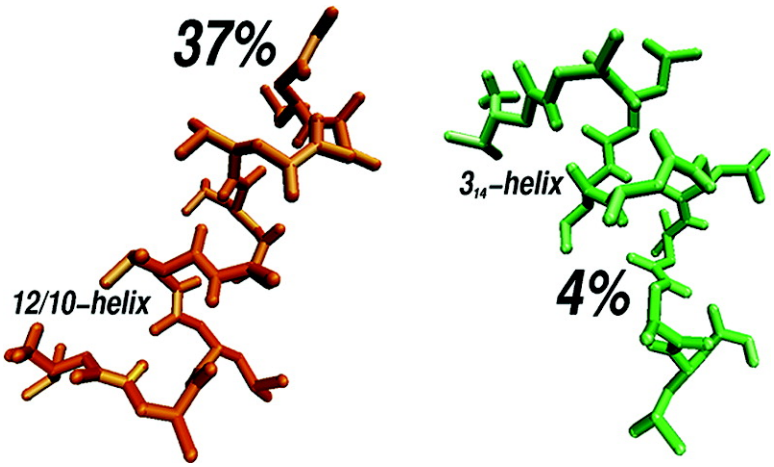
Article

Interpreting NMR Data for β-Peptides Using Molecular Dynamics Simulations

Daniel Trzesniak, Alice Glttli, Bernhard Jaun, and Wilfred F. van Gunsteren

J. Am. Chem. Soc., **2005**, 127 (41), 14320-14329• DOI: 10.1021/ja044285h • Publication Date (Web): 27 September 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 2 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 09/27/2005

Interpreting NMR Data for β-Peptides Using Molecular Dynamics Simulations

Daniel Trzesniak,[†] Alice Glättli,[†] Bernhard Jaun,[‡] and Wilfred F. van Gunsteren*,[†]

Contribution form the Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich, ETH-Hönggerberg, CH-8093 Zürich, Switzerland, and Laboratory of Organic Chemistry, Swiss Federal Institute of Technology Zürich, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

Received September 20, 2004; E-mail: wfvgn@igc.phys.chem.ethz.ch

Abstract: NMR is one of the most used techniques to resolve structure of proteins and peptides in solution. However, inconsistencies may occur due to the fact that a polypeptide may adopt more than one conformation. Since the NOE distance bounds and ^3J-values used in such structure determination represent a nonlinear average over the total ensemble of conformers, imposition of NOE or ^3J-value restraints to obtain one unique conformation is not an appropriate procedure in such cases. Here, we show that unrestrained MD simulation of a solute in solution using a high-quality force field yields a conformational ensemble that is largely compatible with the experimental NMR data on the solute. Four 100 ns MD simulations of two forms of a nine-residue β-peptide in methanol at two temperatures produced conformational ensembles that were used to interpret the NMR data on this molecule and resolve inconsistencies between the experimental NOEs. The protected and unprotected forms of the β-peptide adopt predominantly a *12/10*-helix in agreement with the qualitative interpretation of the NMR data. However, a particular NOE was not compatible with this helix indicating the presence of other conformations. The simulations showed that 3_{14}-helical structures were present in the ensemble of the unprotected form and that their presence correlates with the fulfillment of the particular NOE. Additionally, all inter-hydrogen distances were calculated to compare NOEs predicted by the simulations to the ones observed experimentally. The MD conformational ensembles allowed for a detailed and consistent interpretation of the experimental data and showed the small but specific conformational differences between the protected and unprotected forms of the peptide.

Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy has established itself as a potent method for structure determination of biological molecules in solution.[1−3] The procedure for structure evaluation by NMR involves a number of steps. Different NMR experiments produce NOESY or ROESY spectral intensities, which can be converted into a set of spatial distance bounds for specific pairs of atoms. However, the interatomic distances and their fluctuations are not the only quantities which determine NMR relaxation rates and thus NOESY or ROESY cross-peak intensities, but there is also an additional contribution from intramolecular motions to the observed signal by their time scales and their orientational correlations. These contributions have been investigated earlier,[4−6] and are left out of consideration in the present article, in which neither the effect of spin diffusion is investigated. After acquiring as many as possible of such distance bounds, model structures are derived and should not violate those distances bounds. This is standardly done by performing a simulated annealing geometry optimization in the presence of the distance restraints representing the experimental NMR Nuclear Overhauser Effect (NOE) distance bounds, ideal bond distances, and angles values.[7] This procedure generally works fine for proteins and peptides that adopt a single dominant conformation, but if this is not the case, problems may arise. First, ambiguities may arise when different conformations reproduce the experimental data likewise. A striking example was observed when considering the CD spectra of different conformations of β-peptides.[8] A similar example using NMR data has also been reported.[6] In the case of NMR, there are generally not enough independent NOE distance bounds measurable to establish a complete structure uniquely.[9,10] Often parts of the protein remain structurally undetermined. These are left unspecified or determined

* Corresponding author.
[†] Laboratory of Physical Chemistry.
[‡] Laboratory of Organic Chemistry.

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.
(2) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: San Diego, 1996.
(3) Wüthrich, K. *Nat. Struct. Biol.* **2001**, *8*, 923−925.
(4) Peter, C.; Daura, X.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *20*, 297−310.
(5) Feenstra, K. A.; Peter, C.; Scheek, R. M.; van Gunsteren, W. F.; Mark, A. E. *J. Biomol. NMR* **2002**, *23*, 181−194.
(6) Peter, C.; Rueping, M.; Worner, H. J.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. *Chem.-Eur. J.* **2003**, *9*, 5838−5849.
(7) Brünger, A. T.; Krukowski, A.; Erickson, J. W. *Acta Crystallogr. Sect. A* **1990**, *46*, 585−593.
(8) Glättli, A.; Daura, X.; Seebach, D.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2002**, *124*, 12972−12978.
(9) Jardetzky, O. *Biochim. Biophys. Acta* **1980**, *621*, 227−232.
(10) Glättli, A.; van Gunsteren, W. F. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 6312−6316.

10.1021/ja044285h CCC: $30.25

using only force-field data. In the latter case, such parts of a protein or peptide are characterized by a large structural variety in the set or bundle of NMR model structures reported in structure data banks. Second, different mutually excluding NOE signals may arise from distinct structures that are comparably relevant to the ensemble of structures in solution. These generate inconsistencies between NOEs that cannot be resolved using a single structure, as pointed out before in several studies.[9,11−13] This problem occurs especially when dealing with small and flexible peptides, which are characterized by a variety of internal motions and conformations. NOE intensities can be related to nonlinear inter-proton distance averages of the molecule during the NMR experiment. When it happens that the ensemble of structures in solution is dominated by more than one conformation, the commonly used assumption that all NOEs originate exclusively from one structure is no longer valid and the nonlinear distance averages contain contributions from different conformers.

In this case, irrespective of whether the multiple conformations do or do not lead to structurally inconsistent NOE signals, Molecular Dynamics (MD) simulations[14] using a high-quality force field can contribute to a correct interpretation of the experimental data by providing an atomic resolution picture based on a statistical-mechanically correct ensemble of structures and may in addition yield information about the dynamic processes involved. MD can therefore be a complementary and reliable tool in structure refinement of bio-molecules. For instance, it is nowadays feasible to simulate the reversible folding of small peptides starting from an extended conformation and to generate an accurate (un)folding equilibrium.[6,8,10,11,15−26] With such information it is possible to identify the relevant conformations and assess their weight in the ensemble of structures in solution. In the present work we address an improved interpretation of NMR experiments by the use of long-time MD simulations of bio-molecules in solution based on a high-quality force field, which enables the resolution of seeming inconsistencies in the experimental data. We show the predictive power of unrestrained MD simulation using a thermodynamically calibrated force field to interpret NMR data and to elucidate the relation between measured values of NMR observables and the corresponding three-dimensional structures for a β-nonapeptide in solution.

β-Peptides belong to a class of compounds sometimes alluded to as foldamers,[27,28] because of their ability to form stable secondary structure elements even with as few as only four β−amino acids.[29] This exceptional characteristic makes them very convenient for folding studies.[23,26,30−32] They differ from a natural α-amino acid by an extra carbon atom in the backbone. This opens up the possibility to study folding as a function of the type of side chain, the stereochemistry and the position of the side chain at the backbone (α or β). The molecules chosen for this study were protected and unprotected forms of the β-nonapeptide shown in Figure 1.[32] The β-peptide chain is comprised of nine residues of alternating β^3/β^2 substituted amino acids and consists of three sequences of three amino acids, namely valine, alanine and leucine. The protected form bears the Boc group at the N-terminus and the Bn group at the C-terminus. The unprotected form has both termini protonated, the most probable charge state in methanol. It was concluded based on NMR and CD experiments that this peptide most likely adopts a *12/10*-helical conformation independent of its protection state.[32] The protecting groups seemed only to make the *12/10*-helix more stable. It was suggested that this is due to the Boc ester carbonyl oxygen participating in an additional *12*-membered hydrogen bonded ring.[32] However, three NOEs of the unprotected form were mentioned to be incompatible with a *12/10*-helix and were not considered in the structure determination.[32] This indicates that the molecule may also adopt other relevant conformations than this helix. To investigate the conformational ensembles of both forms of the β-peptide nonamer in methanol, simulations at 298 and 340 K without any conformational NOE or ^3J-value restraints were carried out (for further details see the methods section). The ensembles of structures from the simulation trajectories were analyzed in terms of conformational space sampled, structural properties such as hydrogen bonding and in terms of the level of agreement with the available NMR data (NOE intensities and ^3J-coupling constants). The inconsistencies mentioned above were considered and further investigated by analyzing inter-proton distances after performing a conformational cluster analysis. In addition, all possible inter-proton distances on the peptide were calculated to verify whether the structures obtained would predict NOEs not observed experimentally. The uniqueness of the NMR data for each form of the peptide was examined by cross comparisons of the experimental and simulated data for both forms (the experimental data of the protected form of the peptide was compared with the simulated data of the unprotected form and vice versa). This analysis would indicate if the data available would be able to distinguish between the protected and unprotected forms.

Methods

All MD trajectories were generated using the GROMOS96 software package.[33,34] The force field parameters for the protected and unprotected β-peptides were taken from the 45A3 GROMOS united atom

(11) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(12) Daura, X.; Antes, I.; van Gunsteren, W. F.; Thiel, W.; Mark, A. E. *Proteins* **1999**, *36*, 542−555.
(13) Bürgi, R.; Pitera, J.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *19*, 305−320.
(14) Norberg, J.; Nilsson, L. *Q. Rev. Biophys.* **2003**, *36*, 257−306.
(15) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925−932.
(16) Takano, M.; Yamato, T.; Higo, J.; Suyama, A.; Nagayama, K. *J. Am. Chem. Soc.* **1999**, *121*, 605−612.
(17) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062−9067.
(18) Daura, X.; van Gunsteren, W. F.; Mark, A. E. *Proteins* **1999**, *34*, 269−280.
(19) Ma, B. Y.; Nussinov, R. *J. Mol. Biol.* **2000**, *296*, 1091−1104.
(20) Wang, H. W.; Sung, S. S. *J. Am. Chem. Soc.* **2000**, *122*, 1999−2009.
(21) Hummer, G.; Garcia, A. E.; Garde, S. *Proteins* **2001**, *42*, 77−84.
(22) Daura, X.; Gademann, K.; Schafer, H.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 2393−2404.
(23) van Gunsteren, W. F.; Bürgi, P.; Peter, C.; Daura, X. *Angew. Chem.-Int. Edit.* **2001**, *40*, 351−355, 4616−4618.
(24) Colombo, G.; Roccatano, D.; Mark, A. E. *Proteins* **2002**, *46*, 380−392.
(25) Wu, H. W.; Wang, S. M.; Brooks, B. R. *J. Am. Chem. Soc.* **2002**, *124*, 5282−5283.
(26) Daura, X.; Glättli, A.; Gee, P.; Peter, C.; van Gunsteren, W. F. *Adv. Protein Chem.* **2002**, *62*, 341−360.
(27) Gellman, S. H. *Acc. Chem. Res.* **1998**, *31*, 173−180.
(28) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. *Chem. Rev.* **2001**, *101*, 3893−4011.
(29) Seebach, D.; Matthews, J. L. *Chem. Commun.* **1997**, 2015−2022.
(30) Seebach, D.; Schreiber, J. V.; Abele, S.; Daura, X.; van Gunsteren, W. F. *Helv. Chim. Acta* **2000**, *83*, 34−57.
(31) Cheng, R. P.; Gellman, S. H.; DeGrado, W. F. *Chem. Rev.* **2001**, *101*, 3219−3232.
(32) Rueping, M.; Schreiber, J. V.; Lelais, G.; Jaun, B.; Seebach, D. *Helv. Chim. Acta* **2002**, *85*, 2577−2593.

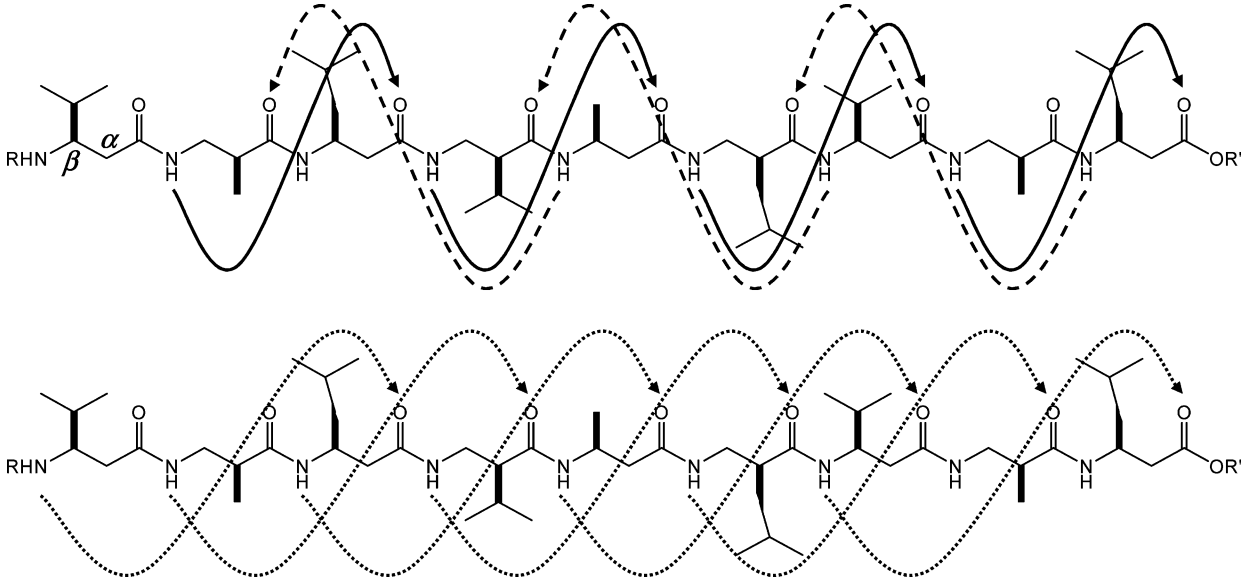


Figure 1. Molecular formula of the β-nonapeptide studied. The terminal groups are omitted and denoted by R and R'. In the protected case R = Boc ($(CH_3)-O-(C=O)-$) and R' = Bn ($-CH_2-C_6H_6$) and in the unprotected case $R = H_2$ and $R' = H$. The arrows indicate the hydrogen bond pattern characteristic of the *12/10*- and 3_{14}-helices. The *12/10*-helix (upper part) is characterized by *10* (solid line) and *12* (dashed line) membered hydrogen-bonded rings, whereas the 3_{14}-helix (lower part) would show *14* membered hydrogen bonded rings (dotted line).

force field.[33,35,36] The nonstandard building blocks for the protecting groups Boc and Bn were constructed based on the same set of parameters. The starting structure for the β−peptide was a fully extended conformation (in which all backbone dihedral angles are set to 180°) placed in the center of a truncated octahedron box (minimum solute to (square) wall distance of 1.4 nm and periodic boundary conditions) filled with 3319 (protected form) or 2797 (unprotected form) methanol molecules (standard GROMOS96 methanol solvent model).[33,37] The MD simulation parameter settings were similar to those of other β−peptide simulations done previously in our group.[6,8,11,15,22] Preceded by 1 ns of equilibration, two 100 ns long simulations at constant temperature (298 and 340 K) and pressure (1 atm), held constant using the weak coupling technique[38] with relaxation times of 0.1 and 0.5 ps respectively, were performed for each of the two forms of the β-peptide. The experimental data had been obtained at room temperature and the other higher temperature (340 K) was chosen to improve conformational sampling. All bonds were kept rigid using the SHAKE[39] method with a relative geometric tolerance of 10^{-4} and the MD equations of motion were integrated using 2 fs time steps. Long-range interactions were treated using a triple-range scheme with cutoff radii of 0.8 nm (where the interactions were updated every time step) and 1.4 nm (for which interactions were updated every fifth time step) and a reaction-field permittivity of 32 for methanol. Configurations were saved at every 2 ps for analysis. Atom-positional root-mean-square difference (RMSD) between pairs of structures were calculated after superposition of the backbone atoms of residues 2 to 8. The structures used as helical references were built based on ideal values for the backbone dihedral angles for the *12/10*- and 3_{14}-helices.[40] A conformational cluster analysis, as described in earlier studies,[18] was performed on the separate and combined trajectories of the protected and unprotected β-peptides using structures at 20 ps intervals and a RMSD similarity criterion of 0.12 nm. The criterion for defining a hydrogen bond was the standard one used with GROMOS, where the maximum hydrogen-acceptor atom distance is 0.25 nm and the minimum donor atom-hydrogen-acceptor atom angle is 135°. NOE distance bound violations were obtained by comparing the nonlinear averaged proton−proton distances (direct r^{-6} averaging, which is appropriate for small molecules)[12,41] from the simulation with the upper bound distances derived from the experimental NOE intensities[32] (strong intensity 0.30 nm, medium intensity 0.35 nm, weak intensity 0.45 nm). Pseudo-atom corrections (0.09 nm for CH_2, 0.10 nm for CH_3, 0.22 nm for $(CH_3)_2$ and $(CH_3)_3$) were applied to the experimental upper bounds and the virtual atom creation technique[33,34] was used for the hydrogen atoms attached to carbon atoms, because a united-atom force-field was used. ^3J-coupling constants were calculated from the simulations via the Karplus relation[42]

$$^3J(\mathrm{HN,H}\beta) = a\cos^2\theta + b\cos\theta + c$$

where a, b, and c are equal to 6.4, −1.4, and 1.9 Hz,[43] respectively. We note that these values for the parameters have been calibrated to data for α-peptides and not β-peptides.

Results and Discussion

Figure 2 shows the atom-positional RMSD for the backbone atoms with respect to *12/10*- (black lines) and 3_{14}-(red lines) helical model structures. Indeed, eventually all the simulations sample predominantly the *12/10*-helix. Nonetheless, the mode and frequency in which this happens are distinct. At 298 K, the protected form (upper left) samples the 3_{14}-helix for approximately 40 ns before adopting the *12/10*-helix. At 340 K (bottom left), the protected molecule reaches the dominant native *12/10*-helical structure 10 times faster. The 3_{14}-helix is not

(33) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf Hochschulverslag: Zürich, 1996.
(34) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.
(35) Daura, X.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **1998**, *19*, 535−547.
(36) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205−1218.
(37) Walser, R.; Mark, A. E.; van Gunsteren, W. F.; Lauterbach, M.; Wipff, G. *J. Chem. Phys.* **2000**, *112*, 10450−10459.
(38) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(39) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. **1977**, *23*, 327−341.
(40) Wu, Y. D.; Lin, J. Q.; Zhao, Y. L. *Helv. Chim. Acta* **2002**, *85*, 3144−3160.
(41) Tropp, J. *J. Chem. Phys.* **1980**, *72*, 6035−6043.
(42) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11−15.
(43) Pardi, A.; Billeter, M.; Wüthrich, K. *J. Mol. Biol.* **1984**, *180*, 741−751.

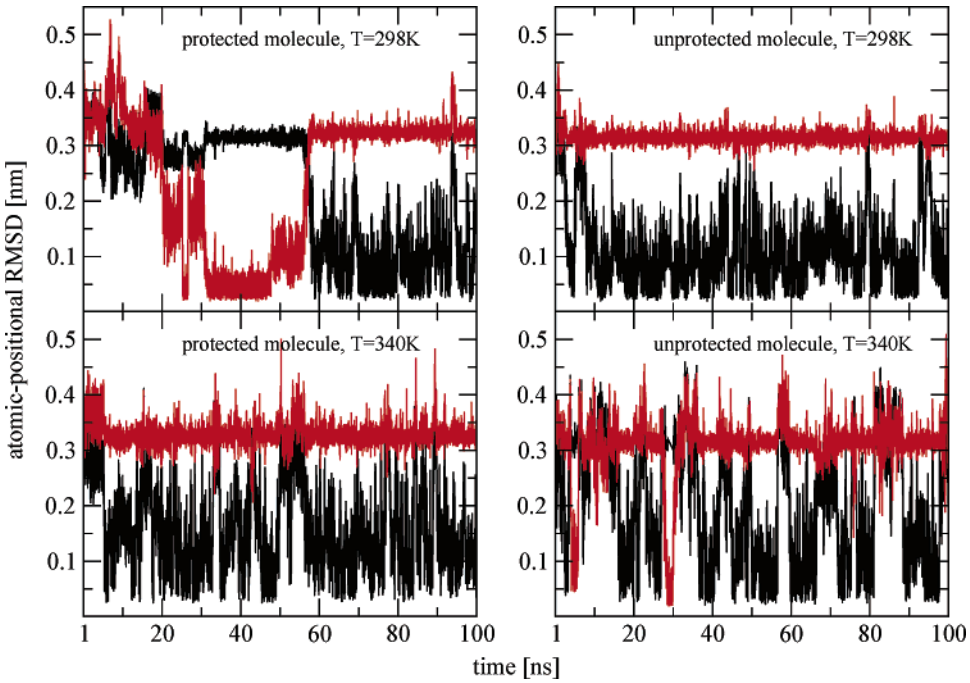


Figure 2. Time series of the backbone atom (N, Cβ, Cα, C)-positional root-mean-square distance (RMSD) with respect to ideal *12/10*- (black) and 3_{14}-(red) helices. Only residues 2−8 are included in the RMSD calculation. The results of the protected β-peptide stand on the left side and those of the unprotected on the right. The top panels refer to the 298 K simulations, the bottom panels refer to the 340 K ones.

sampled. When comparing the simulations at different temperatures after the point (60 ns) at which at 298 K the molecule folds into a *12/10*-helical conformation, it can be seen that at high temperature the (un)folding equilibrium with respect to the *12/10*-helix is shifted to the unfolded state and the folding/unfolding process is occurring more frequently as one would have anticipated. The same is observed for the unprotected molecule which folds almost directly into a *12/10*-helix at low (upper right) and high (bottom right) temperatures. The 3_{14}-helix is not seen at all in the 298 K simulation and is sampled twice at 340 K. Sampling more diverse conformations at a higher temperature is expected because the molecule has more energy to explore other conformations. The *12/10*-helical conformation seems to be slightly more stable for the protected molecule than for the unprotected one (after it is formed for the first time), which might be due to the formation of an additional hydrogen bond to the carbonyl oxygen of the Boc group mentioned before. The occurrence of hydrogen bonds with a population of at least 5% is given in Table 1. The first column identifies the residues and respective hydrogen bond donor and acceptor atoms involved, the subsequent columns show the hydrogen bond populations for the two simulations at both temperatures and for the NMR derived model structures of both forms at room temperature. The conclusions drawn from the RMSD analysis are consistent with the hydrogen bond analysis. The 3_{14}-helix is only well formed in the protected case at low temperature while at high temperature there is just one hydrogen bond at low population. Formation of *14*-membered hydrogen bonded rings is also observed in the unprotected 340 K simulation at the N-terminal half of the molecule. *12*- and *10*-membered hydrogen bonded rings are dominant in all cases, confirming once more that the peptide prefers a *12/10*-helix. The hydrogen bonds sampled are in agreement with the expected hydrogen bond pattern[32] shown in Figure 1 (upper panel). A direct comparison of the hydrogen bonds and the *12/10*-helix formed in the simulations of the protected molecule with those of the unprotected one is not possible at 298 K, because at this temperature the simulations do not sample the ensembles sufficiently well to detect differences. However, at 340 K such a comparison between the two forms of the molecule seems warranted. The *12/10*-helix appears slightly more stable for the protected peptide and the corresponding hydrogen bonds are on average more populated. The hydrogen bonds present in the NMR derived model structures indicate the same.

So far, the results from the RMSD and hydrogen bond analyses appear to indicate that the preferred conformation of the peptide in methanol is a *12/10*-helix as was proposed based

Table 1. Occurrence (in %) of Backbone Hydrogen Bonds[a]

	protected molecule			unprotected molecule			
	MD		NMR (8)	MD		NMR (13)	
H–bond	298 K (%)	340 K (%)	298 K (%)	298 K (%)	340 K (%)	298 K (%)	type of H-bond ring
2NH − 3O	32	30	100	43	25	8	*10*
4NH − 5O	39	69	63	83	54	23	*10*
6NH − 7O	43	62	88	75	55	69	*10*
8NH − 9O	23	22	25	53	43		*10*
3NH − BocO			13				*12*
5NH − 2O	25	48	13	62	38	46	*12*
7NH − 4O	43	69	25	85	48	54	*12*
9NH − 6O	36	46	38	53	41	39	*12*
1NH − 3O	19	7			5	8	*14*
2NH − 4O	19				5		*14*
3NH − 5O	27				8		*14*
4NH − 6O	33				5		*14*
5NH − 7O	35						*14*
6NH − 8O	25						*14*
7NH − 9O	22						*14*

[a] The first column identifies the hydrogen bond and the subsequent columns show the populations larger than 5%. The values are grouped according to the size and type of the resulting hydrogen-bonded ring (last column). The averages over the experimentally derived NMR model structures[32] are also calculated, the number of model structures being shown within parentheses.

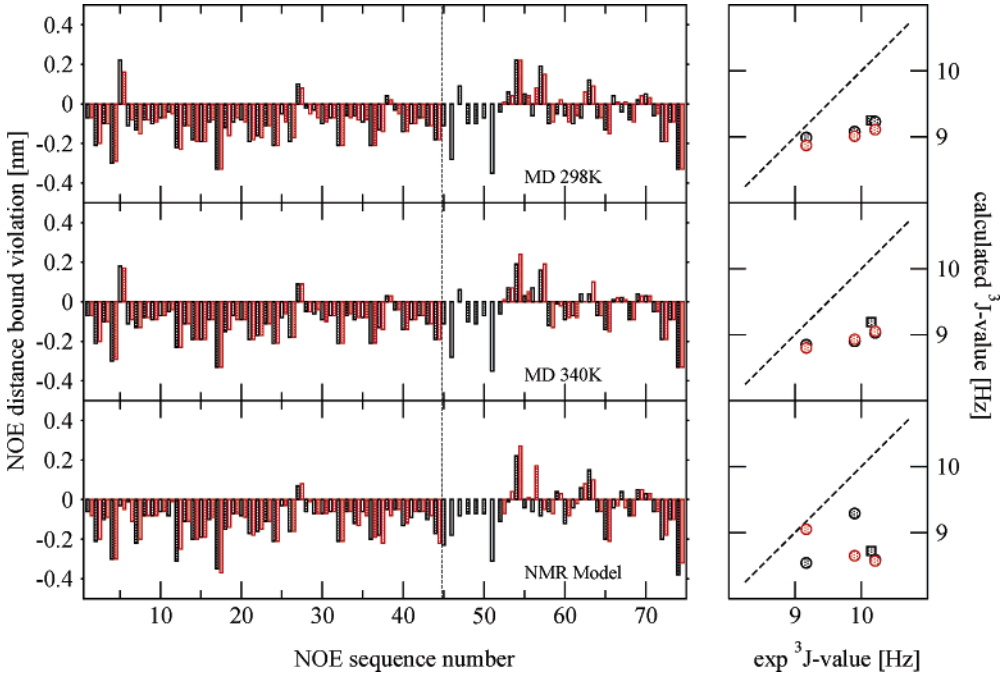


Figure 3. NOE distance bound violations (left-hand panels) and ^3J-coupling constants (right-hand panels). The experimental NMR data (74 NOE bounds and 4 ^3J-coupling constants) comes from the **protected** molecule.[32] The topmost panels refer to the 298 K simulation, the intermediate panels to the 340 K simulation and the bottom panels to the set of model structures derived by structure refinement based on the NMR data.[32] The black color represents structures the protected molecule while red stands for structures of the unprotected form. In the NOE distance violation panels the dashed line divides the graphs such that the NOEs to the left were measured for both the protected and the unprotected molecules (44 NOEs) and those to the right were only observed for the protected peptide. NOEs 45−51 involve the HN of residue 1 and were not calculated for the unprotected form. In the ^3J-coupling constant panels the dashed line represents perfect correlation between measured and calculated values. The circles are ^3J-values that were measured for both peptides whereas the squares were measured only for the **protected** form. The ^3J-value for the first valine residue was only calculated for the protected form.

on the experimental NMR data. However, it was seen in the simulations that other conformations are populated too, which may contribute to the NMR signals. Moreover, the experimental structure refinement of the unprotected molecule did not make use of all available NOEs, because three of them were mentioned to be inconsistent with the *12/10*-helix indicated by the majority of the NOEs. These three NOEs connect atoms that are rather far apart in the residue sequence: residues i and $i + j$ with $j = 3$. Experimentally, those were the only three observed NOEs for which $j = 3$ (there were none with $j > 3$). So all longest range NOEs were omitted in the NMR structure refinement of the unprotected molecule.[32] Therefore, a reinterpretation of the NMR data seems warranted. Are the simulations consistent with the NOEs, the primary experimental data? If yes, can they be used to interpret the experimental data and to determine relevant conformations? To address the first question the NOE distances and ^3J-coupling constants were calculated from the trajectories and from the set of 8 (protected) and 13 (unprotected) NMR model structures[32] (see Figures 3 and 4). A test of the uniqueness of the NMR data for each molecule was also performed by a cross comparison of the NOE distance bounds and ^3J-coupling constants for one molecule with these quantities calculated from the trajectory or NMR model structures for the other molecule and vice-versa, see Figures 3 and 4. In both Figures, black color refers to structures of the protected molecule and red to structures of the unprotected one. A complete list of the NOEs, and NOE bound violations can be found in the Supporting Information. In Figure 3, where the experimental data comes from the protected molecule, the ^3J-coupling constants for the protected molecule (black symbols) compare well with experiment. ^3J-coupling constants for the unprotected peptide (red symbols), do not differ much from those calculated for the protected form of the molecule. In Figure 4, where the experimental data originates from the unprotected molecule, the calculated ^3J-values do agree better with experiment than in the previous case. Surprisingly, considering both the protected and unprotected forms of the molecules with the corresponding experimental ^3J-values, both MD simulations agree better with experiment than the set of NMR model structures derived from the data.

The panels on the left of Figures 3 and 4 show the NOE distance bound violations. The MD trajectory of the protected molecule at 298 K satisfies most of the 74 NOE distance bounds measured for this molecule (Figure 3, black bars in the upper panel). Only four violations longer than 0.1 nm of NOE bounds are present: 1Hβ−3Hα_{Re} (NOE number 5), 2HN−3Hδ (NOE number 54) and 3HN−1Hβ (NOE number 57) and 5HN−2Hβ_{Re} (NOE number 63), where the residue sequence number is given in front of the atom name. At 340 K only the first three violations are still larger than 0.1 nm. The set of NMR model structures shows two violations larger than 0.1 nm, the second and forth ones mentioned above. The patterns of NOE distances as observed for the two MD trajectories and the set of NMR model structures are rather similar. Comparing these patterns for the unprotected (red bars) and the protected (black bars) trajectory structures, the same observation holds. The data for the protected molecule cannot really distinguish between the MD or NMR structures of the protected molecule on one hand and of the unprotected molecule on the other. When considering the NOE data for the unprotected molecule (Figure 4), the same conclusion cannot be drawn. Most of the 65 NOE distance bounds are satisfied by the MD trajectories and NMR model

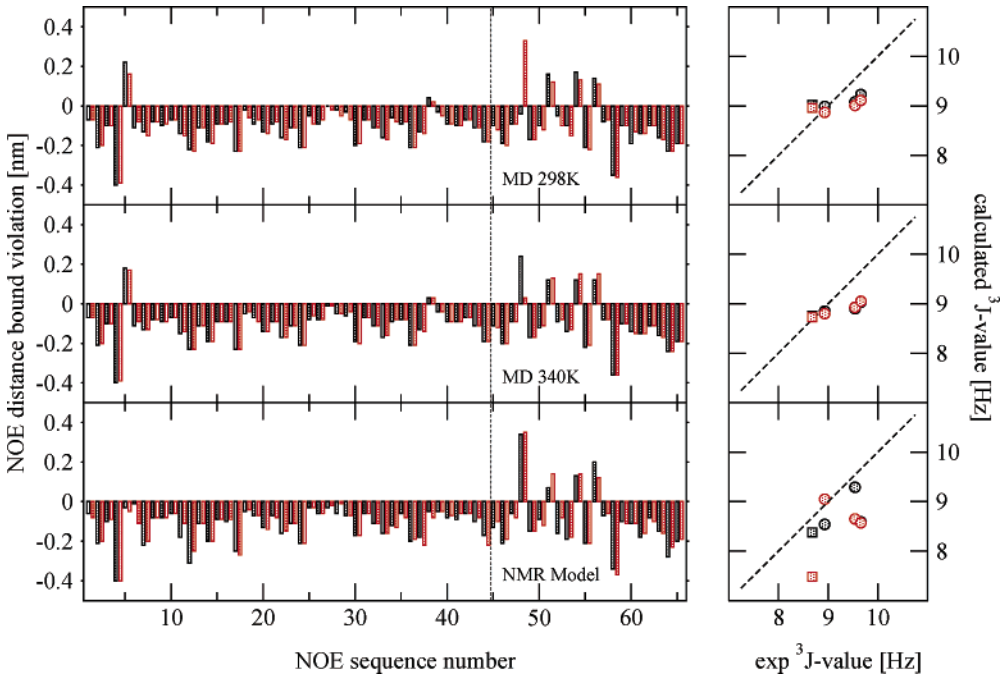


Figure 4. NOE distance bound violations (left-hand panels) and ^3J-coupling constants (right-hand panels). The experimental NMR data (65 NOE bounds and 4 ^3J-coupling constants) comes from the **unprotected** molecule.[32] The topmost panels refer to the 298 K simulation, the intermediate panels to the 340 K simulation and the bottom panels to the set of model structures derived by structure refinement based on NMR data.[32] The black color represents structures of the protected molecule while red stands for structures of the unprotected form. In the NOE distance violation panels the dashed line divides the graphs such that the NOEs to the left were measured for both the protected and the unprotected molecules (44 NOEs) and those to the right were only observed for the unprotected peptide. In the ^3J-coupling constant panels the dashed line represents perfect correlation between measured and calculated values. The circles are ^3J-values that were measured for both peptides whereas the squares were measured only for the **unprotected** form. See also the caption of Figure 3.

structures of both protected and unprotected molecules, five NOEs show violations larger than 0.1 nm: 1Hβ−3Hα_{Re} (NOE number 5), 2HN−5Hβ (NOE number 48), 4HN−5Hα_{Si} (NOE number 51), 6HN−3Hβ (NOE number 54) and 7HN−4Hβ_{Re} (NOE number 56). Of these, a single long-range NOE (number 48), 2HN−5Hβ , stands out. At 298 K, it is violated by the MD trajectory of the unprotected molecule; at 340 K, it is violated by the MD trajectory of the protected molecule, and it is violated by the NMR model structures of both molecules. Interestingly, this is one of the three weak NOE signals measured only for the unprotected molecule that were omitted in the NMR structure refinement (NOEs 48, 54, 56) because of their incompatibility with the *12/10*-helix.[32] The violation of this NOE bound is correlated with the absence of 3_{14}-helical structures or hydrogen bonds in the MD trajectories (Figure 2 and Table 1). The rank order of the 4 MD trajectories with respect to the absence of 3_{14}-helical content is (i) unprotected at 298 K, (ii) protected at 340 K, (iii) unprotected at 340 K, and (iv) protected at 298 K. The same rank order is observed with respect to the size of the violation of NOE bound number 48 in Figure 4: (i) unprotected at 298 K shows the largest violation (red bar in the upper panel), (ii) protected at 340 K shows the second largest violation (black bar in the middle panel), (iii) unprotected at 340 K shows a minor violation, and (iv) protected at 298 K show no violation. The violations of NOEs 54 and 56 are small and not correlated with the absence of 3_{14}-helical structures.

To check the relation between NOE 48 and a 3_{14}-helical conformation further, a conformational cluster analysis was performed and the distance 2HN-5Hβ determining this NOE was calculated for each cluster. This kind of conformational analysis has been described elsewhere.[18] The 2HN-5Hβ distance distribution for all the trajectory structures was computed and the results for the four most populated conformational clusters (numbered from 1 to 4) are depicted in Figure 5. The vertical lines represent the experimental NOE distance bound (dotted line) and the corresponding $\langle r^{-6}\rangle^{-1/6}$ distances for the 298 K (solid lines) and 340 K (dashed lines) simulations. The black color stands for the protected peptide and the red for the unprotected one. Most of the clusters shown represent a complete or partial *12/10*-helix, where partial means that some hydrogen bonds that determine the full helix (see Figure 1, upper panel) are missing. In all those cases, the 2HN−5Hβ r^{-6} averaged distance lies beyond the NOE distance so the NOE is violated. The only two cases in which the cluster is either a full or partial 3_{14}-helix are cluster 2 of the protected molecule at 298 K (solid black line) and cluster 4 of the unprotected molecule at 340 K (dashed red line) which fulfill the NOE bound. When adding up all the contributions for all clusters (bottom panel), we see that the bumps around 0.35 nm due to the 3_{14}-helical structures make the total r^{-6} averaged distance (nearly) satisfy the NOE for those two simulations. In all other cases where the 3_{14}-helix is not or barely sampled and so does not bring enough weight to the total average, the NOE is violated. These results are a striking example of how the presence of pertinent different conformations may complicate the interpretation of NMR data and may invalidate standard single-structure refinement. Not only could the unrestrained MD simulations largely reproduce the experimental data, but they also provided conformational insights and explained seeming inconsistencies between different NOEs.

So far, we have looked whether the available structures fulfill the experimentally observed NOEs. This test is normally the only one used to assess how good structures are. However, an

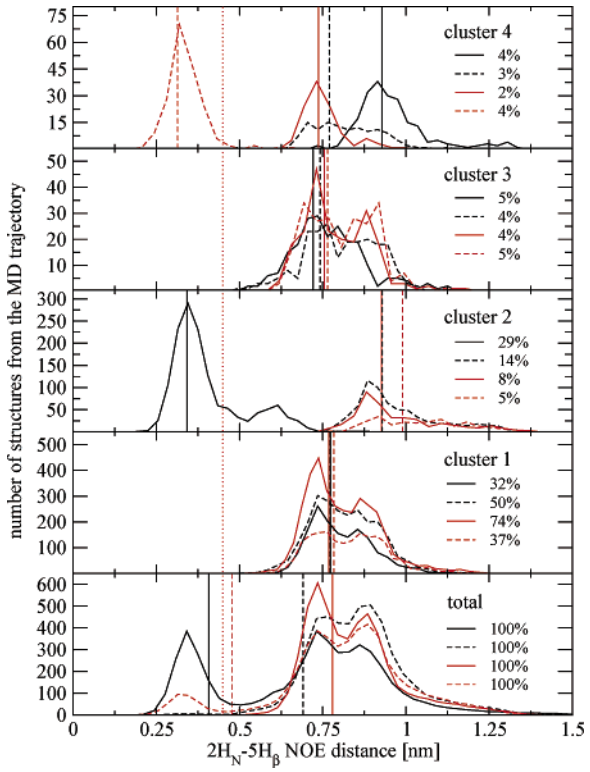


Figure 5. Distance distributions for the proton pair 2HN−5Hβ (NOE number 48 measured only for the unprotected form) calculated from four 100 ns MD trajectories: black color refers to the protected molecule and red to the unprotected form. Solid lines are designated to 298 K simulations and dashed lines to 340 K. The red dotted vertical line represents the experimentally determined NOE bound. The other vertical lines represent the $\langle r^{-6}\rangle^{-1/6}$ averages of the respective distributions. Bottom panel: total distribution based on all trajectory structures. Other panels: distributions for the four most populated conformational clusters. The percentage populations are indicated.

additional check would be to measure all possible hydrogen−hydrogen r^{-6} averaged distances and check whether we are or are not predicting NOEs that are not seen experimentally. In other words, if two hydrogen atoms among the structures show an r^{-6} averaged distance closer than, let us say 0.3 nm, a corresponding NOE signal should appear, if not obscured by phenomena such as peak overlap, spin diffusion or extinction due to rotational tumbling. Typically one-fifth of the NOE cross-peaks cannot be unambiguously interpreted because of overlap between different peaks or because of signal contamination by TOCSY transfers or chemical exchange contributions. Therefore, the absence of predicted NOEs from the list of observed NOEs is not always significant. This means that the following analysis should be read with these caveats in mind. Table 2 labels all hydrogen atoms in the β-peptide. A r^{-6} averaged distance calculation has been done for all possible pairs of hydrogen atoms except hydrogen 1 (NH). In Figures 6 and 7 the results are shown depending on the value obtained: red if the distance was smaller than 0.3 nm, green if it is between 0.3 and 0.35 nm, and blue if it is between 0.35 and 0.45 nm. These figures allow one to distinguish the short range contacts (close to the main diagonal) and long range ones, which characterize the secondary structure. Figure 6 shows that for both molecules the MD simulations and the NMR model structures predict many more NOEs peaks than are actually observed. As mentioned, this may be due to a variety of causes, which will not be discussed here. The two upper panels show the maps for structures taken from clusters 1 and 2 for the MD simulation at 298 K that represent *12/10*- and 3_{14}-helices, respectively. Only 8 structures from each cluster were considered because this is the number of NMR model structures available for the protected molecule. The close contacts look very similar, as expected. In cluster 2 the long-range ones lie further away from the main diagonal, because in the 3_{14}-helix the rings in the structure are wider and thus bring hydrogen atoms farther along the chain closer to each other than in the *12/10*-helix. We also notice that NOE 48 (2HN−5Hβ, pair 8−30) appears only in the 3_{14}-helix being a weak intensity contact. The distance maps for the NMR derived model structures are presented in the middle panels. Both maps show the pattern of a *12/10*-helix with great

Table 2. All Hydrogen Atoms of the β-Peptides, Identified by Residue Number, Hydrogen Name and Hydrogen Sequence Number as Used in Figures 6 and 7[a]

residue no.	hydrogen name	hydrogen sequence no.r	residue no.	hydrogen name	hydrogen sequence no.	residue no.	hydrogen name	hydrogen sequence no.
1	HN	1	4	HN	22	7	HN	43
1	HB	2	4	HB1	23	7	HB	44
1	HA1	3	4	HB2	24	7	HA1	45
1	HA2	4	4	HA	25	7	HA2	46
1	HG	5	4	HG	26	7	HG	47
1	HD1*	6	4	HD1*	27	7	HD1*	48
1	HD2*	7	4	HD2*	28	7	HD2*	49
2	HN	8	5	HN	29	8	HN	50
2	HB1	9	5	HB	30	8	HB1	51
2	HB2	10	5	HA1	31	8	HB2	52
2	HA	11	5	HA2	32	8	HA	53
2	HG*	12	5	HG*	33	8	HG*	54
3	HN	13	6	HN	34	9	HN	55
3	HB	14	6	HB1	35	9	HB	56
3	HA1	15	6	HB2	36	9	HA1	57
3	HA2	16	6	HA	37	9	HA2	58
3	HG1	17	6	HG1	38	9	HG1	59
3	HG2	18	6	HG2	39	9	HG2	60
3	HD	19	6	HD	40	9	HD	61
3	HE1*	20	6	HE1*	41	9	HE1*	62
3	HE2*	21	6	HE2*	42	9	HE2*	63

[a] Methyl hydrogen atoms that are treated as one pseudo atom are indicated by an asterisk.

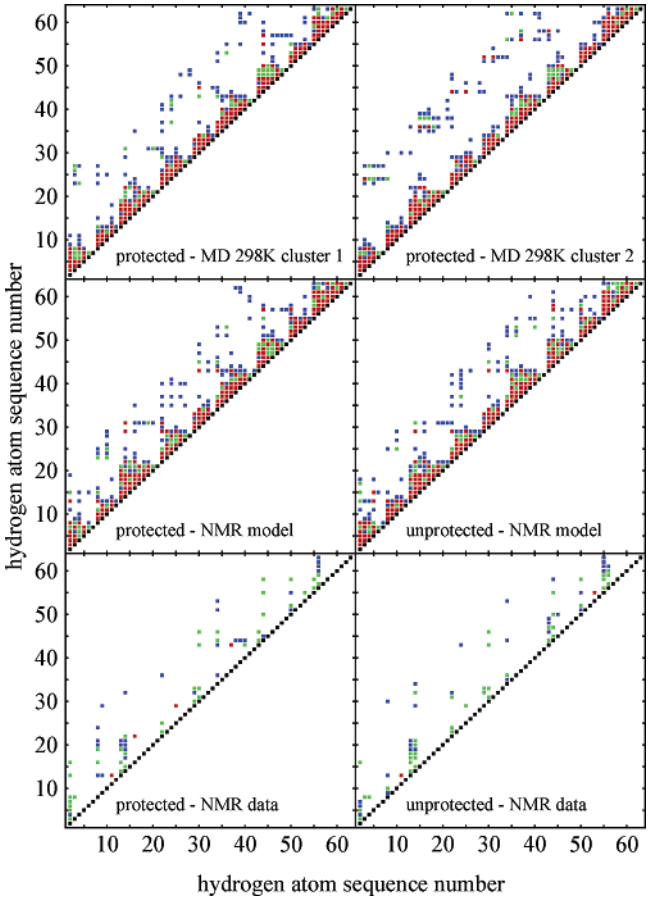


Figure 6. Inter-hydrogen r^{-6} averaged distance maps for all hydrogen atoms (excluding terminal groups) in the β-peptides. The upper panels show the results for (eight) *12/10*- and (eight) *14*-helical structures taken from cluster 1 and from cluster 2, respectively, of the MD simulation at 298 K of the protected molecule. The middle panels show the results for NMR model structures for the protected and unprotected molecules (8 and 13 structures respectively). The bottom panels show the experimentally derived distance bounds of the NMR experiment. The hydrogen sequence numbers on the axes are defined in Table 2. The color code used is the following: red if the distance is below 0.3 nm (corresponds to a strong NOE), green if it is between 0.3 and 0.35 nm (medium) or blue if it lies between 0.35 and 0.45 nm (weak). The black squares are placed along the diagonal.

similarity to the equivalent one discussed before. The two lower panels display the NOE distance bounds for the protected and unprotected molecules. Comparing in detail all panels to one another, the discrepancies between predictions of NOEs can be found. Here we analyze only the protected molecule and short contacts (strong NOEs, red dots), while excluding intraresidue hydrogen atom pairs. The *12/10*-helix from the MD simulation (upper left panel) predicts no strong NOE that is not seen at all experimentally or is not arising from hydrogen atoms of neighboring residues on the backbone. The three red dots seen in the long range portion of the graph, for instance, are seen experimentally as medium intensity NOEs. The NMR model structures (middle left panel), predict a few more strong NOEs, e.g., two that involve hydrogen atoms in residues three and five. These contacts are also present in the map of the *12/10*-helix (cluster 1) taken from the simulation, but not as strong ones. In the case of the unprotected molecule, the result is similar. While 13 structures taken from the first cluster of the simulation at 298 K only predict a few strong long-range NOEs, the NMR derived model structures predict a strong NOE between residues one and three. Figure 7 shows the distance maps calculated over

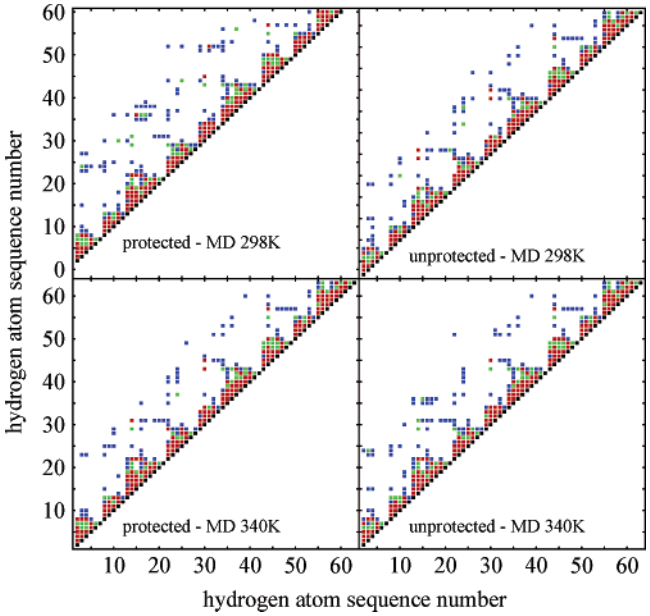


Figure 7. Inter-hydrogen r^{-6} averaged distance maps for all hydrogen atoms (excluding terminal groups) in the $\beta-$peptides as obtained from the 100 ns MD simulations. The upper panels show the results for the protect and unprotected forms of the peptide at 298 K while the bottom panels show the corresponding results at 340. See also caption of Figure 6.

the whole trajectories for all the simulations. All of them exhibit the *12/10*-helix pattern described before except the simulation of the protected molecule at 298 K, which rather shows a mixture of *12/10*- and *14*-helix patterns, as could be expected. The predicted strong contacts of the NMR model mentioned above are present in all panels with different averaged distances.

The conformational clustering analysis in combination with a NOE distance calculation for each cluster provides indications of which conformations of the ensemble are relevant to the reproduction of the experimental data. To know whether the many different structures sampled during the simulations are common to both forms of the molecule, a combined clustering analysis (over the combined trajectories) was done. Such an analysis indicates how much of the conformational spaces sampled by each of the simulations for the protected and unprotected form overlap. For each conformational cluster found, its type of helix was determined according to the appearance (>5%) of a particular type of hydrogen bond, see Figure 8. The first (most populated) cluster is a pure *12/10*-helix which is adopted by both molecules. In addition, *10*-membered and to a lesser extend *12*-membered hydrogen-bonded rings appear in most of the other clusters. Few clusters of both molecules with a relative low population contain *14*-membered hydrogen-bonded rings. The pure 3_{14}-helix is found as cluster 2 in the protected form and as cluster 6 in the unprotected molecule. As expected, the overlap in the configurational space between the two molecules is bigger for the higher temperatures.

Conclusion

Molecular Dynamics simulations of the (un)folding equilibrium of the protected and unprotected forms of the β-nonapeptide shown in Figure 1 were presented with the aim of assessing how much such MD simulations can contribute to a realistic interpretation of experimental NMR data. For each form of the

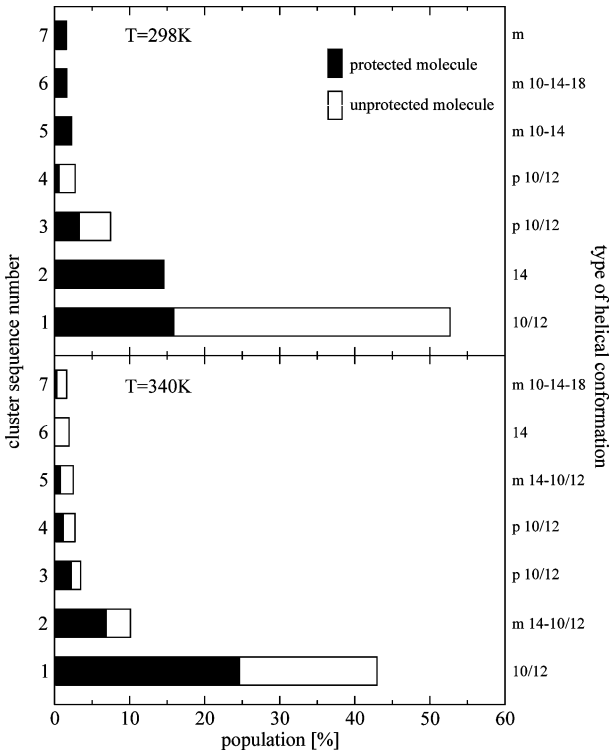


Figure 8. Relative populations of the 7 most populated conformational clusters from the combined 100 ns MD trajectories of the protected (black) and unprotected (white) molecules at 298 K (upper panel) and at 340 K (lower panel). On the right axis the type of helix is indicated: the number is the type of hydrogen-bonded rings (with a population larger than 5% present), *m* stands for "mixed" and *p* for "partial".

molecule two 100 ns long simulations at 298 and 340 K were performed and the structures sampled were compared to primary experimental data such as proton−proton NOE intensities and ^3J-coupling constants and to the NMR model structures (secondary data) that had been obtained by standard single-structure NMR refinement. The primary experimental NMR data, NOE distance bounds and ^3J-values were well reproduced by the unrestrained MD simulations, equally well as these primary data were reproduced by the sets of NMR model structures. Moreover, the MD trajectories seem to predict fewer strong long-range NOEs than the NMR model structures. The molecular structures were analyzed in terms of hydrogen bonds and particular helical conformations. The peptide had been expected to adopt a *12/10*-helix according to the NMR data and model structures, which was indeed observed in the simulations. However, the MD trajectories showed that other conformations were important as well. A 3_{14}-helix was sampled by both, the protected and unprotected, peptides indicating that this conformation must have some weight when evaluating trajectory or ensemble averages of observables such as NOE intensities and ^3J-coupling constants. The possibility of other than *12/10*-helical conformations had been deliberately excluded in the NMR single-structure refinement leading to the model structures, because a single dominant conformation had been assumed and so three NOEs for the unprotected molecule mentioned to be inconsistent with the *12/10*-helix were not included in the original structure determination based on simulated annealing. The unrestrained MD simulations, however, could resolve the seeming inconsistencies between the NOE intensities. It appeared that a small admixture of 3_{14}-helical conformation to the predominant *12/10*-helical conformation in the ensemble would reproduce the NMR data.

The dominant *12/10*-helical conformation seems to be slightly more stable for the protected molecule than for the unprotected one, in agreement with earlier analysis[32] of the relative NOE intensities for the various proton pairs. However, such an analysis could not provide information about alternative conformations being present in the structural ensembles for the molecules, because only (nonlinear) averages of observables over the ensemble of solution structures can be measured and a variety of different conformational distributions may lead to the same average. MD simulations provide not only averages, but also conformational distributions, which can be analyzed to obtain the weight of particular, e.g., *12/10*- or 3_{14}-helical conformations. Moreover, the necessarily different sensitivities of measured, nonlinearly averaged, observables to different conformations in the ensemble can be straightforwardly determined. For the β-nonapeptide, we showed that the intensity of one particular long-range NOE, the proton pair 2HN−5Hβ, is highly correlated with the admixture of 3_{14}-helical conformation in the ensemble of solution structures. This leads to the conclusion that although both forms of the peptide adopt predominantly a *12/10*-helical conformation in solution, the unprotected one also adopts a 3_{14}-helical conformation to a small extent.

The presented MD simulations of the β-peptide illustrate the increasing usefulness of MD simulation of bio-molecules from a practical point of view.

1. Unrestrained MD simulation of bio-molecules in explicit solvent using a thermodynamically calibrated force field in which solute and solvent are consistent with each other, can reproduce experimental data, such as NOE intensities and ^3J-values rather well;

2. The nonlinear ensemble averages over the unrestrained MD trajectories that correspond to the mentioned observables may agree equally well with the experimental data averages over model structures obtained from standard single-structure refinement that uses the experimental data as structural restraints;

3. When the conformational ensemble of a molecule in solution is dominated by more than one conformation, standard single-structure refinement against experimental data that reflect multiple conformations will lead to inconsistent or unlikely structures, whereas unrestrained MD simulation will generate the proper ensemble provided the use of a high-quality force field, explicit treatment of solvent degrees of freedom, and sufficient sampling;

4. Analysis of the sensitivity of nonlinear averages corresponding to measurable observables to the underlying conformational (Boltzmann) distribution as generated by MD simulation may lead to identification of those observables that are most sensitive to particular conformations. The small conformational differences between the protected and unprotected forms of the β-nonapeptide could be identified by such an analysis.

In summary, unrestrained MD simulation using a consistent, high-quality force field for both solute and solvent, and including explicitly solvent degrees of freedom can contribute significantly to a correct interpretation of experimental data in terms of conformational distributions.

Acknowledgment. We thank Prof. Dieter Seebach for suggesting to simulate the peptide. We also thank Dr. Bojan Zagrovic for valuable discussions on the inter-hydrogen distance calculations. Financial support by the National Centre of Competence in Research (NCCR) in Structural Biology of the Swiss National Science Foundation (SNF) is gratefully acknowledged.

Supporting Information Available: The complete list of NOEs for both β-peptides shown in Figure 3 and Figure 4 is available. This material is available free of charge via the Internet at http://pubs.acs.org.

JA044285H