Journal of Peptide Science
J. Peptide Sci. **11**: 74–84 (2005)
Published online 15 September 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.597

JOURNAL OF PEPTIDE SCIENCE

Principles of carbopeptoid folding: a molecular dynamics simulation study

RICCARDO BARON, DIRK BAKOWIES and WILFRED F. VAN GUNSTEREN*

Laboratorium für Physikalische Chemie, ETH, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

Received 6 May 2004; Accepted 19 May 2004

Abstract: The conformational spaces of five oligomers of tetrahydrofuran-based carbopeptoids in chloroform and dimethyl sulfoxide were investigated through nine molecular dynamics simulations. Prompted by nuclear magnetic resonance experiments that indicated various stable folds for some but not all of these carbopeptoids, their folding behaviour was investigated as a function of stereochemistry, chain length and solvent. The conformational distributions of these molecules were analysed in terms of occurrence of hydrogen bonds, backbone torsional-angle distributions, conformational clustering and solute configurational entropy. While a *cis*-linkage across the tetrahydrofuran ring favours right-handed helical structures, a *trans*-linkage results in a larger conformational variability. Intra-solute hydrogen bonding is reduced with increasing chain length and with increasing solvent polarity. Solute configurational entropies confirm the picture obtained: they are smaller for *cis*- than for *trans*-linked peptides, for chloroform than for dimethyl sulfoxide as solvent and for shorter peptide chains. The simulations provide an atomic picture of molecular conformational variability that is consistent with the available experimental data.

Keywords: computer simulation; carbopeptoids; force field; GROMOS; molecular dynamics; peptide folding; sugar; tetrahydrofuran

INTRODUCTION

Explaining how proteins and peptides in solution fold to relatively stable structures is one of the most challenging problems in computational biochemistry. Whenever an atomic picture can be obtained from measurement, it appears that proteins adopt 3-dimensional structures, which enable them to function in specific ways, for example as biologically active enzymes. Shorter polypeptides generally adopt less compact and less stable structures in solution, making the experimental characterization of their preferred folds a non-trivial task. However, in the recent past, significant progress has been made theoretically and experimentally in predicting and measuring the formation of secondary structure of polypeptides of various types.

Molecular dynamics (MD) simulations have been shown to be a reliable instrument to investigate the process of reversible folding of various α- and β-peptides, using an atomistic model for both solute and solvent [1–9]. Implicit solvation models have also been applied with some success [10,11]. The dynamics of reversible folding can nowadays be simulated on the ns to μs timescale for peptides and small proteins [12–16]. Encouraging agreement with experiment has been obtained showing that state-of-the-art biomolecular force fields mimic these microscopic processes very well. Experimentally measurable quantities can be obtained from MD simulations by averaging over the generated ensemble of structures. Once agreement with experimental data has been observed, MD simulations can be used to estimate physical properties of the system not accessible to direct experimental observation.

On the experimental side, much progress has been achieved in the design and synthesis of molecules that mimic secondary structure elements of naturally occurring proteins [17–19]. For example, carbopeptoids of various chain lengths have been synthesized from furanose-based amino acid residues [20,21]. These δ-peptides may formally be constructed from α-peptides by replacement of every second $C(=O)-N-C_\alpha$ fragment with a substituted tetrahydrofuran (THF) ring. Folding preferences for this class of carbopeptoids have been studied in nuclear magnetic resonance experiments [20–22]. Oligomers with a 2,5-*cis* stereochemistry across the THF ring tend to form a right-handed helix composed of repeating β-turns and are characterized by $NH(i) - O(i-2)$ hydrogen bonds. This hydrogen bond pattern was observed for the *cis*-tetramer **1** (Figure 1a) in both chloroform and dimethyl sulfoxide (DMSO) and seems present also for the *cis*-hexamer **2** and the *cis*-octamer **3**. A *trans*-tetramer with a 2,5-*trans* stereochemistry on the THF ring was reported to have no preferred stable structure [23]. Increasing the chain length to six and eight residues (molecules **4** and **5**) changes the folding properties of *trans*-linked carbopeptoids. The formation of a rare type of left-handed helix stabilized by inter-residue

* Correspondence to: Wilfred F. van Gunsteren, Laboratorium für Physikalische Chemie, ETH, ETH-Hönggerberg, CH-8093 Zürich, Switzerland; e-mail: wfvgn@igc.phys.chem.ethz.ch

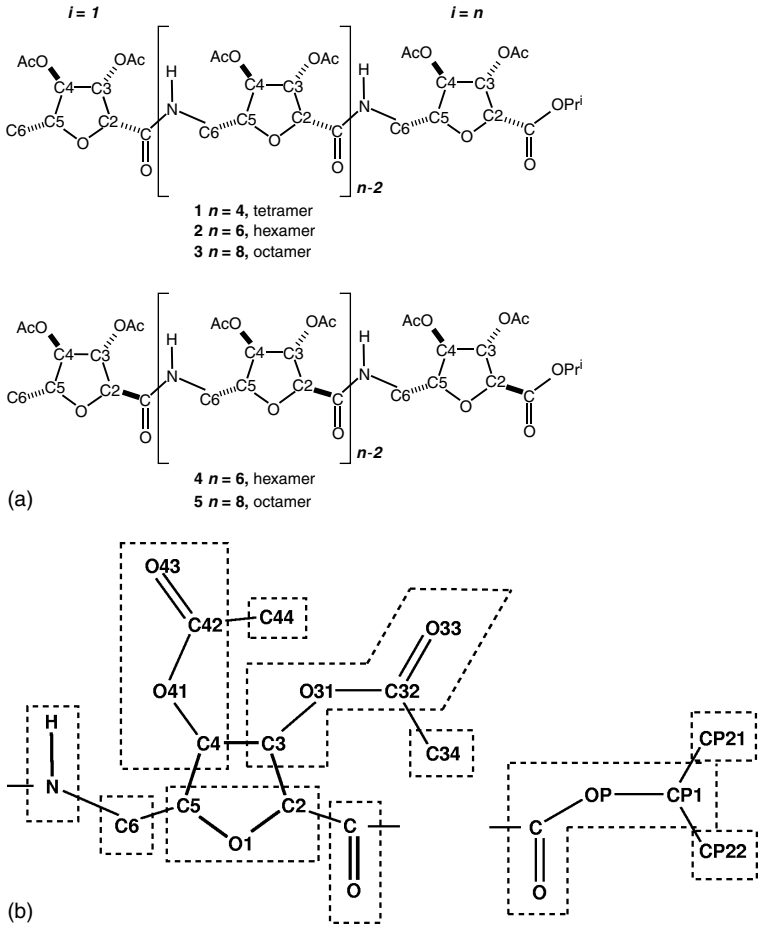

Figure 1 (a) Chemical formulae and atom names of the five carbopeptoids studied: *cis*-tetramer **1**, *cis*-hexamer **2**, *cis*-octamer **3**, *trans*-hexamer **4** and *trans*-octamer **5**. Residue numbers are indicated by *i*. (b) Atom naming convention for the force-field parameters. The dashed lines indicate the charge groups.

NH(i) – O(i – 3) hydrogen bonds has been reported for the octamer with ketal protecting groups [24]. The apparent dependence of structural motifs on chain length and stereochemistry makes these small systems an interesting target for MD simulations to study the microscopic mechanism of peptide folding.

Previously reported MD simulation results [25] showed encouraging agreement with experimental nuclear magnetic resonance (NMR) work and provided an interpretation of nuclear Overhauser effect (NOE) data in terms of conformational distributions. Here additional analysis is presented focusing on how different properties of flexible peptide chains are correlated to differences in their folding behaviour. The carbopeptoids considered are shown in Figure 1a and simulation details are summarized in Table 1. In addition to the six MD simulations discussed previously [25], three further MD simulations were performed and are presented here. The solvent effect on fold stability was investigated through an additional 100 ns simulation of the *cis*-hexamer **2** and 50 ns of the *cis*-octamer **3** with DMSO as solvent instead of chloroform. To assess the convergence over 100 ns of simulation for the fairly long *trans*-octamer **5**, another 100 ns run was carried out starting from a different initial structure with all torsional angles set to random values. Conformational analyses of the extended set of nine simulations at 298 K are presented. Solute entropies are calculated and related to the folding properties of this family of carbopeptoids. Finally an explanation is proposed for the differences in folding observed between the *trans*-linked hexamer **4** and octamer **5**.

METHODS

The molecular dynamics simulations were performed with the GROMOS96 package of programs using the 45A3 force field [26,27], augmented by ester group parameters originally derived for lipids [28]. The chloroform and DMSO models were taken from Tironi *et al.* [29] and Liu *et al.* [30], respectively. Aliphatic CH_n groups were generally treated as united atoms, except for chloroform. A summary of the force field parameters for the solutes is given in Table 2.

Two minor modifications were made to the peptide substitution pattern reported experimentally [20–24]. First, the *N*-terminal azide group was replaced by a methyl group for

Table 1 Simulated Systems. Each peptide is numbered as in Figure 1. The structures suggested by experiment are a right-handed helix (R) composed of repeating β-turns and a left-handed helix (L). Initial structures are either extended (ext.) or built from random torsional-angle values (random)

systems simulated									
peptide	1	1	2	2	3	3	4	5	5
THF linkage	*cis*	*cis*	*cis*	*cis*	*cis*	*cis*	*trans*	*trans*	*trans*
number of sugar rings	4	4	6	6	8	8	6	8	8
solvent	$CHCl_3$	DMSO	$CHCl_3$	DMSO	$CHCl_3$	DMSO	$CHCl_3$	$CHCl_3$	$CHCl_3$
type of helix derived from exp.	R	R	R	R	R	R	—	L	L
simulation parameters									
number of peptide atoms	74	74	110	110	146	146	110	146	146
number of solvent molecules	1567	1745	2579	2918	5479	6148	2807	6203	6206
number of atoms in the system	7909	7054	13 005	11 782	27 541	24 738	14 145	31 161	31 176
initial box volume [nm^3]	209	209	347	347	732	732	377	829	829
simulation length [ns]	100	100	100	100	100	50	100	100	100
initial structure	ext.	ext.	ext.	ext.	ext.	ext.	ext.	ext.	random

which force field parameters are readily available. Second, acetate groups were used consistently for all C3 and C4 substituents of all peptides (Figure 1a), whereas experimental work [20,21,23,24] considers various ketal protecting groups for different peptides. The type of protecting group, however, appears to have little influence on the folding properties of the peptide [24], and consistent substitution patterns simplify the comparison of peptides with different chain length and stereochemistry.

The initial structure for each peptide was chosen extended (all backbone torsional angles in *trans* configuration). For the *trans*-octamer **5** an additional simulation was performed starting from a structure with random torsional-angle values.

Periodic boundary conditions were applied with truncated octahedra as unit cells. Non-bonded interactions were truncated at $R = 1.4$ nm, they were updated every step for $R \leq 0.8$ nm and every five steps for $0.8 < R \leq 1.4$ nm. A reaction field correction was applied beyond the cutoff to account for the dielectric response of chloroform and DMSO (ε_{RF} of 5 and 30, respectively).

The peptide molecules were immersed in boxes of solvent large enough to avoid any interactions between mirror images of the solute. Solvent molecules in close contact with the solute were deleted if the minimum distance between non-hydrogen atoms fell below 0.23 nm. Table 1 details the resulting initial box volumes and numbers of solvent molecules for all simulations. The initial solute-solvent configurations were subjected to two subsequent energy minimization runs, each using the steepest descent algorithm with an energy convergence criterion of 0.1 kJ mol^{-1}. The first was run to relax the solvent and had a solute positional restraint imposed with a force constant of 250 kJ mol^{-1} nm^{-2}. The second was run without restraints to remove any residual strain.

The MD simulations were initialized with atomic velocities taken from a Maxwell-Boltzmann distribution at 298 K. Newton's equations of motion were integrated using the leapfrog algorithm with a time step of 2 fs, while keeping constrained the bond lengths and the geometry of the solvent molecules using the SHAKE algorithm [31] with a relative geometric tolerance of 10^{-4}. The weak coupling algorithm [32] was used with relaxation times of 0.1 ps and 0.5 ps, respectively, to maintain the reference temperature (298 K) and pressure (1 atm). A value of 4.574×10^{-4} [kJ mol^{-1} nm^{-3}]$^{-1}$ was used for the isothermal compressibility of the system.

Trajectory snapshots of the solute were taken every 0.01 ns and clustered into batches of similar configurations, using backbone atom-positional root-mean-square difference (RMSD) values as similarity criteria and the clustering algorithm described previously [33] with cutoffs of 0.13 nm (tetramer), 0.17 nm (hexamers) and 0.28 nm (octamers). All backbone atoms (N, C6, C5, O, C2, C) were used for the RMSD calculations except for C6, C5 and O of the first residue and C of the last. Cluster analyses were also performed on combined trajectories, i.e. on sets of solute configurations originating from two different MD simulations. This allows us to assess the overlap of conformational spaces visited in the various trajectories [34].

Model structures for the *cis*-linked carbopeptoids were generated through 1 ns MD simulations in which proton–proton distances were (instantaneously) restrained to distance bounds inferred from NOE NMR experiments, using a force constant of 1000 kJ mol^{-1} nm^{-2} (see reference [25] for the NOEs used). Clustering analysis was then performed to select single representative model structures from the 1 ns trajectories. All *cis*-linked model structures have right-handed helical character. The 16-membered left-handed helix for the *trans*-octamer **5** was modelled based on the qualitative information reported in NMR work [20], as no detailed NOE information was available. The resulting structure was refined by steepest descent energy minimization in chloroform to eliminate any residual strain.

Configurational entropies were calculated following the procedure of Schlitter [35] as implemented by Schäfer *et al.* [36,37]. A translational and rotational least-squares fit on all solute atom positions was carried out to eliminate

Table 2a Force-field Parameters used for the Tetrahydrofuran Carbopeptoids. The force field is based on the standard GROMOS functional form [26] and parameters are taken from refs. [26–28]. See Figure 1b for the atom naming convention. The GROMOS non-bonded interaction code defines the van der Waals parameters used

atom	description	GROMOS non-bonded interaction code	partial charge [e]
C	carbonyl C	11	0.38
O	carbonyl O	1	−0.38
N	amide N	5	−0.28
H	amide H	18	0.28
O1, O31, O41, OP	sugar or ester oxygen	3	−0.36
C2, C3, C4	sugar CH	12	0.16
C5, CP1	sugar CH, aliphatic carbon	12	0.20
C6	aliphatic CH2	13	0.00
C32, C42	carbonyl C (ester)	11	0.58
O33, O43	carbonyl O (ester)	1	−0.38
C34, C44, CP21, CP22	methyl (ester or aliphatic)	14	0.00

Table 2b

Bond [atom names]	b_0 [nm]	K_b [10^6 kJ mol^{-1} nm^{-4}]
N—H	0.100	18.7
C=O, C32=O33, C42=O43	0.123	16.6
N—C, O31—C32, O41—C42, OP—C	0.133	11.8
O1—C2, O1—C5	0.1435	6.10
N—C6, O31—C3, O41—C4, OP—CP1	0.147	8.71
C32—C34, C42—C44	0.148	7.64
C2—C3, C3—C4, C4—C5, C5—C6	0.152	5.43
C2—C, CP1—CP21, CP1—CP22	0.153	7.15
bond angle [atom names]	**θ_0 [degree]**	**K_θ [kJ mol^{-1}]**
C∗ —C∗—C∗ (furanose)	109.5	285
C∗ —C∗ —O∗ (furanose)	109.5	320
C2—O1—C5	109.5	380
C—C2—O1, C—C2—C3, C5—C6—N	109.5	520
OP—CP1—CP2∗, CP21—CP1—CP22	111.0	530
O31—C32—C34, O41—C42—C44, OP—C—C2	113.0	545
H—N—C6	115.0	460
N—C—C2	115.0	610
C3—O31—C32, C4—O41—C42, C—OP—CP1	117.0	635
C2—C—O	121.0	685
O31—C32—O33, O41—C42—O43, C—N—C6, OP—C—O	122.0	700
C—N—H	123.0	415
N—C—O	124.0	730
O33—C32—C34, O43—C42—C44, C2—C—O (C-terminus)	125.0	750

overall translation and rotation. All fits were performed with reference to the first configuration of the analysed trajectory. Side-chain, backbone and correlation contributions were calculated by applying the same procedure to subsets of solute atoms, including either all acetate side-chain atoms or all other atoms [37]. Complete trajectories were used for the entropy calculations, including the initial equilibration phase. Solute snapshots were taken at 5 ps intervals, and entropies were recalculated after every nanosecond.

Table 2c

improper (harmonic) dihedral angle [atom names]	ξ_0 [degree]	K_ξ [kJ mol^{-1} degree^{-2}]
N—C—C6—H	0.0	0.051
C—C2—N—O, C—C2—OP—O	0.0	0.051
C2—C3—C—O1 (*cis*), C2—C3—O1—C (*trans*)	35.26439	0.102
C5—C6—C4—O1	35.26439	0.102
C3—C4—O31—C2, C4—C3—O41—C5	35.26439	0.102
C32—C34—O31—O33, C42—C44—O41—O43	0.0	0.051
CP1—OP—CP22—CP21	35.26439	0.102

proper (trigonometric) torsional angle [atom names]	cos(δ)	m	K_ϕ [kJ mol^{-1}]
C3—C2—C—N, C3—C2—C—OP	+1	6	1.0
C2—C—N—C6	−1	2	33.5
C2—C—OP—CP1	−1	2	16.7
C—N—C6—C5	−1	6	1.0
N—C6—C5—C4	+1	3	5.86
N—C6—C5—O1	+1	2	0.418
O1—C2—C3—C4, C6—C5—C4—C3	+1	3	5.86
O1—C2—C3—C4, O1—C5—C4—C3	+1	2	0.418
C—C2—C3—O31, C6—C5—C4—O41	+1	2	0.418
O1—C2—C3—O31, O1—C5—C4—O41	+1	2	2.09
C2—C3—C4—C5	+1	3	5.86
C2—C3—C4—O41	+1	2	0.418
O31—C3—C4—C5	+1	2	0.418
O31—C3—C4—O41	+1	2	2.09
C5—O1—C2—C3, C2—O1—C5—C4	+1	3	3.77
C2—C3—O31—C32, C5—C4—O41—C42	+1	3	3.77
C3—O31—C32—C34, C4—O41—C42—C44	−1	2	16.7
C—OP—CP1—CP21	+1	3	3.77

RESULTS

Secondary Structures

Figure 2 shows as a time series the atom-positional RMSD of trajectory structures from the model structure of the *cis*-hexamer **2** in DMSO (upper panel), the *cis*-octamer **3** in DMSO (middle panel) and the *trans*-octamer **5** in chloroform which has been simulated starting from a structure with random torsional-angle values (lower panel). These plots complement the analysis for the other six simulations presented in reference [25].

During 100 ns of simulation in DMSO, the *cis*-hexamer **2** samples configurations characteristic of its stable fold in chloroform just two times, after 4 and 70 ns (Figure 2, upper panel). The strong hydrogen-bonding character of DMSO disfavours the intramolecular hydrogen bonding typical of the solution structure in chloroform (Table 3 and reference [25]). In DMSO the *cis*-linked furanose amino acid oligomers do not lose completely the typical intramolecular backbone NH(i) – O($i-2$) hydrogen bond pattern observed in chloroform [25], but the dominant helical fold consisting of a sequence of linked β-turns becomes less stable. This was also suggested experimentally based on an analysis of temperature coefficients and chemical shifts [22]. Similar conclusions apply to the solvent-dependent folding behaviour of the *cis*-tetramer **1** [25]. In the case of the *cis*-octamer **3**, neither the simulation with chloroform [25] nor that with DMSO (Figure 2, middle panel) as solvent fully approaches the model structure derived from experimental information. The analysis of hydrogen bonds, however, still confirms that the more polar solvent DMSO disfavours the formation of the typical hydrogen-bond pattern found for *cis*-linked carbopeptoids (see Table 3).

Finally, the two trajectories for the *trans*-linked octamer **5**, starting from either an extended [25] or a randomized conformation (Figure 2, lower panel) show only one single folding event each, casting some doubt on the vague experimental information available to build the reference model structure.

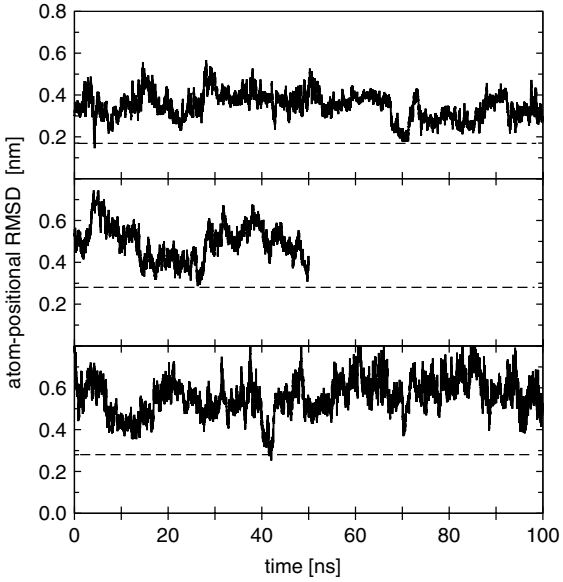

Figure 2 Backbone atom-positional root-mean-square deviation (RMSD) of trajectory structures from right- handed helical model structures as function of time: *cis*-hexamer **2** in DMSO (upper panel) and *cis*-octamer **3** in DMSO (middle panel). RMSD from left-handed helical model structure for the *trans*-octamer **5** in chloroform starting from a configuration with random torsional angles (lower panel). The dashed lines indicate the backbone RMSD similarity criterion chosen to define the folded structures. The construction of the model structures is described in the Methods section. All structures were superimposed using the backbone atoms (N, C6, C5, O, C2, C), excluding C6, C5 and O of the first residue and C of the last one.

Clustering trajectory structures into batches of highly similar configurations (see Methods) generally shows the dominant configurations sampled in an MD simulation. For the *cis*-hexamer in DMSO we find 7% of all trajectory structures in the most populated cluster, and a total of 29 significantly (>1% each) populated clusters, summing to 73% of all configurations. For the same molecule in chloroform we found previously [25] 11% of all structures in the first cluster and only 20 significantly populated clusters summing to 76% of the ensemble. In agreement with the RMSD and hydrogen-bond analysis presented above, these numbers demonstrate a larger flexibility of the peptide in the more polar solvent DMSO. The values found for the *trans*-hexamer **4** in chloroform (5%, 30, 62%) indicate that the change to a 2,5-*trans* stereochemistry on the THF ring has a similar effect on folding as the change in solvent polarity.

The *trans*-octamer **5** shows considerable conformational variability in chloroform as solvent (Figure 2 and reference [25]). This poses the question of the degree of sampling reached within 100 ns for this fairly large carbopeptoid whose length equals that of 16 α-amino acids. Comparison of the two 100 ns trajectories starting from different initial configurations may help to answer this question [38]. The first simulation shows larger conformational diversity: 27 vs 15 clusters are populated significantly, totalling 82% vs 87% of the entire ensemble. This may be explained by the extended starting structure from which the trajectory must pass through different high free energy conformers to reach its more compact equilibrium ensemble of structures. The intramolecular hydrogen bonding differs between the two simulations, although differences may be not too significant in view of the small percentages (Table 3). Yet, the variety of structures sampled in the two simulations of the *trans*-octamer **5** indicates that the 16-membered left-handed helix NH(i) – O($i-3$) suggested experimentally [24] is not the only possible stable configuration. Other hydrogen bonds NH(i) – O($i+k$) do occur in the two simulations of the *trans*-octamer **5** which are characteristic for hydrogen-bonded rings of size 20 ($k=2$), 22 ($k=-4$), 26 ($k=3$), 28 ($k=-5$), 32 ($k=4$), 34 ($k=-6$) and 38 ($k=5$).

In view of the differences observed between the two trajectories, some caution must be exercised when calculating thermodynamic properties from statistical averages. Hence for the calculation of entropies, both trajectories are considered and the results obtained are compared (see below).

Characterization of Conformational Spaces

Figure 3 shows the backbone torsional-angle distributions of the dihedral angles ω (C2—C—N—C6), ϕ (C—N—C6—C5) and ψ (N—C6—C5—O) for *cis*-linked carbopeptoids **1**, **2** and **3** (from left to right) from MD simulations with chloroform as solvent. No large variations emerge between different residues in the same molecule and between corresponding residues in different molecules. Only some bias was noted in the ϕ distribution towards gauche (+) conformations for *N*-terminal residues.

Figure 4 compares pairs of simulation trajectories for the *cis*-linked carbopeptoids. The conformational spaces accessible to the segments constituted by residues 1–4 and 3–6 of the *cis*-hexamer **2** and *cis*-octamer **3** show sizeable overlap with the conformational space explored by the *cis*-tetramer **1** (Figure 4a, b and d, e). None of the first ten most populated clusters from the combined trajectories is populated exclusively by one of the two peptide chains. The dominant secondary structure motifs displayed by the *cis*-tetramer **1** in chloroform are also preferred by the four-residue segments of the longer carbopeptoids, the only exception being the *C*-terminal segment of the octamer **3** (residues 5–8, Figure 4h). A common feature of many of the clusters appearing in Figure 4a, b, d, e is that their members share at least one of the two NH(i) – O($i-2$) hydrogen bonds characteristic of the stable fold of the *cis*-tetramer **1**. These results agree with the experimental observation of a characteristic periodic architecture of repeating β-turns secondary structures for

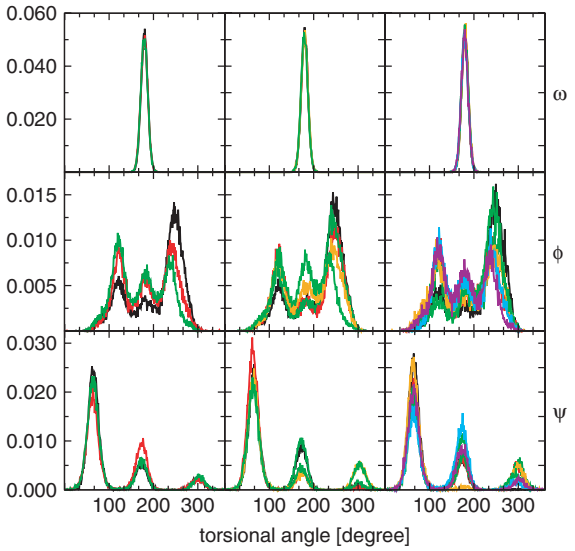

Figure 3 Torsional-angle distributions: ω (C2—C—N—C6), ϕ (C—N—C6—C5) and ψ (N—C6—C5—O) in residues 2 (black), 3 (red), 4 (green), 5 (orange), 6 (light green), 7 (cyan) and 8 (violet) of *cis*-tetramer **1** (left panels), *cis*-hexamer **2** (middle panels) and *cis*-octamer **3** (right panels) in chloroform.

all *cis*-linked carbopeptoids **1**, **2** and **3** [20,21]. Less conformational overlap is observed between the larger six-residue segments of the *cis*-octamer **3** and the peptide chain of the *cis*-hexamer **2** (Figure 4c, f and i). The most populated cluster is again shared by all pairs of trajectories, but the overlap is much smaller for the less populated clusters.

Finally the cluster analysis for the combined trajectories of the *cis*-tetramer **1** in chloroform and DMSO shows considerable overlap (Figure 4g). Structures from both trajectories contribute to each of the 20 most populated clusters. Apparently, changes in solvent polarity affect structural details such as certain hydrogen-bond frequencies (see above), but they do not have a marked impact on the overall folding behaviour.

Analogous analysis for the *trans*-linked carbopeptoids is presented in Figure 5. The hexamer **4** and octamer **5** show rather similar torsional-angle distributions (Figure 5, left and middle panels), but slightly more variation between residues is discernible for the longer carbopeptoid. In contrast to their *cis*-linked counterparts, peptides **4** and **5** contain several residues with ϕ angle distributions biased towards gauche (–) conformations. Clustering analysis shows little conformational overlap between *trans*-hexamer **4** and any

Table 3 Intramolecular Backbone NH—CO Hydrogen-bond Occurrences from 100 ns Simulations. The hydrogen bonds are specified in the second columns with residue numbers given in parentheses. Percentage values in parentheses refer to a 50 ns simulation. Hydrogen bonds are defined to have a maximum hydrogen-acceptor distance of 0.25 nm and a minimum donor-hydrogen-acceptor angle of 135°

molecule	hydrogen bond	occurrence (%)		molecule	hydrogen bond	occurrence (%)	
		solvent				initial structure	
		chloroform	DMSO			extended	random
cis-tetramer 1	NH(3)—O(1)	11.1	1.0	*trans*-octamer 5	NH(2)—O(4)	1.1	0.8
	NH(4)—O(2)	6.0	3.3		NH(2)—O(5)	0.7	0.1
cis-hexamer 2	NH(3)—O(1)	11.7	0.1		NH(2)—O(6)	1.1	1.1
	NH(4)—O(2)	14.0	1.0		NH(2)—O(7)	7.8	0.8
	NH(5)—O(3)	7.6	1.3		NH(3)—O(5)	0.6	1.2
	NH(6)—O(4)	9.7	3.2		NH(3)—O(6)	8.3	0.6
cis-octamer 3	NH(3)—O(1)	12.0	(0.7)		NH(3)—O(7)	2.9	0.3
	NH(4)—O(2)	7.6	(0.6)		NH(4)—O(6)	3.9	0.0
	NH(5)—O(3)	13.8	(2.3)		NH(4)—O(7)	0.3	4.1
	NH(6)—O(4)	14.9	(1.8)		NH(4)—O(8)	5.7	0.3
	NH(7)—O(5)	9.2	(0.6)		NH(5)—O(7)	1.2	1.2
	NH(8)—O(6)	6.2	(1.5)		NH(5)—O(8)	0.8	0.8
trans-hexamer 4	NH(2)—O(4)	1.6	—		NH(6)—O(8)	0.1	2.7
	NH(2)—O(5)	1.8	—		NH(7)—O(1)	0.5	1.0
	NH(2)—O(6)	0.5	—		NH(7)—O(2)	2.9	2.5
	NH(3)—O(5)	0.5	—		NH(7)—O(3)	3.6	0.0
	NH(3)—O(6)	2.4	—		NH(7)—O(4)	0.3	0.1
	NH(4)—O(6)	1.1	—		NH(8)—O(1)	0.5	0.5
	NH(6)—O(1)	0.4	—		NH(8)—O(2)	1.2	0.3
	NH(6)—O(2)	1.4	—		NH(8)—O(3)	3.7	2.7

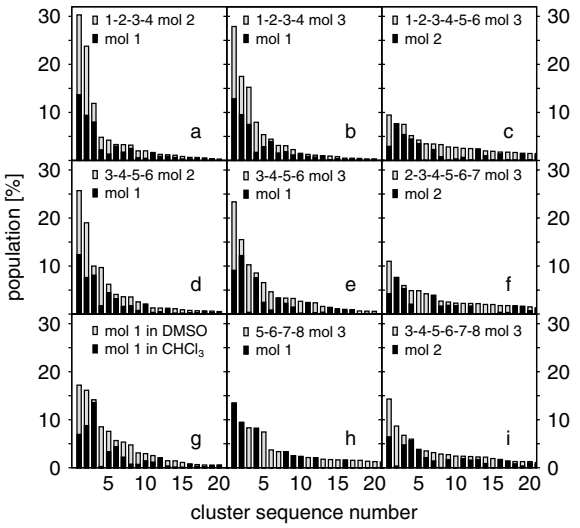

Figure 4 Populations of the first 20 clusters obtained from cluster analysis for combined trajectories from two simulations. (a) and (d) show the results for a combination of the trajectory of peptide **1** (black fill) with trajectories for residues 1–4 and 3–6 of peptide **2** (grey fill). Analogously, peptide **1** (black fill) is compared with residues 1–4, 3–6 and 5–8 of peptide **3** (grey fill) in (b), (e) and (h). The comparisons between peptide **2** (black fill) and residues 1–6, 2–7 and 3–8 of peptide **3** (grey fill) are given in (c), (f) and (i). (g) compares trajectories for peptide **1** in chloroform (black fill) and DMSO (grey fill).

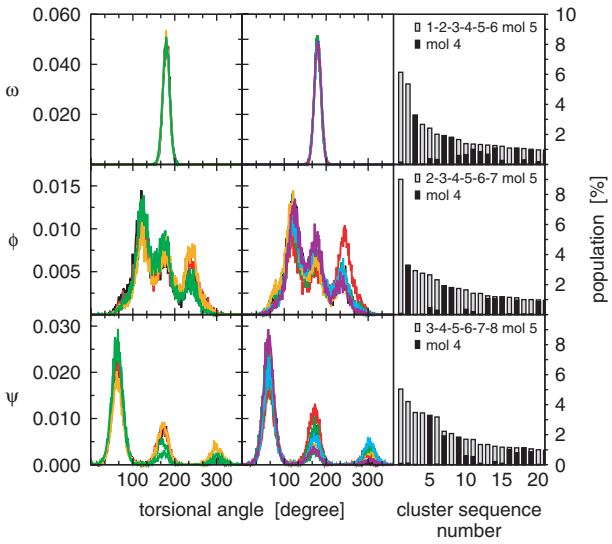

Figure 5 Torsional-angle distributions: ω, ϕ and ψ of *trans*-hexamer **4** (left panels) and *trans*-octamer **5** (middle panels). For colour coding of the residues see caption of Figure 3. The right-hand panels show populations of the first 20 clusters of combined trajectories from the simulations of hexamer **4** (black fill) and octamer **5** (grey fill). Residues 1–6 (upper panel), 2–7 (middle panel) and 3–8 (lower panel) of the octamer **5** are considered for this analysis.

six-residue segment of the *trans*-octamer **5** (Figure 5, right panels). Almost all of the most populated clusters refer to conformations only adopted by the octamer. This result contrasts the observations made for the set of *cis*-isomers and shows that the *trans*-octamer has a stronger tendency to adopt particular conformations than the *trans*-hexamer. Carbopeptoids with a *trans*-linkage apparently require longer chains to generate stable folds than do the *cis*-linked isomers. This conclusion is in line with experimental observations reported previously [20,21]. Finally, it is noted that the conformational overlap is smallest for the *C*-terminal segment 3–8, the part of the chain mainly involved in the formation of left-handed hydrogen bonded rings.

Configurational Entropies

An alternative way to characterize the extent of configurational space accessible to a molecule is to calculate its configurational entropy. The approximation proposed by Schlitter [35] was used to this end, and implemented as described earlier in an application to a β-peptide [37] and to a protein [39]. Estimates of the solute configurational entropies are reported in Figure 6 and Table 4. The calculated entropy converges within 50 ns for the *cis*-tetramer **1** and within 100 ns for the larger molecules. Starting from rather different initial configurations of the longest molecule, the *trans*-octamer **5**, entropies were calculated that differed by only 0.2%, and entropies per residue [37] that differed by less than 3% (not shown). The picture emerging from the results shown in Figure 6 and Table 4 is consistent with the trends observed before in the hydrogen-bond and conformational clustering analyses: the solute entropy of a given peptide is smaller in chloroform than in DMSO. It is also smaller for a *cis*-linked carbopeptoid than for the corresponding *trans*-linked molecule. This confirms the reduced tendency of the *trans*-linked carbopeptoids to form stable folds (see above). The hexamer **4** lacks any stable secondary structure and the octamer **5** prefers folds containing fairly large hydrogen-bonded rings which allow for a much higher backbone flexibility than the shorter turns of the *cis*-linked isomers. Splitting the total solute entropy S_{all} into three terms, the backbone entropy S_{bb}, the side-chain entropy S_{sc}, and the contribution from correlation between backbone and side-chain motion $\Delta S^{corr}_{sc/bb}$, it is seen that the latter amounts to about 15% and that backbone and side-chain contributions are about equal. The side-chains are still more flexible than the backbone, and this is reflected in their higher entropy value per atom.

DISCUSSION AND CONCLUSIONS

MD simulations have been used to characterize in atomic detail the folding behaviour of several carbopeptoids. Previously [25] good agreement was found between simulation and experimental (NMR)

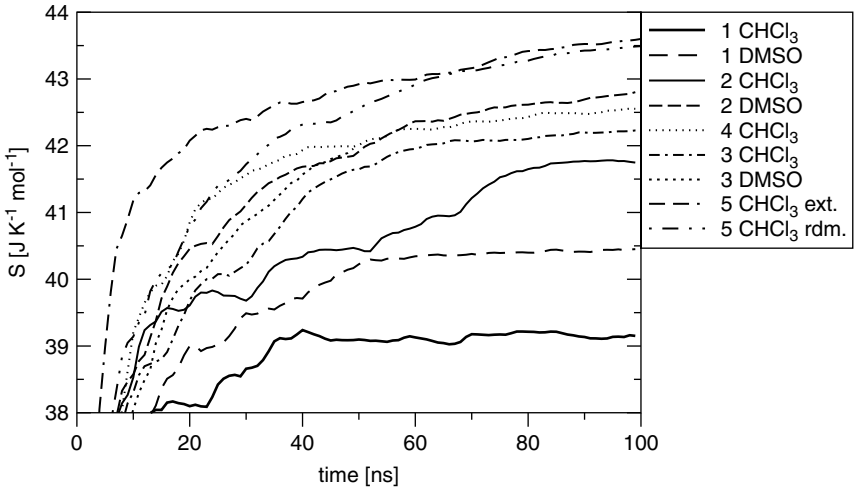

Figure 6 Solute configurational entropies S_{all} per atom, excluding overall translation and rotation, for the *cis*-tetramer **1**, *cis*- and *trans*-hexamers **2** and **4**, *cis*-and *trans*-octamers **3** and **5**, in $CHCl_3$ and DMSO (see legend box and Table 1).

Table 4 Solute Configurational Entropies Calculated for all Five Peptides. Rotational and translational degrees of freedom have been projected out by superposition of trajectory structures. For each system, we report the entropies of the entire peptide (S_{all}), its backbone (S_{bb}), its side chains (S_{sc}) and contributions due to the correlation between backbone and side chains ($\Delta S^{corr}_{sc/bb}$) as defined in reference [37]. Entropies per atom are given in parentheses. The two results for peptide **5** in $CHCl_3$ refer to simulations starting from either an extended configuration or a configuration with random torsional-angle values (last row). All entropies are given in [J K^{-1} mol^{-1}]

molecule	solvent	S_{all}	S_{bb}	S_{sc}	$\Delta S^{corr}_{sc/bb}$	$\Delta S^{corr}_{sc/bb}/S_{all}$ (%)	S_{bb} (%)	S_{sc} (%)
cis-tetramer 1	$CHCl_3$	2975 (40)	1637 (39)	1761 (55)	423 (5)	14 (12)	41	44
	DMSO	3074 (41)	1716 (41)	1805 (56)	447 (6)	14 (15)	42	44
cis-hexamer 2	$CHCl_3$	4592 (42)	2543 (41)	2762 (57)	713 (6)	15 (14)	40	44
	DMSO	4711 (43)	2634 (42)	2807 (58)	730 (7)	15 (16)	41	43
cis-octamer 3	$CHCl_3$	6166 (42)	3444 (42)	3750 (59)	1028 (7)	17 (17)	40	43
trans-hexamer 4	$CHCl_3$	4681 (42)	2694 (43)	2753 (57)	766 (7)	16 (17)	41	42
trans-octamer 5	$CHCl_3$	6365 (44)	3681 (45)	3780 (59)	1096 (7)	17 (16)	41	42
	$CHCl_3$	6350 (43)	3681 (45)	3774 (59)	1105 (7)	17 (16)	41	42

data for this class of molecules, showing that the biomolecular force field used was accurate enough to describe their conformational distributions reliably.

In the present study, the conformational distributions were analysed in terms of structural properties (hydrogen bonds, backbone torsional angles), conformational similarity (cluster analysis) and conformational flexibility (configurational entropy). Encouragingly, the three types of analyses provide a very consistent picture of the folding characteristics of this class of carbopeptoids. Molecules with 2,5-*cis* stereochemistry tend to form fairly rigid 16 membered hydrogen-bonded rings connecting residues (i) and ($i - 2$). This structural motif is repeated throughout the peptide chain, forming right-handed helices composed of successive β-turns. As seen in the cluster analyses, the tetramer, the hexamer and the octamer all show the same repeating hydrogen-bonding pattern, they all prefer the same helical motif and only differ in length.

Carbopeptoids with 2,5-*trans* stereochemistry appear to be unable to form similarly small hydrogen-bonded rings. The formation of stable secondary structures thus requires longer chains, and in turn involves larger and thus less rigid and stable hydrogen-bonded rings. The hexamer is rather structureless, and even the octamer shows only weak conformational preferences, forming various hydrogen-bonded rings of size 20 to 38. Consequently, the *trans* isomers are characterized by higher solute configurational entropies than their *cis*-counterparts. The lack of a repeating structural motif for the *trans*-isomers is reflected both in the hydrogen-bonding and in the cluster analyses.

Finally, all three types of analyses show that the conformational variability increases and the tendency to form intra-solute hydrogen bonds decreases (a) upon

increasing the chain length, (b) upon inverting the stereochemistry from 2,5-*cis* to 2,5-*trans*, and (c) upon increasing solvent polarity.

It is noted that most of these conclusions are perfectly in line with conclusions drawn from experimental work. In contrast to experimental (NMR, NOE) data, however, MD simulations provide not only average structural information, but also distributions of conformations. Experimental data are often interpreted in terms of single structures, while MD simulations typically show several conformations to be relevant. Often — but not always — there is one particularly dominant conformation which can then be associated with the structure derived from experimental work. If no such dominant conformation is found, the MD simulations may still be in good agreement with the raw experimental data, but the structural analysis may be at variance with a single structure derived from experimental work. In particular, the work suggests that the octamer **5** with 2,5-*trans* stereochemistry is structurally much less well-defined than suggested experimentally.

Solute configurational entropies are an excellent and direct measure of solute flexibility. While entropies in general used to be notoriously difficult to obtain from MD simulation, the current work shows again that Schlitter's formula provides a fairly easy and robust access. The excellent agreement between solute entropies calculated for two quite different trajectories of the octamer **5** may be somewhat fortuitous. However, very good convergence of entropies was observed for all trajectories analysed in this work. This promising result shows that 100 ns long MD simulations provide enough sampling for reliable entropy estimates of small to medium-sized peptides.

Acknowledgement

Financial support was obtained from the Schweizer Nationalfonds, project number 2000-063590.00, and through the National Center of Competence in Research (NCCR) Structural Biology of the Swiss National Science Foundation, which is gratefully acknowledged.

REFERENCES

1. 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.
2. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed.* 1999; **38**: 236–240.
3. Takano M, Yamato T, Higo J, Suyama A, Nagayama K. Molecular dynamics of a 15-residue poly(L-alanine) in water: helix formation and energetics. *J. Am. Chem. Soc.* 1999; **121**: 605–612.
4. Roccatano D, Amadei A, Di Nola A, Berendsen HJC. A molecular dynamics study of the 41–56 β-hairpin from B1 domain of protein G. *Protein Sci.* 1999; **8**: 2130–2143.
5. Bonvin AMJJ, van Gunsteren WF. β-Hairpin stability and folding: molecular dynamics studies of the first β-hairpin of tendamistat. *J. Mol. Biol.* 2000; **296**: 255–268.
6. van Gunsteren WF, Bürgi R, Peter C, Daura X. The key to solving the protein-folding problem lies in an accurate description of the denatured state. *Angew. Chem. Int. Ed.* 2001; **40**: 351–355.
7. Glättli A, Daura X, Seebach D, van Gunsteren WF. Can one derive the conformational preference of a β-peptide from its CD spectrum? *J. Am. Chem. Soc.* 2002; **124**: 12 972–12 978.
8. Colombo G, Roccatano D, Mark AE. Folding and stability of the three-stranded β-sheet peptide betanova: insights from molecular dynamics simulations. *Proteins Struct. Funct. Genet.* 2002; **46**: 380–392.
9. Wu HW, Wang SM, Brooks BR. Direct observation of the folding and unfolding of a β-hairpin in explicit water through computer simulation. *J. Am. Chem. Soc.* 2002; **124**: 5282–5283.
10. Schäfer M, Bartels C, Karplus M. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* 1998; **284**: 835–848.
11. Ferrara P, Caflisch A. Folding simulations of a three-stranded antiparallel β-sheet peptide. *Proc. Natl. Acad. Sci. USA* 2000; **97**: 10 780–10 785.
12. Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, Garcia AE. Peptide folding simulations. *Curr. Opin. Struct. Biol.* 2003; **13**: 168–174.
13. Hansson T, Oostenbrink C, van Gunsteren WF. Molecular dynamics simulations. *Curr. Opin Struct. Biol.* 2002; **12**: 190–196.
14. Pande VS, Rokhsar DS. Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G. *Proc. Natl. Acad. Sci. USA* 1999; **96**: 9062–9067.
15. Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 2002; **124**: 11 258–11 259.
16. Snow CD, Nguyen N, Pande VS, Gruebele M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* 2002; **420**: 102–106.
17. Stigers KD, Soth MJ, Nowick JS. Designed molecules that fold to mimic protein secondary structures. *Curr. Opin. Chem. Biol.* 1999; **3**: 714–723.
18. Cubberley MS, Iverson BL. Models of higher-order structure: foldamers and beyond. *Curr. Opin. Chem. Biol.* 2001; **5**: 650–653.
19. Hill DJ, Mio MJ, Prince RB, Hughes TS, Moore JS. A field guide to foldamers. *Chem. Rev.* 2001; **101**: 3893–4011.
20. Smith MD, Fleet GWJ. Designing secondary structures: 5-azidomethyl tetrahydrofuran-2-carboxylates as carbohydrate-derived dipeptide isosteres. *J. Peptide Sci.* 1999; **5**: 425–441.
21. Smith MD, Claridge TDW, Sansom MSP, Fleet GWJ. Bend ribbon-forming tetrahydrofuran amino acids. *Org. Biomol. Chem.* 2003; **1**: 3647–3655.
22. Smith MD, Claridge TDW, Tranter GE, Sansom MSP, Fleet GWJ. Secondary structure in oligomers of carbohydrate amino acids. *Chem. Commun.* 1998; 2041–2042.
23. Smith MD, Long DD, Martin A, Marquess DG, Claridge TDW, Fleet GWJ. Absence of secondary structure in a carbopeptoid tetramer of a trans-5-aminomethyl-tetrahydrofuran-2-carboxylate. *Tetrahedron Lett.* 1999; **40**: 2191–2194.
24. Claridge TDW, Long DD, Hungerford NL, Aplin RT, Smith MD, Marquess DG, Fleet GWJ. An octameric carbopeptoid; Secondary structure in octameric and tetrameric 5-aminomethyl-tetrahydrofuran-2-carboxylates. *Tetrahedron Lett.* 1999; **40**: 2199–2202.
25. Baron R, Bakowies D, van Gunsteren WF. Carbopeptoid folding: effects of stereochemistry, chain length and solvent. *Angew. Chem. Int. Ed.* 2004; **43**: 4055–4059.
26. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. *The GROMOS96 Manual and User Guide*. Vdf Hochschulverlag AG: Zürich, 1996.

27. Schuler LD, Daura X, van Gunsteren WF. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* 2001; **22**: 1205–1218.
28. Chandrasekhar I, Kastenholz M, Lins RD, Oostenbrink C, Schuler LD, Tieleman DP, van Gunsteren WF. A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field. *Eur. Biophys. J.* 2003; **32**: 67–77.
29. Tironi IG, van Gunsteren WF. A molecular dynamics simulation study of chloroform. *Mol. Phys.* 1994; **83**: 381–403.
30. Liu H, Müller-Plathe F, van Gunsteren WF. A force field for liquid dimethyl sulfoxide and physical properties of liquid dimethyl sulfoxide calculated using molecular dynamics simulations. *J. Am. Chem. Soc.* 1995; **117**: 4363–4366.
31. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Chem.* 1977; **81**: 3684–3698.
32. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 1984; **81**: 3684–3698.
33. Daura X, van Gunsteren WF, Mark AE. Folding-unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. *Proteins Struct. Funct. Genet.* 1999; **34**: 269–280.
34. Baron R, Bakowies D, van Gunsteren WF, Daura X. β-peptides with different secondary-structure preferences: how different are their conformational spaces? *Helv. Chim. Acta* 2002; **85**: 3872–3882.
35. Schlitter J. Estimation of absolute and relative entropies of macromolecules using the covariance-matrix. *Chem. Phys. Lett.* 1993; **215**: 617–621.
36. Schäfer H, Mark AE, van Gunsteren WF. Absolute entropies from molecular dynamics simulation trajectories. *J. Chem. Phys.* 2000; **113**: 7809–7817.
37. Schäfer H, Daura X, Mark AE, van Gunsteren WF. Entropy calculations on a reversibly folding peptide: changes in solute free energy cannot explain folding behavior. *Proteins Struct. Funct. Genet.* 2001; **43**: 45–56.
38. Smith LJ, Daura X, van Gunsteren WF. Assessing equilibration and convergence in biomolecular simulations. *Proteins Struct. Funct. Genet.* 2002; **48**: 487–496.
39. Schäfer H, Smith LJ, Mark AE, van Gunsteren WF. Entropy calculations on the molten globule state of a protein: side-chain entropies of α-lactalbumin. *Prot. Struct. Funct. Genet.* 2002; **46**: 215–224.

