fig. 7. For the simulation at 340 K: **a)** The average over all 42 average violations of the experimental average inter-proton distances measured at 298 K[19,31] as a function of the number of clusters considered. The solid line corresponds to $< r^{-6} >^{-1/6}$ averaging of the calculated inter-proton distances. The dashed line corresponds to $< r^{-3} >^{-1/3}$ averaging of the calculated inter-proton distances. Note that the value for 158 clusters corresponds to $< v >_{42}$ (column $< v >_{cls}$ at 340 K) in Table I. **b)** The average over all 21 average violations of the experimental average ^3J-coupling constants measured at 298 K[19] as a function of the number of clusters considered. Note that the value for 158 clusters corresponds to $< v >_{21}$ (column $< v >_{cls}$ at 340 K) in Table II.

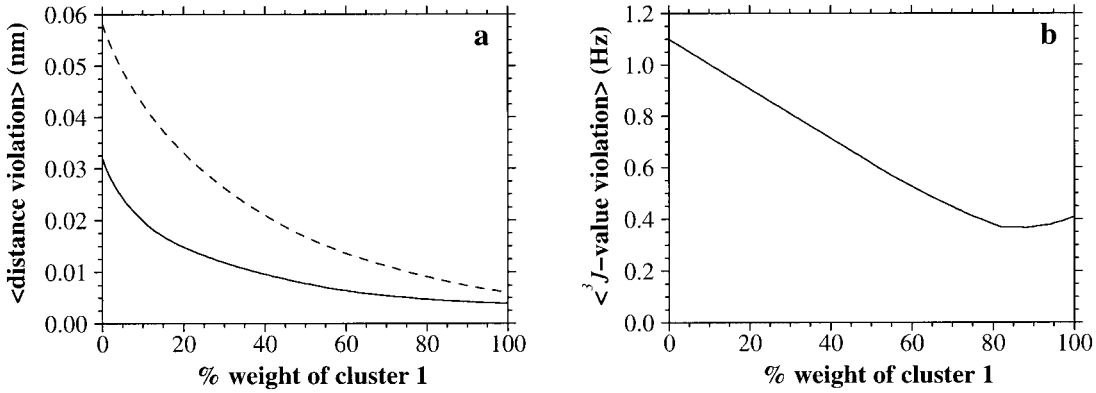

Fig. 8. For the simulation at 340 K: **a)** The average over all 42 average violations of the experimental average inter-proton distances measured at 298 K[19,31] as a function of the percentage weight of cluster 1 (see Materials and Methods). The solid line corresponds to $< r^{-6} >^{-1/6}$ averaging of the calculated inter-proton distances. The dashed line corresponds to $< r^{-3} >^{-1/3}$ averaging of the calculated inter-proton distances. Note that the value at 50% corresponds to $< v >_{42}$ (column $< v >_{cls}$ at 340 K) in Table I. **b)** The average over all 21 average violations of the experimental average ^3J-coupling constants measured at 298 K[19] as a function of the percentage weight of cluster 1. Note that the value at 50% corresponds to $< v_{21} >$ (column $< v >_{cls}$ at 340 K) in Table II.

linear functions of w_f (and hence of the %weight of cluster 1), the average ^3J-coupling constant violation, which runs over the absolute values of the 21 average ^3J-coupling constant violations, is not. Interestingly, Figures 7b and 8b suggest that the average ^3J-coupling constant violation is lowest when some weight is given to certain clusters of unfolded conformations (especially cluster 2).

Chemical Shifts

Figure 9 shows the calculated chemical shifts for the backbone amide nitrogen, α-carbon and carbonyl carbon atoms for selected structures from the trajectory at 340 K. The structures chosen were the same as those used in Figure 2, and are shown in Figure 3. They represent the structure with the minimum backbone atom-positional RMSD from the model 3_1-helix, a hairpin-like structure, the structure with the minimum radius of gyration, and the structure with the maximum radius of gyration. Included in Figure 9 for comparison are also results for the central member structure of cluster 1, an average over the central member structures of all clusters with a minimum of 20 members weighted by the number of members of each cluster, and an average over 500 structures sampled at 0.1-ns intervals.

It is clear from the results for the individual structures (Fig. 3a–d) that the calculated chemical shifts are sensitive to conformation. The corresponding variations in the computed shifts for residues 2–6 range from 6 to 21 ppm for the backbone amide nitrogen, 2 to 5 ppm for the α-carbon, and 3 to 8 ppm for the carbonyl carbon. Similar variations are found for the calculated chemical shifts of the terminal atoms (not shown in Fig. 9), which fall between −368 and −374 ppm for the terminal nitrogen, and between 183 and 186 ppm for the terminal carboxylic carbon. These variations for different conformations are relatively small, and there is no systematic trend that

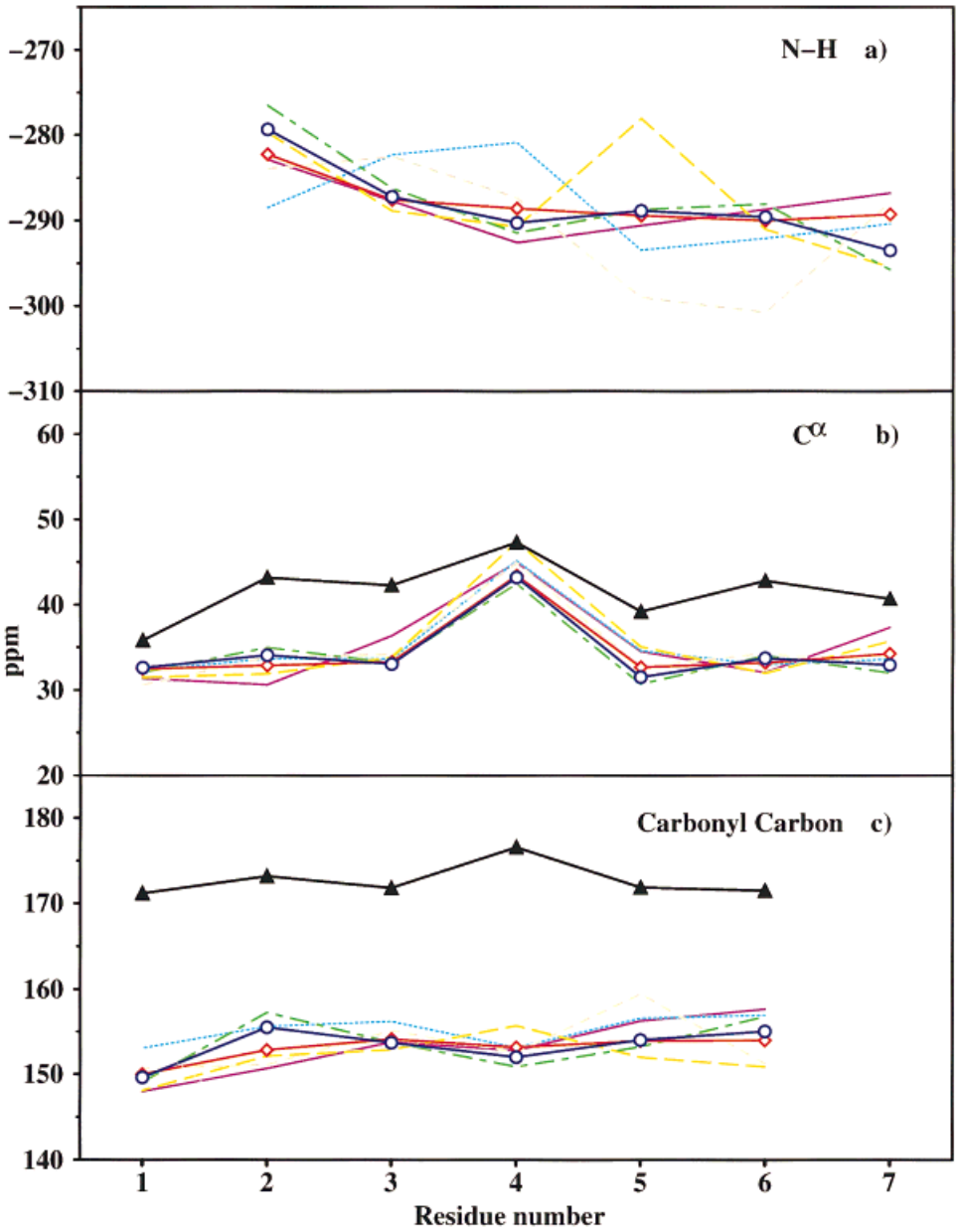

Fig. 9. Chemical shifts estimated from MNDO calculations for four structures selected from the trajectory at 340 K shown in Figure 3: **a)** Backbone amide nitrogen; **b)** backbone α-carbon; **c)** backbone carbonyl carbon. Dashed light blue line: The structure with the minimum atom-positional root-mean-square-deviation (RMSD) from the model 3_1-helical structure for the backbone atoms of residues 1 to 7 (Fig. 3a); dashed brown line: A hairpin-like structure (Fig. 3b); solid violet line: The structure with the minimum radius of gyration (Fig. 3c); dashed orange line: The structure with the maximum radius of gyration (Fig. 3d); dashed green line: The central member structure of cluster 1 (see Materials and Methods); solid blue line with circles: An average over the central member structures of all clusters with a minimum of 20 members, weighted by the number of members of each cluster; solid red line with diamonds: An average over 500 structures taken at 0.1 ns intervals; solid black line with triangles: The experimental chemical shifts measured by NMR at 298 K.[31]

would enable us to distinguish helical from non-helical conformations based on these calculations. The changes in the computed chemical shifts associated with differences in chemical bonding are larger than those associated with conformational differences. This is evident from the α-carbon atom of residue 4, which is methylated, and particularly for the terminal nitrogen and carboxylic-carbon atoms (see above).

As a check on these semi-empirical results, single-point ab initio RHF/6–311G* calculations have been carried out for each of the four structures in Figure 3a–d (relative energies at this level of 0, 33, 42, and 62 kcal mol^{-1}, respectively), and chemical shifts have been computed for the α-carbon and carbonyl carbon atoms (data not shown). The ab initio shifts are generally larger than the MNDO values (typically by 10 to 15 ppm for the α-carbon and 20 to

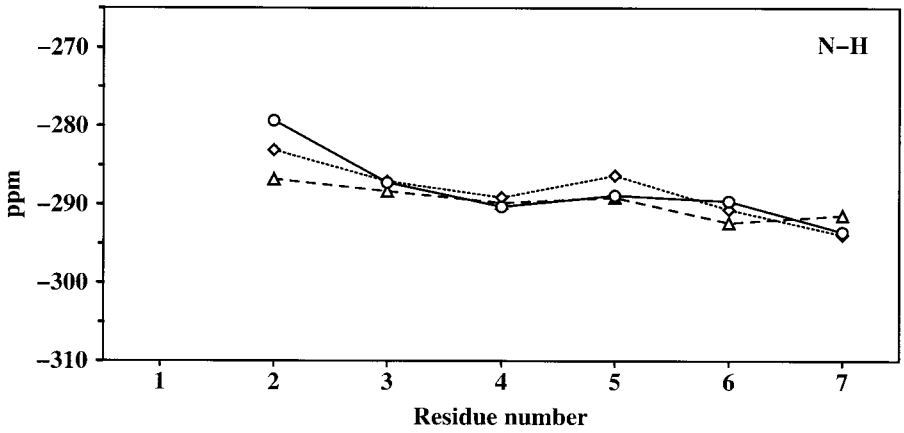

Fig. 10. Backbone amide-nitrogen chemical shifts estimated from MNDO calculations. Average over the central member structures of all clusters with a minimum of 20 members, weighted by the number of members of each cluster. Circles: MD simulation at 340 K; diamonds: MD simulation at 350 K; triangles: MD simulation at 360 K.

30 ppm for the carbonyl carbon), and they span a larger range for the four conformations considered (variations of 5 to 16 ppm for the α-carbon and 4 to 13 ppm for the carbonyl carbon, often about twice as large as in the MNDO case). However, in analogy to the MNDO results, there is again no obvious trend that would discriminate between helical and non-helical conformations, and the variations due to differences in chemical bonding remain larger than those arising from conformational differences (see above).

The four structures considered (Fig. 3a–d) represent qualitatively different conformations, e.g., a folded 3_1-helix (Fig. 3a) and an unfolded extended structure (Fig. 3d). It is also instructive to compare qualitatively similar conformations such as that with the lowest RMSD from the model 3_1-helix (Fig. 3a, light blue dashed line in Fig. 9) and the center of cluster one (green dashed line). Even though these two structures are quite similar and both have a 3_1-helical fold, the computed chemical shifts vary considerably: the differences are between 2 to 12 ppm for nitrogen and 0 to 4 ppm for carbon at the MNDO level, and between 1 to 12 ppm for α-carbon and 1 to 8 ppm for carbonyl carbon at the ab initio level. These variations are of the same order as those found for qualitatively different conformations (see above). Hence, the finer details of the structure can result in as much scatter in the calculated chemical shift as the overall fold does.

The effects of motional averaging on the computed chemical shifts can only be evaluated at the semi-empirical MNDO level since the corresponding ab initio calculations would be too expensive (see Materials and Methods, Chemical Shift Calculations). The ensemble of structures sampled during the MD simulation was approximated by taking a subset of 500 structures sampled at 0.1-ns intervals (average chemical shifts: solid red line with diamonds in Figure 9). The resulting curves are reasonably close to those obtained from a weighted average of the chemical shifts of the center structures of the clusters (solid blue line with circles in Figure 9), but there are perceptible differences (up to 4.2, 1.3, and 2.7 ppm for the backbone amide nitrogen, α-carbon, and carbonyl carbon, respectively) so that this latter average does not represent the true ensemble average with sufficient precision. A comparison of the best theoretical average at 340 K (solid red line) with the available experimental data at 298 K (solid black line) shows deviations of 3 to 10 ppm for the α-carbon atom and 17 to 23 ppm for the carbonyl carbon atom, consistent with expectations.[23] It should also be noted in this connection that the theoretical calculations refer to isolated molecules and neglect environmental effects such as interactions with the solvent.

Considering that the computed chemical shifts show only moderate variations for different folds and smaller, but still sizable, variations for different structures in a given fold (see above), the accuracy of the quantum-chemical calculations is insufficient to be diagnostic of local conformational changes, especially when motional averaging is applied. However, it may still be possible to use such calculations to predict trends in chemical shifts associated with the sampling of different regions of conformational space. Figure 10 shows the amide nitrogen chemical shift as an average over the central member structures of all clusters containing at least 20 members weighted by the number of members of each cluster, from the simulations at 340 K, 350 K, and 360 K. Despite the frequency of sampling conformations other than the 3_1-helical conformation increasing with temperature, the weighted average of the chemical shift for each atom is similar at the three temperatures and well within the uncertainty of the approach, especially in conjunction with cluster weighted averages. Averaging over a representative ensemble may improve the fidelity of the results, but the change in chemical shift as a function of temperature due to the sampling of alternate conformations is predicted to be slight.

DISCUSSION

When modelling the structure of a peptide based on NMR data, it is common to search for a structure (or group of structures) that individually satisfy the experimental

data. Such structures represent individual solutions to a set of constraints rather than necessarily reflecting the ensemble of structures accessible to the peptide in solution. Given the average nature of the inter-proton distances inferred from NOE intensities and of the NMR ^3J-coupling constants, the modelled structures will contain structural information from spatially ($> 10^{15}$ molecules) and temporally (typically ms) independent peptide conformations. Populations of potentially very different conformations may be hidden within the average. In fact, there may well exist no specific structure in the sample that conforms to the proposed model, i.e., a structure that simultaneously fulfills all the NMR-derived constraints. This is clearly illustrated in Figures 2 and 4. At 340 K, the average distance violation calculated over the entire trajectory (0.008 nm) is below the average distance violation for the structure with the lowest backbone atom-positional RMSD deviation from the model left-handed 3_1-helix (0.009 nm) despite only 50% of structures in the trajectory having the 3_1-helical fold. Moreover, the structure with the lowest RMSD violates one of the distances inferred from the NMR data (number 42 in Table I) by 0.13 nm, a distance which is not violated in the trajectory average. Even at 360 K, where only 25% of the sampled structures have the 3_1-helical fold, there are few violations, none larger than 0.1 nm. A similar conclusion is reached regarding the ^3J-coupling constants. The calculated average ^3J-coupling constants at 340 K and even at 360 K are closer to the experimentally measured ^3J-coupling constants than those for the specific structure with the lowest RMSD from the model 3_1-helix.

When analyzing the ensembles of conformations at different temperatures, the calculated average inter-proton distances show a greater sensitivity to differences in the populations of conformations than do the ^3J-coupling constants. The latter were calculated using the empirically parameterized Karplus relation.[14] This relates ^3J-coupling constants between protons connected through three bonds to the corresponding torsional dihedral angle. Results based on the Karplus relation are ambiguous in the sense that the sinusoidal form of this relation means that for certain values there is more than one solution (dihedral angle) for a given ^3J-coupling constant. Due to energetic considerations the range of highly populated dihedral angles is also limited. This means that although the populations of conformations vary between the four different temperatures studied, the actual populations of dihedral angles sampled may vary only marginally and the apparent ^3J-coupling constant to an even lower degree.

The averages displayed in Figures 5 to 8 are given for inter-proton distance violations calculated using an $<r^{-6}>^{-1/6}$ as well as an $<r^{-3}>^{-1/3}$ weighting. Standardly, when inferring inter-proton distances from NOE crosspeaks it is assumed that the rate of magnetization transfer is proportional to r^{-6}, r being the inter-proton distance. Tropp,[1] however, has shown that in cases where there is fast internal motion but a comparatively long rotational correlation time for the molecule overall, such as might occur in globular macromolecules, the expression for the cross-relaxation rates reduces to an r^{-3} dependence. For this reason, especially when the time scale of the sampling is short, an $<r^{-3}>^{-1/3}$ weighting is frequently used to calculate inter-proton distance violations in MD simulations. This is despite the fact that the reference distances derived from the NMR data were in most cases obtained assuming an r^{-6} dependence. This has two consequences. First, the assumption of an r^{-6} dependence in the experimental data restricts the potential range of the NOE, resulting in tight experimental constraints. Second, the use of $<r^{-3}>^{-1/3}$ in the analysis of MD trajectories reduces the effect of motional averaging on the apparent distance violations, again enhancing the degree to which the experimental constraints infer limited conformational sampling. In cases where the overall tumbling rate is fast compared to the internal motion, Tropp showed that an $<r^{-6}>^{-1/6}$ weighting is appropriate. For the β-heptapeptide studied here, conformational transitions occur in the simulations on a time scale of nanoseconds to tens of nanoseconds, whereas the rotational correlation time for the molecule is in the order of 0.5-ns or less (as implied by the use of ROESY spectra), and an $<r^{-6}>^{-1/6}$ weighting is appropriate. In this case it is very clear that the presence of specific NOE's which can constrain the molecule to a particular fold (in this case a 3_1-helix), cannot be taken as sufficient evidence to claim the molecule is predominately folded under a given set of conditions. In densely packed systems an equilibrium between two or more states might be indicated by the presence of mutually conflicting distance restraints. In the case of the β-peptide at 340 K, where the molecule is only folded approximately 50% of the time, the second most populated conformation is present only 6% of the time. Only where a very close approach between two atoms widely separated in sequence and in the folded state would occur, an NOE restraint conflict would be detected.

In principle, NMR chemical shifts also contain structural information, and the presence of chemical shift dispersion is one of the primary aspects of protein and peptide NMR spectra which is used to identify persistent structure in solution. The degree of electronic shielding of nuclei is, however, dependent on very slight differences in electronic structure which cannot be reliably approximated using classical models. In this study we have applied a recently developed semi-empirical approach based on a specifically parameterized MNDO method to analyze the influence of motional averaging on computed chemical shifts. The semi-empirical as well as single-point ab initio calculations confirm that the shifts depend on the conformation, with moderate variations between different folds and still sizable differences between different structures of the same fold. Under these circumstances, the accuracy of the chosen quantum-chemical approach is insufficient to monitor the changes in the populations of specific conformers through motional averaging.

In conclusion, it has been shown that when motional averaging is taken into account NMR data has only limited capacity to distinguish between the peptide adopting a single folded conformation and various mixtures of folded and unfolded conformations.

ACKNOWLEDGMENTS

We thank K. Gademann, B. Jaun, and D. Seebach for providing the experimental data on the β-heptapeptide, and D. Bakowies for helpful comments on the manuscript.

REFERENCES

1. Tropp J. Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances. J Chem Phys 1980;72:6035–6043.
2. Hoch JC, Dobson CM, Karplus M. Fluctuations and averaging of proton chemical shifts in the bovine pancreatic trypsin inhibitor. Biochemistry 1982;21:1118–1125.
3. Olejniczak ET, Dobson CM, Karplus M, Levy RM. Motional averaging of proton nuclear Overhauser effects in proteins. Predictions from a molecular dynamics simulation of lysozyme. J Am Chem Soc 1984;106:1923–1930.
4. Hoch JC, Dobson CM, Karplus M. Vicinal coupling constants and protein dynamics. Biochemistry 1985;24:3831–3841.
5. Kessler H, Griesinger C, Lautz J, Müller A, van Gunsteren WF, Berendsen HJC. Conformational dynamics detected by nuclear magnetic resonance NOE values and *J* coupling constants. J Am Chem Soc 1988;110:3393–3396.
6. Van Gunsteren WF, Brunne RM, Gros P, van Schaik RC, Schiffer CA, Torda AE. Accounting for molecular mobility in structure determination based on nuclear magnetic resonance spectroscopic and X-ray diffraction data. In: James TL, Oppenheimer NJ, editors. Methods in enzymology: nuclear magnetic resonance, Vol 239. New York: Academic Press; 1994. p 619–654.
7. Philippopoulos M, Lim C. Internal motions in the molecular tumbling regime. Effect on NMR dipolar cross-relaxation and interproton distance determination. J Phys Chem 1994;98:8264–8273.
8. Brüschweiler R, Case DA. Characterization of biomolecular structure and dynamics by NMR cross relaxation. Prog NMR Spectroscopy 1994;26:27–58.
9. Abseher R, Lüdemann S, Schreiber H, Steinhauser O. NMR cross-relaxation investigation by molecular dynamics simulation: a case study of ubiquitin in solution. J Mol Biol 1995;249:604–624.
10. Schneider TR, Brünger AT, Nilges M. Influence of internal dynamics on accuracy of protein NMR structures: derivation of realistic model distance data from a long molecular dynamics trajectory. J Mol Biol 1999;285:727–740.
11. Wüthrich K. NMR of proteins and nucleic acids. New York: JohnWiley & Sons; 1986. 292 p.
12. Torda AE, van Gunsteren WF. Molecular modeling using nuclear magnetic resonance data. In: Lipkowitz KB, Boyd DB, editors. Reviews in computational chemistry, Vol III. New York: VCH Publishers; 1992. p 143–172.
13. Scott WRP, Mark AE, van Gunsteren WF. On using time-averaging restraints in molecular dynamics simulation. J Biomol NMR 1998;12:501–508.
14. Karplus M. Contact electron-spin interactions of nuclear magnetic moments. J Chem Phys 1959;30:11–15.
15. Bax A, Tjandra N. Are proteins even floppier than we thought? Nat Struct Biol 1997;4:254–256.
16. de Groot BL, Amadei A, Scheek RM, van Nuland NAJ, Berendsen HJC. An extended sampling of the configurational space of HPR from *E. coli*. Proteins 1996;26:314–322.
17. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Reversible peptide folding in solution by molecular dynamics simulation. J Mol Biol 1998;280:925–932.
18. van Gunsteren WF, Billeter SR, Eising AA, et al. Biomolecular simulation: the GROMOS96 manual and user guide. Zürich: Vdf Hochschulverlag AG an der ETH Zürich; 1996. p 1–1024.
19. Daura X, van Gunsteren WF, Rigo D, Jaun B, Seebach D. Studying the stability of a helical β-heptapeptide by molecular dynamics simulations. Chem Eur J 1997;3:1410–1417.
20. Pardi A, Billeter M, Wüthrich K. Calibration of the angular dependence of the amide proton-Cα proton coupling constants, ^3JHNα, in a globular protein: Use of ^3JHNα for identification of helical secondary structure. J Mol Biol 1984;180:741–751.
21. de Marco A, Llinás M, Wüthrich K. Analysis of the ^{1}H NMR spectra of ferrichrome peptides I: The non-amide protons. Biopolymers 1978;17:617–636.
22. Daura X, van Gunsteren WF, Mark AE. Folding-unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. Proteins 1999;34:269–280.
23. Patchkovskii S, Thiel W. NMR chemical shifts in MNDO approximation: Parameters and results for C, H, N, and O. J Comput Chem 1999; in press.
24. Patchkovskii S. Analytical computation of first-order response properties in MNDO methods (dissertation). Zürich: University of Zürich. 1997.
25. Dewar MJS, Thiel W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. J Am Chem Soc 1977;99:4899–4907.
26. Wüthrich, K. NMR in biological research: peptides and proteins. Amsterdam: North Holland; 1976. 379 p.
27. Hehre WJ, Radom L, Schleyer PvR, Pople JA. Ab initio molecular orbital theory. New York: John Wiley & Sons, 1986. 548 p.
28. Frisch MJ, Trucks GW, Schlegel HB, et al. GAUSSIAN94, Revision E.2. Pittsburgh, PA: Gaussian Inc.; 1995.
29. Cheeseman JR, Trucks GW, Keith TA, Frisch MJ. A comparison of models for calculating nuclear magnetic resonance shielding tensors. J Chem Phys 1996;104:5497–5509.
30. Thiel W. Program MNDO97, Version 5.0, Zürich. 1998.
31. Seebach D, Ciceri PE, Overhand M, et al. Probing the helical secondary structure of short-chain β-peptides. Helv Chim Acta 1996;79:2043–2066.
32. Gademann K, Jaun B, Seebach D, Perozzo R, Scapozza L, Folkers G. Temperature-dependent NMR and CD spectra of β-peptides: On the thermal stability of β-peptide helices— Is the folding process of β-peptides non-cooperative? Helv Chim Acta 1999;82:1–11.