Peptide Solution Structures

Are NMR-Derived Model Structures for β-Peptides Representative for the Ensemble of Structures Adopted in Solution?**

*Alice Glättli and Wilfred F. van Gunsteren**

Connecting an NMR observable $\vec{q}^{\,obs}$, such as an interproton distance derived from nuclear Overhauser effect (NOE) intensity or a ^3J-coupling constant, with the underlying conformational ensemble $\{\vec{r}\}$ is a long-standing problem for the structure determination of peptides, proteins, and other biomolecules in solution.[1–9] In particular, for flexible molecules such as peptides the assumption that all NMR signals originate from the same predominant conformer may not hold.[10–13] In that case conventional NMR structure refinement procedures are not sufficient for the correct interpretation of the experimental data.

Molecular dynamics (MD) simulation can often complement the experimental tools for biomolecular structure determination such as NMR and circular dichroism spectroscopy.[8,9,11,14,15] A sufficiently accurate simulation for comparison with and interpretation of experimental data requires:

1) the choice of the essential degrees of freedom appropriate to model the system of interest and to calculate the desired experimental observable,
2) a physically calibrated force field to describe the interactions along and between the chosen degrees of freedom,
3) equations of motion or a sampling method that generates a Boltzmann-weighted ensemble of conformers, and
4) the knowledge of the relation $\vec{q}(\vec{r})$ between the experimental observable $\vec{q}$ and the molecular structure $\vec{r}$ of the system of interest.

If the molecular model and the atomic interaction function (i.e. the force field) were perfect, and if a simulation could be carried out for an infinitely long time, the generated ensemble would exactly represent the real molecular system and there would be no need to carry out an NMR experiment on the system of interest. Furthermore, the connection between experimentally measured observables and the underlying conformational ensemble would be perfectly understood, provided that the relation $\vec{q}(\vec{r})$, which connects the experimentally measured observables $\vec{q}$ to a conformation $\vec{r}$, is exactly known. In practice, however, none of these conditions are completely fulfilled. Despite the enormous increase in computation power in recent years, the time required to sample the complete populated conformational space of a biomolecule is still beyond affordable simulation (sampling) for all but the shorter peptides. Additionally, even though current force fields for biomolecules are steadily improving in correctly modeling the structure, dynamics, and energetics of biomolecular systems,[16] they are based on various approximations concerning polarizability, many-body interactions, quantum effects, etc. On the other hand, conventional structure determination makes use of force fields because the number of experimental observables is generally too small to uniquely derive the underlying set of conformations. Here, we compare two methods for the conformational interpretation of the NMR data of a β-hexapeptide (Figure 1):

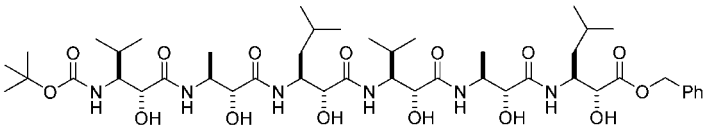

Figure 1. Chemical formula of the hydroxy β-hexapeptide studied. The peptide is protected by a *tert*-butoxycarbonyl group (N-Boc) at the N terminus and by a carbobenzoxy group (Z) at the C terminus. The hydroxy groups are attached to the C_α atoms and the side chains to the C_β atoms of the backbone.

firstly, conventional structure determination by simulated annealing in vacuo applying NOE distance and dihedral-angle restraints using a simple force field widely employed for structure determination of polypeptides, and secondly, free (unrestrained) MD simulation, which inherently generates Boltzmann-weighted ensembles, using a physically, thermodynamically calibrated force field and explicit treatment of solvent degrees of freedom. We show that the two methods yield rather different results and that the former one may lead to erroneous interpretation of NMR experiments.

The β-hexapeptide under investigation has been suggested by conventional NMR structure refinement to adopt a (*P*)-2_8-helical conformation in solution[17] (Figure 2), representing the first example of the fourth helical secondary structure element adopted by β-peptides. By forming intramolecular hydrogen bonds, β-peptides adopt secondary structure elements that are very similar to those found in α-peptides and proteins,[18,19] namely left- and right-handed helices (3_{14}-,[20–22] 2.5_{12}-,[23] 10/12-,[24] and the previously mentioned 2_8-helix[17]), turns, and sheets.[25] Starting from a fully extended conformation, the β-hexapeptide displayed in Figure 1 was simulated in explicit methanol solution at two different temperatures (298 K and 340 K) and at constant pressure (1 atm) using the GROMOS simulation package[26,27] and the GROMOS biomolecular force field (version 45A3[26,28]). For simulation details, see the Supporting Information. The ensembles of structures from the two 100-ns

[*] A. Glättli, Prof. W. F. van Gunsteren
Laboratorium für Physikalische Chemie, ETH
ETH Hönggerberg, HCI, 8093 Zürich (Switzerland)
Fax: (+41) 1-632-1039
E-mail: wfvgn@igc.phys.chem.ethz.ch

[**] We thank Prof. D. Seebach for challenging us to simulate the behavior of the peptide and Dr. K. Gademann for providing us with the 20 NMR model structures. Financial support from the Schweizer Nationalfonds (project number 2000-063590.00) and from the National Center of Competence in Research (NCCR) in Structural Biology of the Swiss National Science Foundation is gratefully acknowledged.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

  DOI: 10.1002/ange.200460384

Figure 2. Superposition of the 20 NMR model structures with lowest energy from the ab initio simulated-annealing structure refinement runs[17] with the 40 upper-bound distance restraints and 12 torsional-angle restraints derived from NMR data at 298 K in methanol.[17] The simulated annealing runs were performed using the X-PLOR program and force field.[29] The model structures suggest that the peptide adopts a (*P*)-2_8-helical conformation. The structures are superimposed using the backbone atoms of residues 2–5. The protection groups are omitted for clarity. Color scheme: C = yellow, H = white, N = blue, O = red.

trajectories were analyzed regarding the level of agreement with the NMR-derived data and regarding the conformational space sampled by the peptide in the simulation compared to the one covered by the 20 NMR model structures obtained by standard structure refinement using the program X-PLOR.[29]

In Figure 3 the simulated and the NMR-derived data are compared in terms of 40 NOE distances and 12 ^3J-coupling constants. The average effective violations of the upper-bound distances from all recorded structures in both simulations and from the 20 X-PLOR structures are displayed in Figure 3 A–C. The upper-bound nature of NOE-derived distances implies that only violations with positive values are true violations. Essentially all experimentally observed NOEs are satisfied both in the X-PLOR structures and in the simulations at 298 and 340 K. The ensemble of structures at 298 K marginally violates two interresidue NOE distances: $HC_\gamma(2)/HC_\beta(1)$ (NOE sequence number 3) and NH(2)/NH(3) (NOE sequence number 16) are both violated by 0.05 nm. The simulation at 340 K violates only one NOE distance (sequence number 3) by 0.03 nm. We also checked whether long proton–proton distances correspond to the absence of measured NOEs, although the latter may have other causes such as fast exchange, a particular overall tumbling time, and extensive spin–spin splitting. The 12 experimentally measured ^3J-coupling constants are compared to the average ^3J-coupling constants calculated for the trajectory structures from the simulations at 298 and 340 K and for the 20 X-PLOR structures using the Karplus relation[30] as shown in Figure 3 D–F. With an average absolute deviation of 0.4 Hz, the average ^3J-coupling constants calculated from the ensemble of structures generated at room temperature agree very well with the experimentally measured values. Only one deviation exceeding 1 Hz, for HN–HC_β of residue 1, is observed. The 20 X-PLOR structures agree slightly worse with the experimental values than the simulation at room temperature with deviations of more than 1 Hz for the H-C_β-C_α-H torsions of residues 1 and 4, resulting in an average absolute deviation of 0.6 Hz.

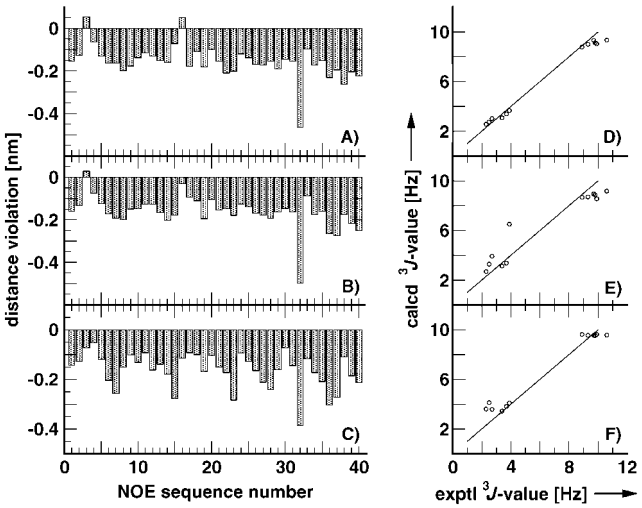

Figure 3. Panels A–C: Average (r^{-6}-averaging) distance violations of the upper-bound distances over all the recorded structures (2×10^5) in the MD simulations at 298 (A) and 340 K (B) and the over the 20 X-PLOR structures[17] (C). The upper-bound distances are inferred from 40 experimental NOE intensities observed in the ROESY NMR spectrum at 298 K.[17] Panels D–F: Comparison of the 12 ^3J-coupling constants (6 HN–HC_β, 6 HC_β–HC_α) extracted from the one-dimensional ^{1}H NMR spectrum measured at 298 K[17] with the corresponding calculated ^3J-coupling constants averaged over all structures from the MD simulations at 298 K (D) and at 340 K (E) and averaged over the 20 X-PLOR structures[17] (F). The Karplus equation[30] with $a = 6.4$ Hz, $b = -1.4$ Hz, and $c = 1.9$ Hz[40] for ^3J(HN,HC) and with $a = 9.5$ Hz, $b = -1.6$ Hz, and $c = 1.8$ Hz[41] for ^3J(HC,HC) was used. Tables of the experimental NOE upper-bound distances and ^3J-coupling constants, of NOE distance bound violations of the r^{-6}-averaged distances and the calculated ^3J-coupling constants averaged over the X-PLOR structures and the structures from the MD simulations are included in the Supporting Information.

Based on purely geometric criteria, the occurrence of hydrogen bonds in the 20 X-PLOR structures and in the ensemble of structures recorded in the simulations at 298 and 340 K has been determined (Table 1). Interestingly, eight-membered hydrogen-bonded rings (HB_8), characteristic for a (*P*)-2_8-helix, appear only at very low percentages in the two simulations ($\leq 4\,\%$), while they are to some extent present in the X-PLOR bundle of structures. This indicates that the suggested (*P*)-2_8-helix is only scarcely, if at all, populated in the simulations at 298 and 340 K. The hydrogen-bond analysis also shows that no regular secondary structure is sampled at room temperature, while at elevated temperature the occurrence of four 12-membered hydrogen-bonded rings hints at the formation of a (*P*)-2.5_{12}-helix. Additional simulations starting from either a (*P*)-2.5_{12}-helix or one of the 20 X-PLOR structure, which are suggested to represent a (*P*)-2_8-helix,[17] show the (*P*)-2.5_{12}-helical conformation to be stable once the peptide has adopted it, while the X-PLOR structures are rather unstable (see the Supporting Information). In addition to the backbone–backbone hydrogen bonds, various hydrogen bonds between the α-hydroxy hydrogen and carbonyl oxygen atoms are present at the two temperatures. However, hydrogen bonds of the type OH(*i*)–O(*i*) (HB_5), proposed to

stabilize the (*P*)-2_8-helical conformation,[17] are observed neither in the simulations nor in the 20 X-PLOR structures.

The results of the conformational clustering analysis over the combined ensembles (10000 structures from each simulation (at 298 K and 340 K) and the 500 copies of each of the 20 X-PLOR structures) are displayed in Figure 4. The conformational clustering analysis groups the structures of the ensembles according to their positional root-mean-square deviation (rmsd) of the backbone (N, C_β, C_α, C) atoms (excluding the first and last residue with the protecting groups) between each other structure favoring the most populated cluster.[31] When a rather stringent rmsd similarity criterion of 0.04 nm is used (Figure 4A), the three ensembles do not share any part of conformational space; each cluster is populated by members of only one of the three ensembles. With a larger rmsd criterion of 0.08 nm the ensembles slightly overlap at their peripheries (Figure 4B). Cluster 8 represents the only cluster containing structures from all three ensembles (MD at 298 K: 12%; MD at 340 K: 12%; X-PLOR: 76%). Figure 4 also illustrates that the Boltzmann-weighted ensembles generated by free (unrestrained) molecular dynamics simulations bear a much larger conformational variability than the bundle of structures obtained by restrained simulated annealing.

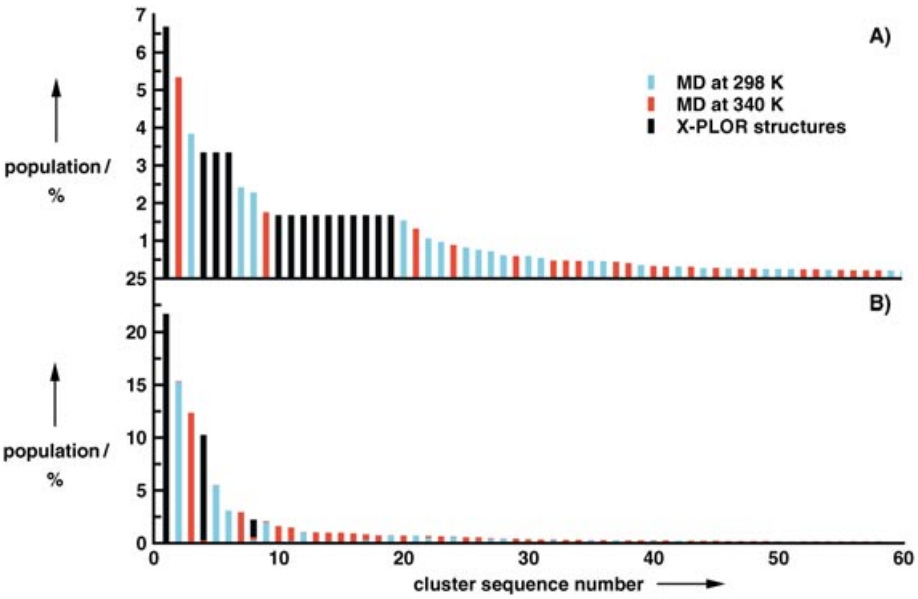

Figure 4. Conformational clustering analysis combining the "ensemble" of 500 copies of each of the 20 NMR model structures (i.e. equally weighing each NMR model structure) with the ensembles of 10000 structures each (1 per 10 ps) sampled in the MD simulations at 298 and at 340 K. The plots show the population in percentage per cluster (conformers) and the portion of structures per cluster belonging to the "ensemble" of NMR model structures (black) and to the ensemble of structures generated at 298 K (blue) and at 340 K (red) by unrestrained MD simulation. In order to illustrate the conformational spread of the various ensembles, two different rmsd-similarity criteria are used: A more stringent one of 0.04 nm (A) and the standard criterion for a β-hexapeptide of 0.08 nm (B). With the first criterion a total of 2772 clusters are found, of which clusters 1–20 represent more than 50%, clusters 1–120 more than 75%, clusters 1–625 more than 90%, and clusters 1–1272 more than 95% of the total population. With the second criterion a total of 237 clusters are found, of which clusters 1–4 represent more than 50%, clusters 1–10 more than 75%, clusters 1–28 more than 90%, and clusters 1–51 more than 95% of the total population. For both rmsd criteria, only the populations of the first 60 most populated clusters are shown.

Table 1: Occurrence of intramolecular hydrogen bonds. A hydrogen bond is considered to exist when the donor-hydrogen-acceptor angle is larger than 135° and the hydrogen–acceptor distance is less than 0.25 nm. The hydrogen bonds are grouped according to the type of hydrogen-donor (NH or OH group) and the size (in terms of number of atoms) of the resulting hydrogen-bonded ring (e.g. a hydrogen bond between NH of residue *i* and C=O of residue (*i*−2) results in an eight-membered hydrogen-bonded ring, denoted as HB_8). O(0) corresponds to the carbonyl oxygen of the Boc protecting group. Only hydrogen bonds with a population larger than 10% in one of the sets of structures are shown.

	Occurrence of hydrogen bonds [%]		
	MD simulation		Refinement
Donor–Acceptor	298 K	340 K	X-PLOR
NH(*i*)–O(*i*−2) [HB_8]			
NH(3)–O(1)	0	1	20
NH(4)–O(2)	0	1	25
NH(5)–O(3)	2	4	10
NH(*i*)–O(*i*−3) [HB_{12}]			
NH(3)–O(0)	0	30	0
NH(4)–O(1)	0	26	0
NH(5)–O(2)	0	35	0
NH(6)–O(3)	1	18	0
NH(*i*)–O(*i*+1) [HB_{10}]			
NH(2)–O(3)	11	0	0
NH(5)–O(6)	11	1	0
OH(*i*)–O(*i*−1) [HB_7]			
OH(6)–O(5)	0	14	0
OH(*i*)–O(*i*−2) [HB_{11}]			
OH(4)–O(2)	0	8	10
OH(5)–O(3)	1	22	0
OH(6)–O(4)	1	10	0
OH(*i*)–O(*i*−3) [HB_{15}]			
OH(4)–O(1)	1	26	0
OH(5)–O(2)	0	10	0
OH(*i*)–O(*i*+2) [HB_{13}]			
OH(3)–O(5)	38	0	0

Summarizing, we find that even though the 20 model structures obtained from simulated annealing and the two ensembles generated by free MD simulations show different hydrogen-bond patterns and access different parts of the conformational space of the peptide, they all agree with the available NMR data. First, this indicates that for a given set of experimental observables, depending on the structural properties of a peptide, more than one solution structure is possible. Consequently, a single structure may not be representative for the ensemble of structures in solution.

Second, this study demonstrates that unbiased MD simulation using a thermodynamically calibrated force field reproduces experimental NMR data such as NOE upper-bound distances and ^3J-coupling constants just as well or even better than a set of NMR model structures derived by classical single-structure simulated-annealing refinement techniques using a simple force field in vacuo and the NOE upper bounds and J-value derived torsional-angle values as restraints. The fact that the MD simulations using the GROMOS force field do not need $40+12$ restraints to satisfy the NMR data on the peptide demonstrates the accuracy of this force field and of the inclusion of explicit solvent compared to results from the use of the X-PLOR force field in vacuo. Therefore, standard NMR refinement procedures for flexible molecules such as small peptides as well as for proteins should be revised by completing the refinement process with molecular dynamics simulation in explicit solvent with a thermodynamically calibrated force field in order to generate a proper Boltzmann-weighted ensemble of structures. This should lead to a more reliable structural interpretation of the experimentally measured NMR observables. In the present case of the β-hexapeptide the MD simulations show that a (P)-2_8-helical conformation is not stable and therefore probably not representative for the ensemble of solution structures. It cannot be excluded that the finding that the simulation using the GROMOS force field prefers the formation of a (P)-2.5_{12}-helix over a (P)-2_8-helix might be an artifact of the force field. Yet, the simulations agree with the experimental data and the GROMOS force field has very well reproduced experimental findings in previous peptide folding investigations.[11,14,15,31–39] Finally, we note once more that the comparison of modeling or simulation results with experiment should always be done with primary, measured data such as NOE intensities, and maybe distances, or ^3J-values and not only with secondary, derived data such as molecular structures and torsional-angle values in order to avoid spurious conclusions regarding (dis)agreement with experimental data.

Received: April 20, 2004

Keywords: β-peptides · conformation analysis · molecular dynamics · NMR spectroscopy · structure elucidation

[1] O. Jardetzky, *Biochim. Biophys. Acta* **1980**, *621*, 227 – 232.
[2] J. Tropp, *J. Chem. Phys.* **1980**, *76*, 6035 – 6043.
[3] W. F. van Gunsteren, R. M. Brunne, P. Gros, R. C. van Schaik, C. A. Schiffer, A. E. Torda in *Methods in Enzymology: Nuclear Magnetic Resonance*, *Vol. 239* (Eds.: T. James, N. Oppenheimer), Academic Press, New York, **1994**, pp. 619 – 654.
[4] R. Abseher, S. Lüdemann, H. Schreiber, O. Steinhauser, *J. Mol. Biol.* **1995**, *249*, 604 – 624.
[5] A. M. J. J. Bonvin, A. T. Brünger, *J. Biomol. NMR* **1996**, *7*, 72 – 76.
[6] T. R. Schneider, A. T. Brünger, M. Nilges, *J. Mol. Biol.* **1999**, *285*, 727 – 740.
[7] C. A. E. M. Spronk, B. Sander, A. M. J. J. Bonvin, E. Krieger, G. W. Vuister, G. Vriend, *J. Biomol. NMR* **2003**, *25*, 225 – 234.
[8] A. E. Torda, W. F. van Gunsteren in *Reviews in Computational Chemistry*, *Vol. III* (Eds.: K. Lipkowitz, D. Boyd), VCH, New York, **1992**, pp. 143 – 172.
[9] W. R. P. Scott, A. E. Mark, W. F. van Gunsteren, *J. Biomol. NMR* **1998**, *12*, 501 – 508.
[10] R. Abseher, S. Lüdemann, H. Schreiber, O. Steinhauser, *J. Am. Chem. Soc.* **1994**, *116*, 4006 – 4018.
[11] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Angew. Chem.* **1999**, *111*, 249 – 253; *Angew. Chem. Int. Ed.* **1999**, *38*, 236 – 240.
[12] X. Daura, I. Antes, W. F. van Gunsteren, A. E. Mark, *Proteins Struct. Funct. Genet.* **1999**, *36*, 542 – 555.
[13] R. Bürgi, J. Pitera, W. F. van Gunsteren, *J. Biomol. NMR* **2001**, *19*, 305 – 320.
[14] C. Peter, X. Daura, W. F. van Gunsteren, *J. Biomol. NMR* **2001**, *20*, 297 – 310.
[15] A. Glättli, X. Daura, D. Seebach, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2002**, *124*, 12972 – 12978.
[16] T. Hansson, C. Oostenbrink, W. F. van Gunsteren, *Curr. Opin. Struct. Biol.* **2002**, *12*, 190 – 196.
[17] K. Gademann, A. Häne, M. Rueping, B. Jaun, D. Seebach, *Angew. Chem.* **2003**, *115*, 1573 – 1575; *Angew. Chem. Int. Ed.* **2003**, *42*, 1534 – 1537.
[18] D. Seebach, J. L. Matthews, *Chem. Commun.* **1997**, *79*, 2015 – 2022.
[19] R. P. Cheng, S. H. Gellman, W. F. DeGrado, *Chem. Rev.* **2001**, *101*, 3219 – 3232.
[20] D. Seebach, M. Overhand, F. N. M. Kühnle, B. Martinoni, L. Oberer, U. Hommel, H. Widmer, *Helv. Chim. Acta* **1996**, *79*, 913 – 941.
[21] D. Seebach, P. E. Ciceri, M. Overhand, B. Jaun, D. Rigo, L. Oberer, U. Hommel, H. Widmer, *Helv. Chim. Acta* **1996**, *79*, 2043 – 2066.
[22] D. H. Appella, L. A. Christianson, I. L. Karle, D. R. Powell, S. H. Gellman, *J. Am. Chem. Soc.* **1996**, *118*, 13071 – 13072.
[23] D. H. Appella, L. A. Christianson, D. A. Klein, D. R. Powell, X. Huang, J. J. Brachi, Jr., S. H. Gellman, *Nature* **1997**, *387*, 381 – 384.
[24] D. Seebach, S. Abele, K. Gademann, G. Guichard, T. Hintermann, B. Jaun, J. L. Matthews, J. V. Schreiber, L. Oberer, U. Hommel, H. Widmer, *Helv. Chim. Acta* **1998**, *81*, 932 – 982.
[25] D. Seebach, S. Abele, K. Gademann, B. Jaun, *Angew. Chem.* **1999**, *111*, 1700 – 1703; *Angew. Chem. Int. Ed.* **1999**, *38*, 1595 – 1597.
[26] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, vdf Hochschulverlag, ETH Zürich, Switzerland, **1996**.
[27] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, W. F. van Gunsteren, *J. Phys. Chem.* **1999**, *103*, 3596 – 3607.
[28] L. D. Schuler, X. Daura, W. F. van Gunsteren, *J. Comput. Chem.* **2001**, *22*, 1205 – 1218.
[29] A. T. Brünger, *X-PLOR. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT, USA, **1992**.
[30] M. Karplus, *J. Chem. Phys.* **1959**, *30*, 11 – 15.
[31] X. Daura, W. F. van Gunsteren, A. E. Mark, *Proteins Struct. Funct. Genet.* **1999**, *34*, 269 – 280.
[32] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *J. Mol. Biol.* **1998**, *280*, 925 – 932.
[33] W. F. van Gunsteren, R. Bürgi, C. Peter, X. Daura, *Angew. Chem.* **2001**, *113*, 363 – 367; *Angew. Chem. Int. Ed.* **2001**, *40*, 351 – 355.
[34] X. Daura, K. Gademann, H. Schäfer, B. Jaun, D. Seebach, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2001**, *123*, 2393 – 2404.
[35] C. Peter, X. Daura, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2000**, *122*, 7461 – 7466.
[36] R. Bürgi, X. Daura, A. E. Mark, M. Bellanda, B. Mammi, E. Peggion, W. F. van Gunsteren, *J. Pept. Res.* **2001**, *57*, 107 – 118.

[37] X. Daura, A. Glättli, P. Gee, C. Peter, W. F. van Gunsteren, *Adv. Protein Chem.* **2002**, *62*, 341 – 360.
[38] C. Peter, M. Rueping, H. J. Wörner, B. Jaun, D. Seebach, W. F. van Gunsteren, *Chem. Eur. J.* **2003**, *9*, 5838 – 5849.
[39] H. Yu, X. Daura, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2004**, *54*, 116 – 127.
[40] A. Pardi, M. Billeter, K. Wüthrich, *J. Mol. Biol.* **1984**, *180*, 741 – 751.
[41] A. de Marco, M. Llinás, K. Wüthrich, *Biopolymers* **1978**, *17*, 617 – 636.

