Eur Biophys J (2010) 39:1421–1432
DOI 10.1007/s00249-010-0596-9

ORIGINAL PAPER

Residual structure in a peptide fragment of the outer membrane protein X under denaturing conditions: a molecular dynamics study

Vincent Kräutler · Sebastian Hiller · Philippe H. Hünenberger

Received: 20 August 2009 / Revised: 16 February 2010 / Accepted: 1 March 2010 / Published online: 21 March 2010

Abstract The *Escherichia coli* outer membrane protein X (OmpX) contains two polypeptide segments that present nonrandom residual structure in 8 M aqueous urea, whereas the remainder of the protein is in a flexibly disordered conformation (Tafer et al. in Biochemistry 43:860–869, 2004). In the present study, the results of two long-timescale (0.4 μs) unrestrained explicit-solvent molecular dynamics (MD) simulations of a tetradecapeptide representative of one of these two segments in 8 M aqueous urea are reported and analyzed. The two simulations were initiated either from the conformation of the corresponding segment in an NMR model structure of the unfolded protein or from an entirely extended configuration. The sampled conformational ensembles agree qualitatively with the experimentally observed NOEs, but not quantitatively, suggesting that a number of relevant configurations were not visited on the 2×0.4 μs timescale. Major conformational transitions occur on the 0.1 μs timescale, and the ensembles corresponding to the two independent simulations overlap only to a limited extent. However, both simulations show in multiple events the reversible formation and disruption of α-helical secondary structure (characteristic of the urea-denatured state) and β-turn secondary structure (characteristic of the native state). Events of helix formation are correlated with the appearance of hydrogen bonds between two side chains (Asp75–Ser78) and of a persistent hydrophobic contact (Trp76–Tyr80). They also evidence a peculiar helix stabilization and N-terminal capping role for a negatively charged residue (Asp75). These features are in good qualitative agreement with the NMR model for the structured state of the corresponding segment in the urea-denatured protein. The analysis of the simulations provides a detailed picture of the structural and dynamic features of the considered peptide at atomic resolution that is of high relevance in the understanding of the OmpX folding process.

Keywords Computer simulation · Molecular dynamics · OmpX fragment · Denatured state · Residual structure · Hydrophobic clustering

V. Kräutler · S. Hiller · P. H. Hünenberger (✉)
Laboratory of Physical Chemistry,
ETH Zürich, 8093 Zürich, Switzerland
e-mail: phil@igc.phys.chem.ethz.ch

S. Hiller
Institute of Molecular Biology and Biophysics,
ETH Zürich, 8093 Zürich, Switzerland

Introduction

The determination of the structure of globular proteins by nuclear magnetic resonance (NMR) spectroscopy and crystallography (X-ray) has in many cases become a routine operation (Wüthrich 1995, 2003; Wider and Wüthrich 1999; Montelione et al. 2000). Still, reaching a detailed understanding of the folding process remains one of the major challenges of structural biology. There is significant evidence suggesting that, even under denaturing conditions, many proteins adopt conformations that are not satisfactorily described by a pure random-coil model, but exhibit residual local and nonlocal structure (Radford et al. 1992; Shortle 1993, 1996; Shortle and Ackerman 2001; McCarney et al. 2005). Residual structure elements in the denatured state of proteins (which need not necessarily be native-like)

are believed to play an important role in both the thermodynamics and the kinetics of protein folding. The characterization of these elements is therefore expected to provide valuable insight into the folding process. For example, in the soluble protein lysozyme, residual structure induced by hydrophobic clustering in the denatured state has been shown to provide the seeding for the initial hydrophobic collapse of the protein (Klein-Seetharaman et al. 2002).

The *Escherichia coli* outer membrane protein X (OmpX; Fig. 1a) (Tamm et al. 2001; Wimley 2003; Tafer et al. 2004) is a 148-residue integral β-barrel transmembrane protein involved in cell adhesion and entry into the host cell (Vogt and Schulz 1999). In vivo, the protein is synthesized along with a signal sequence and translocated to the cell membrane in an unfolded form (Mori and Ito 2001; Wimley 2003). There, insertion into the membrane and folding result from a chaperone-facilitated process (Kleinschmidt and Tamm 2002). In vitro, the folded protein can also be studied within various types of detergent micelles (Pautsch et al. 1999; Fernández et al. 2001).

OmpX is the first membrane-bound protein for which the denatured state has been characterized by NMR experiments (Tafer et al. 2004). Under denaturing conditions in 8 M aqueous urea solution, two regions of the polypeptide chain adopt nonrandom conformations. Region I, corresponding to residues 73–82, partially adopts a helical conformation, whereas region II, corresponding to residues 137–145, partially forms a structured hydrophobic cluster with Trp140 at its core (Fig. 1). The relative populations of these structured states in the denatured protein are about 25% (Tafer et al. 2004). These regions are particularly interesting, because (1) the observed conformations differ from the corresponding conformations in the folded membrane-bound protein: region I forms a β-turn between strands β_4 and β_5 and region II a β-strand in the native structure (Fernández et al. 2004) (Fig. 1a); (2) these two hydrophobic regions are known to be specifically involved in the binding of urea-denatured OmpX to micelles and thus likely play an important role in the process of insertion and folding (Hiller et al. 2008).

Molecular dynamics (MD) simulations provide a powerful tool complementary to experiment for investigating the structure and dynamics of peptides and proteins in solution at atomic resolution (Daura et al. 1999). When experimental structural data from NMR spectroscopy or X-ray crystallography are available, this information can be used in three different ways in MD simulations:

1. It can be used to construct a three-dimensional structural model at atomic resolution, which can then serve as an *initial configuration* for the simulation.
2. It can be included as a *boundary condition* during the simulation, by applying [possibly time-averaged (Nanzer et al. 1995) or ensemble-averaged (Fennen et al. 1995)] restraints to enforce agreement between (averaged) microscopic observables and the corresponding experimentally determined values.
3. It can be used for *validation*, i.e., to assess the agreement between the results of an unrestrained simulation (driven by the atomic force-field only) and experimental observations.

When the experimental data are obtained from NMR spectroscopic measurements, interproton distances derived from Nuclear Overhauser Enhancements (NOE) and *J*-coupling constants represent the most common sources of observables for the three approaches.

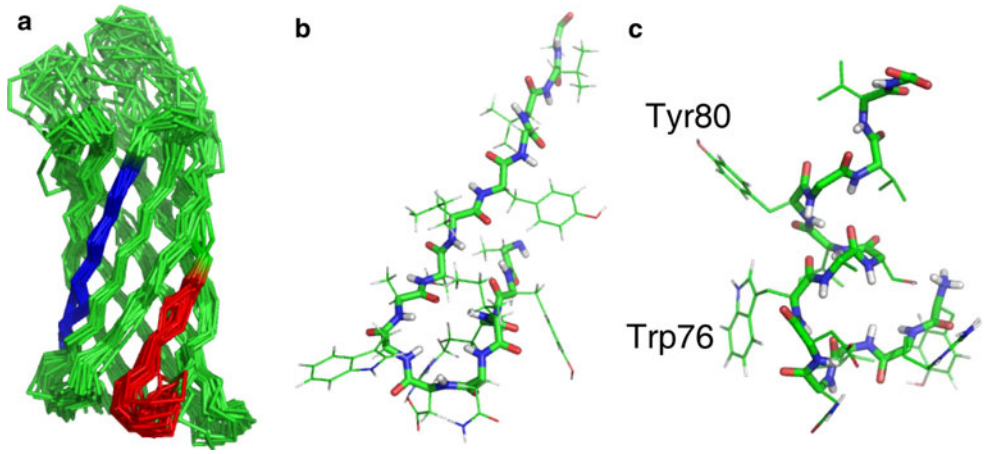

Fig. 1a–c Comparison of folded (native) and unfolded (residual) structures in the membrane protein OmpX. **a** Bundle of 20 NMR model structures representative of the folded membrane-bound protein (Fernández et al. 2004). The highlighted regions comprise residues 73–82 (*red*) and 137–145 (*blue*). **b** NMR model structure of residues 71–84 in the folded protein, taken from the lowest-energy structure of **a**. **c** NMR model structure of same segment in the ordered state of the urea-denatured protein (Tafer et al. 2004), which is also the initial structure of simulation **A**

The goal of the present study is to investigate the structure and dynamics of the tetradecapeptide fragment of OmpX corresponding to region I of the protein (residues 73–82, extended by two residues at both termini; Fig. 1b, c) using long-timescale (0.4 μs) explicit-solvent MD simulations in 8 M aqueous urea and comparison with the experimental data. While the available experimental information in the form of 31 interproton distance upper-bounds is significant, it remains: (1) relatively scarce (underdetermination of the corresponding structural ensemble); (2) potentially representative of different conformations in equilibrium (of which the structured ones only represent a 25% fraction); (3) representative of the denatured OmpX protein (rather than of the simulated tetradecapeptide). The latter point is probably of minor relevance since experimental measurements on an octadecapeptide fragment corresponding to region I in 8 M urea evidenced identical chemical shifts compared to the full-length OmpX protein under the same conditions (Hiller et al. 2007, 2008), suggesting similar ensemble properties in both cases.

For the above reasons, the structure refinement approach (point 2 above) was not adopted in the present work. Instead, the simulations were solely driven by the physical force-field [GROMOS 45A3 (Schuler et al. 2001)], and the experimental data were used for the assessment of the resulting conformational ensemble (point 3 above). Because the experimental data [in conjunction with a highly simplified force-field restricted to atomic volume terms (Güntert et al. 1997)] had previously been used for the determination of a tentative model structure (Tafer et al. 2004) for the ordered state of region I in the urea-denatured OmpX (Fig. 1c), two separate simulations were undertaken starting from either this model (point 1 above) or from an entirely extended conformation.

By analyzing the simulation results and comparing them with the experimental data, the present study has two aims: (1) attempt to generate a physically meaningful ensemble of structures accounting for the experimental data via long-timescale unrestrained MD simulations; (2) provide insight into the process of structure formation at the atomic level for the peptide considered.

Computational details

Two explicit-solvent MD simulations (**A** and **B**) of the tetradecapeptide H^+–YRINDWASIYGVVG–O^- (charged at both termini as well as at the arginine and aspartate residues; net charge of zero) in 8 M aqueous urea were performed at 288 and 310 K, respectively, with initial configurations set to the NMR-derived model structure (Tafer et al. 2004) for the ordered state of region I and to an entirely extended conformation, respectively. This peptide sequence corresponds to a fragment of *Escherichia coli* OmpX comprising residues 71–84 (Fig. 1b, c). Note that the residues are numbered in the present article according to the sequence of the OmpX protein (i.e., the *n*th residue of the peptide is given residue number $n + 70$). The choice of charged termini was made by reference to corresponding measurements on the unblocked octadecapeptide fragment (termini ionized at neutral pH) corresponding to region I (Hiller et al. 2008).

The MD simulations were carried out using the GROMOS96 (van Gunsteren et al. 1996; Scott et al. 1999) package of programs together with the GROMOS 45A3 united-atom force field (Schuler et al. 2001; Chandrasekhar et al. 2003; Lins and Hünenberger 2005; Soares et al. 2005) for biomolecules, GROMOS-compatible urea parameters (Smith et al. 2004, 2005), and the simple-point-charge (SPC) water model (Berendsen et al. 1981). The equations of motion were integrated using the leapfrog scheme (Leimkuhler and Reich 2004) with a time step of 2 fs. All bond lengths were constrained by application of the SHAKE algorithm (Ryckaert et al. 1977) with a relative geometric tolerance of 10^{-4}. The nonbonded interactions were handled using a twin-range cutoff scheme (van Gunsteren et al. 1996) with short- and long-range cutoff radii of 0.8 and 1.4 nm, respectively, and a frequency of five time steps for updating the short-range pairlist and intermediate-range interactions. The mean effect of electrostatic interactions beyond the long-range cutoff distance was included through a reaction-field correction (Tironi et al. 1995), using an approximate relative dielectric permittivity of 30 for the urea-water mixture. The simulations were carried out under periodic boundary conditions based on a truncated-octahedral computational box. Solute and solvent degrees of freedom were separately coupled to a heat bath (Berendsen et al. 1984) with a relaxation time of 0.1 ps. The box dimensions were isotropically coupled to a pressure bath (Berendsen et al. 1984) at 1 atm, with a relaxation time of 0.5 ps and an isothermal compressibility of 45.75×10^{-5} mol nm^3 kJ^{-1}. The center of mass motion was removed every 0.2 ps. Each simulation was carried out for 0.4 μs and coordinates were written to file every 10 ps for later analysis.

Simulation **A** (at 288 K) was initiated from the NMR-derived model structure (Tafer et al. 2004) for the ordered state of region I (Fig. 1c; the temperature of 288 K corresponding to the temperature at which the NMR experiment was performed). The solute in this initial configuration was placed in a cubic box and solvated by 370 urea molecules. The resulting system was relaxed by energy minimization with constrained peptide coordinates. Peptide and relaxed urea molecules were then transferred into a truncated-octahedral box (based on a cubic box of edge length 5.58 nm) and solvated by 1,562 water molecules. The resulting system

Table 1 Experimentally derived (Tafer et al. 2004) NOE interproton distance upper bounds d [from the indicated reference, after application of multiplicity and pseudo-atom corrections (Fletcher et al. 1996; van Gunsteren et al. 1996)] employed in the analysis of the simulations, and corresponding violations Δd observed in the two simulations

Number	Residue	Atom	N_p	Residue	Atom	N_p	d (nm)	Δd (nm)	
								A	**B**
1	72ARG	HB3	1	75ASP	HN	1	0.550	0.007	0.077
2	73ILE	HN	1	76TRP	HA	1	0.550	0.348	0.217
3	73ILE	HN	1	77ALA	QB	3	0.691	0.478	0.111
4	73ILE	QG2	3	76TRP	HE1	1	0.691	−0.071	−0.103
5	73ILE	HG12	1	76TRP	HA	1	0.550	0.064	0.006
6	73ILE	QD1	3	78SER	HB2	1	0.691	−0.104	−0.134
7	74ASN	QB	2	77ALA	QB	3	0.861	0.138	−0.250
8	74ASN	QB	2	79ILE	QD1	3	0.861	−0.301	−0.358
9	75ASP	HB2	1	78SER	HB2	1	0.550	−0.134	−0.062
10	75ASP	HB2	1	79ILE	QG2	3	0.691	−0.010	-0.111
11	75ASP	HB3	1	78SER	HN	1	0.550	−0.263	−0.221
12	76TRP	HA	1	79ILE	HN	1	0.550	−0.135	−0.034
13	76TRP	HA	1	79ILE	HB	1	0.550	−0.154	0.030
14	76TRP	HA	1	79ILE	QD1	3	0.691	−0.250	−0.098
15	76TRP	HD1	1	79ILE	HA	1	0.550	0.284	0.194
16	76TRP	HD1	1	79ILE	HG12	1	0.550	0.095	0.212
17	76TRP	HE3	1	79ILE	QG2	3	0.691	−0.070	−0.013
18	76TRP	HE3	1	79ILE	HG12	1	0.550	0.098	0.015
19	76TRP	HE3	1	79ILE	QD1	3	0.691	−0.094	−0.028
20	76TRP	HZ2	1	79ILE	HG13	1	0.550	0.348	0.097
21	77ALA	HN	1	82VAL	QQG	6	0.961	0.048	0.118
22	77ALA	HA	1	80TYR	HN	1	0.550	−0.125	0.094
23	77ALA	HA	1	81GLY	HN	1	0.550	−0.086	0.292
24	77ALA	HA	1	82VAL	QQG	6	0.961	−0.202	0.265
25	77ALA	QB	3	81GLY	HN	1	0.691	0.050	0.257
26	77ALA	QB	3	82VAL	HN	1	0.691	−0.021	0.541
27	78SER	HN	1	82VAL	QQG	6	0.961	−0.092	0.167
28	78SER	HA	1	81GLY	HN	1	0.550	−0.018	0.251
29	78SER	HA	1	82VAL	QQG	6	0.961	−0.375	0.177
30	78SER	HB3	1	82VAL	QQG	6	0.961	−0.181	0.160
31	79ILE	HN	1	82VAL	QQG	6	0.961	−0.132	−0.084

The notation Q indicates a pseudo-atom site defined by the nonstereospecifically assigned protons of a methyl ($N_p = 3$) or methylene ($N_p = 2$) group. The notation QQ indicates a pseudo-atom site defined by the nonstereospecifically assigned protons of the methyl groups ($N_p = 6$) in an isopropyl sidechain (Val). The violations are calculated based on an inverse-sixth power distance averaging over 4,000 trajectory frames sampled at 0.1 ns intervals from simulations **A** and **B**. Positive values indicate a discrepancy with respect to experimental data, while negative values indicate agreement

was equilibrated by energy minimization followed by 200 ps MD simulation with position restraints on the peptide atom coordinates (harmonic force constant progressively decreased from 10 to 1 kJ mol^{-1} nm^{-2}). This equilibration was completed by 600 ps MD simulation with time-averaged distance restraints (Nanzer et al. 1995) (harmonic force constant of 1 kJ mol^{-1} nm^{-2}, linearization cutoff of 1 nm, averaging time of 10 ps using an inverse-sixth power averaging of the interproton distances) based on the experimentally derived NOE distance upper bounds (Table 1). After this relaxation, the truncated-octahedral box dimensions corresponded to a cube of edge length 5.32 nm, leading to a peptide in an 8.16 M urea solution.

Simulation **B** (at 310 K) was initiated from an entirely extended conformation of the peptide (the slightly higher temperature of 310 K was selected to enhance conformational sampling without significantly altering the simulated properties). This configuration was obtained by subjecting

the starting configuration of simulation **A** (after equilibration) to 100 ps MD simulation with position restraints on the two termini enforcing an end-to-end distance of about 4.5 nm. Although the computational box employed is too small to appropriately accommodate the peptide in this specific configuration without periodicity-induced artifacts (Weber et al. 2000; Reif et al. 2009), the configurations sampled in the two simulations (except at the very onset of simulation **B**) are much more compact. Using a larger box would have increased the computational cost (and, therefore, reduced the sampling time), with little benefit considering that only very few configurations are possibly affected by these artifacts.

The backbone atom-positional root-mean-square distances (RMSD) with respect to the NMR-derived model structures for region I, either within the urea-denatured OmpX [ordered state (Tafer et al. 2004); Fig. 1c] or within the folded membrane-bound protein [native (Fernández et al. 2004); Fig. 1b] were monitored as a function of time based on trajectory frames sampled at 0.1 ns intervals. In addition, RMSD matrices (RMSD between structures sampled at different times along the two simulations) were generated based on trajectory frames sampled at 1 ns intervals. For these analyses, both superposition of the successive trajectory frames (roto-translational least-squares fit) and RMSD calculation were based on the coordinates of the backbone trace atoms (C', C^{α}, N) of residues 73–82.

Analysis of the peptide secondary structure was performed as a function of time using a simplified version of the DSSP algorithm (Kabsch and Sander 1983). Here, a backbone-backbone hydrogen bonding pattern was first assigned using an energetic criterion involving the interaction between dipoles associated with the C = O and N–H bonds (Kabsch and Sander 1983). The resulting pattern was then used to classify residues as being part of either: no secondary structure, antiparallel β-bridge (two residues n and m with $|n - m| > 2$ characterized by hydrogen bonds from both n to m and m to n), or α-turn (sequence of residues n to $n + 4$ involving a hydrogen bond from $n + 4$ to n). The occurrences of π- and 3_{10}-turns were also monitored, but found to cover less than 1% of the generated trajectories. For this reason, they were omitted from the final analysis.

Hydrogen bonds were monitored as a function of time based on a geometric criterion. A hydrogen bond involving a donor atom (O or N), a hydrogen bound to the donor atom, and an acceptor atom (O or N) are assumed to be present if the hydrogen acceptor distance is below 0.25 nm and the donor-hydrogen-acceptor angle is larger than 135°. All backbone-backbone hydrogen bonds were monitored as well as those involving the carboxylate group of the Asp75 side chain.

Medium- or longer-range hydrophobic contacts were monitored as a function of time based on a geometric criterion depending on the peptide side chain coordinates. A (medium- or longer-ranged) hydrophobic contact between two hydrophobic residues (Ile, Trp, Tyr or Val) in the present peptide is assumed to be present if the two residues are separated by at least two other residues along the sequence and the distance between the centers of geometry of the corresponding side chains (i.e., all atoms of the residue except the backbone C', O, C^{α}, N, and H^{N} atoms) is below 0.6 nm.

NOE-derived interproton distance upper bounds were determined based on the available experimental distance estimates, set to a common value of 0.55 nm for all detectable NOEs (Tafer et al. 2004), by applying suitable multiplicity (Fletcher et al. 1996) and pseudo-atom (van Gunsteren et al. 1996) corrections for nonstereospecifically resolved protons. The resulting 31 distance upper bounds are reported in Table 1. Violations with respect to these upper bounds were determined using inverse-sixth power averaging over 4,000 trajectory frames sampled at 0.1 ns intervals. Instantaneously fulfilled bounds were also monitored as a function of time.

All analysis programs were implemented within the open-source molecular mechanics analysis package `esra` (Kräutler et al. 2005), as Mathematica notebooks calling Java routines. Visualizations were performed using PyMol (DeLano 2002).

Results and discussion

The time evolution of the RMSD with respect to the NMR-derived model structures for region I, either within the urea-denatured OmpX [ordered state (Tafer et al. 2004); Fig. 1c] or within the folded membrane-bound protein [native (Fernández et al. 2004); Fig. 1b], are displayed in Fig. 2 for simulations **A** and **B**. For ease of discussion, the helical structure in Fig. 1c is referred to as the urea-denatured model structure and the β-turn structure in Fig. 1b is referred to as the native model structure.

The RMSD matrices associated with the simulations **A** and **B**, as well as the matrix comparing **A** and **B**, are displayed in Fig. 3. Individual points in these matrices indicate the extent of structural similarity between configurations sampled at different times along the same simulation or along the two different simulations.

For simulation **A**, which was initiated from the equilibrated urea-denatured model structure, the sampled configurations remain close to this initial structure (RMSD ≈ 0.2 nm) and to each other (RMSD ≈ 0.0–0.2 nm) for about 150 ns. After this time point, the system evolves to a second region of conformational space, also characterized by mutually similar structures (RMSD ≈ 0.0–0.2 nm), but differing significantly from the urea-denatured model structure

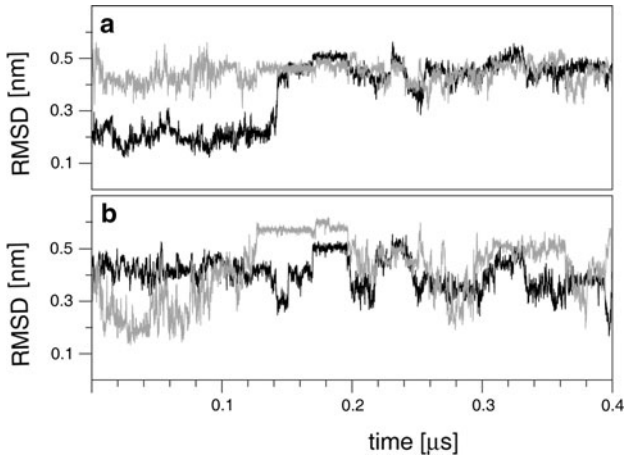

Fig. 2a, b Backbone atom-positional root-mean-square distances (RMSD) with respect to the two different reference structures for 4,000 trajectory frames sampled at 0.1 ns intervals along simulations **A** (*black*) and **B** (*gray*). **a** RMSD with respect to the (helical) urea-denatured model structure (Fig. 1c). **b** RMSD with respect to the (β-turn) native model structure (Fig. 1b). The coordinates of the backbone atoms (C′, C^{α}, N) of residues 73–82 were used for the superposition of structures and the RMSD calculation

(RMSD ≈ 0.5 nm) and from the ones sampled in the first part of the simulation (RMSD ≈ 0.2–0.7 nm). After a very brief (5 ns) transition closer to this model structure at about 250 ns, the system evolves to a third region of conformational space, more similar to the second one than to the first one. All sampled configurations remain dissimilar to the native model structure (RMSD ≈ 0.3–0.5 nm), except for one very brief transition (2 ns; RMSD ≈ 0.2 nm) towards the very end of the simulation.

For simulation **B**, which was initiated from an entirely extended configuration, an initial 90 ns exploration period is observed that involves mutually dissimilar structures (RMSD ≈ 0.2–0.3 nm). During the first 75 ns of this period, regions of conformational space bearing close similarity with the native model structure (RMSD ≈ 0.15–0.2 nm) are repeatedly visited. This initial period is followed by a sudden transition to a different region of conformational space. From this point onward, the sampled configurations differ very significantly from the ones generated in the early explorative stage of the simulation (RMSD ≈ 0.2–0.8 nm), while they remain reasonably similar to each other

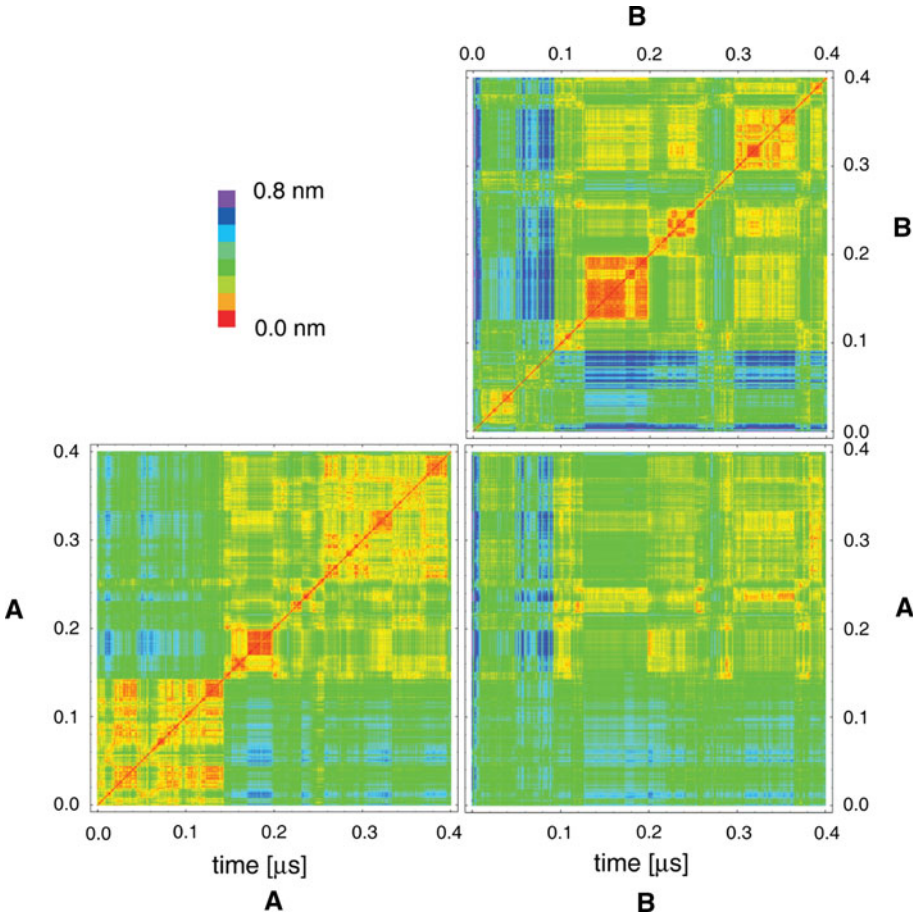

Fig. 3 Backbone atom-positional root-mean-square distance (RMSD) matrices of 400 trajectory frames sampled at 1 ns intervals along simulations **A** (*bottom left*) and **B** (*top right*), together with the corresponding RMSD matrix comparing the two simulations (*bottom right*). The coordinates of the backbone atoms (C′, C^{α}, N) of residues 73–82 were used for the superposition of structures and the RMSD calculation

(RMSD ≈ 0.0–0.3 nm). Within this second period, three distinct subensembles of mutually very similar structures (RMSD ≈ 0.0–0.2 nm) are successively encountered along the trajectory, corresponding to the time periods 130–190, 220–250, and 300–360 ns. Note, however, that the sampled configurations remain quite dissimilar to the urea-denatured model structure over the entire course of the trajectory (RMSD ≈ 0.3–0.5 nm).

The RMSD matrix comparing the structures encountered in simulations **A** and **B** clearly indicates some convergence of the two ensembles after the initial exploration phases of the two simulations (150 and 90 ns for simulations **A** and **B**, respectively). In this region, it is possible to find for each structure of simulation **A** a structure of simulation **B** with an RMSD of 0.2 nm or less, and vice versa. These configurations also appear to differ less significantly from the ones sampled in the initial phase of simulation **A** (initiated from the equilibrated urea-denatured model structure) than in the initial phase of simulation **B** (initiated from an entirely extended configuration).

The time evolution of the secondary structure along the two simulations is illustrated in Fig. 4. Both simulations appear to switch reversibly between two main structural motifs: an α-turn between Asp75 and Tyr80, sometimes extending to Gly81, Val82 or even Val83, and different variants of antiparallel β-bridges, the most persistent of which are Tyr71–Val82, Arg72–Val83, and Ile73–Tyr80 for simulation **A**, or Arg72–Val82, Ile73–Val82, Ile73–Val83, and Asp75–Gly81 for simulation **B**. Note that towards the end of simulation **A**, the α-turn rapidly and reversibly interconverts with a β-turn-like structure involving residues Asp75 to Ser78, simultaneously with the transient but longer-lasting formation of the Tyr71–Val82 and Ile73–Tyr80 antiparallel β-bridges. Such a turn can also be observed transiently towards the end of simulation **B**. Interestingly, the α-turn motif, whenever present, is nearly always stopped at Asp75 on its N-terminal end, although its C-terminal end may extend by one or two (and exceptionally three) residues beyond Tyr80. The intermittent presence of this α-turn in the second part of both simulations (after about 200 ns) is probably largely responsible for the structural similarity of the two corresponding subensembles (Fig. 3).

While the urea-denatured model structure corresponds to a fairly regular helix (Fig. 1c) with two sequential π-turns (Gly81 → Trp76 and Val82 → Ala77 backbone hydrogen bonds) and two sequential 3_{10}-turns (Trp76 → Ile73 and Ala77 → Asn74) involving residues 73–82 (Tafer et al. 2004), the native model structure of the peptide corresponds to the turn connecting the β-strands β_4 and β_5 along with a portion of strand β_5 (Fig. 1b). In the latter structure, Asp75 is located in the turn region (interacting with Arg72), while a Tyr81–Ile79 β-bridge is involved in the

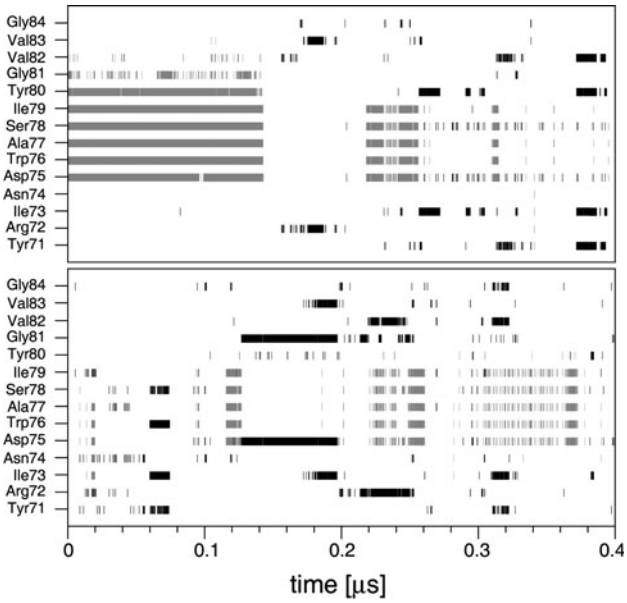

Fig. 4 Time evolution of the peptide secondary structure based on 4,000 trajectory frames sampled at 0.1 ns intervals along simulations **A** (*top*) and **B** (*bottom*). A residue can be either part of an antiparallel β-bridge (*black*) or of an α-turn (*gray*), depending on the local backbone hydrogen bonding pattern (Kabsch and Sander 1983)

formation of a β-hairpin. These exact features are observed in neither of the two simulations. However, a very similar native-like structure is seen in simulation B after about 75 ns (Figs. 2, 3, 4), with the transient formation of the corresponding register-shifted β-bridges Tyr71–Ser78 and Ile73–Trp76.

Because residue Asp75 appears to play a particular role in the conformational behavior of the peptide, it is interesting to look at the hydrogen-bonding properties of its side chain. The time evolution of all hydrogen bonds involving the carboxylate group of this residue is displayed in Fig. 5. Throughout simulation **A**, the carboxylate group of Asp75 is almost continuously hydrogen-bonded to the backbone amide groups of either Ser78 or Ala77 (the latter predominantly at the end of the simulation) or to the side chain hydroxyl group of Ser78 (predominantly at the beginning of the simulation). For simulation **B**, the hydrogen-bonding propensity of Asp75 becomes similar to that observed in simulation **A** after about 90 ns, but also includes alternative patterns (e.g., involving the backbone amide groups of Trp76, Ile79, and Tyr80). These persistent hydrogen bonds probably explain why Asp75 nearly always terminates the α-turn or helix at its N-terminus (Fig. 4), i.e., the aspartate carboxylate effectively caps the N-terminal region of the helix. In addition, the appearance of the Ala77 → Asp75 and Ser78 → Asp75 hydrogen bonds (Fig. 5) appears correlated with the formation of this α-turn (Fig. 4) in simulation **B**, suggesting that these hydrogen-bonded interactions

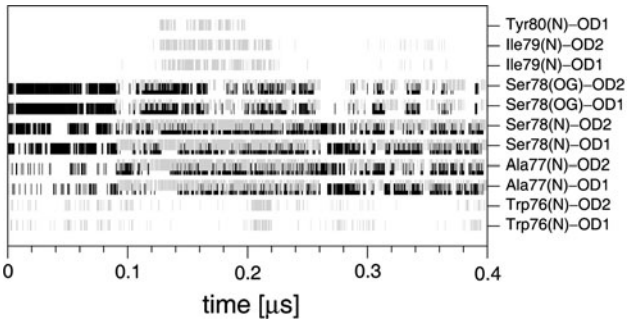

Fig. 5 Time evolution of hydrogen bonds involving the carboxylate group of the Asp75 residue based on 4,000 trajectory frames sampled at 0.1 ns intervals along simulations **A** (*black*) and **B** (*gray*). A hydrogen bond is defined by a hydrogen-acceptor distance smaller than 0.25 nm and a donor-hydrogen-acceptor angle larger than 135°. The corresponding donor residues (and atoms) and acceptor oxygen atoms of the Asp75 carboxylate (OD1 or OD2) are also indicated. Only hydrogen bonds occurring in more than 5% of the frames in either of the two trajectories are shown

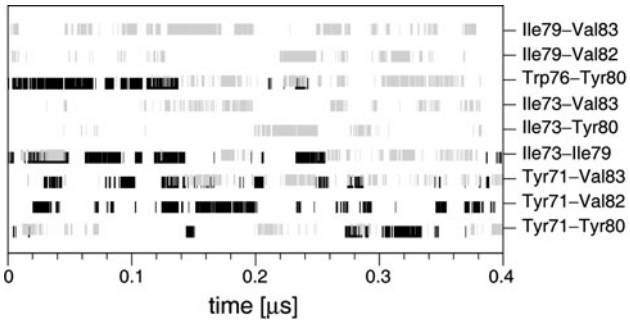

Fig. 6 Time evolution of medium- or longer-range hydrophobic contacts based on 4,000 trajectory frames sampled at 0.1 ns intervals along simulations **A** (*black*) and **B** (*gray*). A (medium- or longer-range) hydrophobic contact exists if two hydrophobic residues (Ile, Trp, Tyr, or Val) are separated by at least two other residues along the sequence and the distance between the centers of geometry of the corresponding sidechains is below 0.6 nm. Only contacts occurring during more than 10% of the frames in either of the two trajectories are shown

both induce (in the N → C direction; probably via a decrease in the overall structural flexibility and a favorable positioning of the charged side chain of Asp75 to stabilize the helix dipole) and limit (in the C → N direction, via capping) helix formation. Interestingly, no significant ion-pairing interactions are found between Asp75 and Arg72 in either simulation. This observation contrasts with both the urea-denatured and the native model structures, where Asp75 is found to be in close proximity with Arg72. Note, however, that in the urea-denatured model structure, the presence of an ion-pair is inferred from model building using weak experimental information, namely the presence of a single observed NOE involving $H_{\beta 3}$ of Arg72 and H_N of Asp75 (Table 1).

The time evolution of (medium- or longer-ranged) hydrophobic contacts between side chains along the two simulations is displayed in Fig. 6. In simulation **A**, a stable contact between Trp76 and Tyr80 is seen in the first half of the simulation. This contact is also present in the urea-denatured model structure, along with Trp76–Ile73 and Trp76–Ile79 contacts, which are not observed in the simulation. Instead, a stable contact between Ile73 and Ile79 is observed at the beginning of the simulation, as well as three contacts involving Tyr71 that appear intermittently over the course of the simulation. In simulation **B**, more numerous but less specific contacts are observed. Hydrophobic interactions between Trp76–Tyr80 and Ile79–Val82, present in the urea-denatured model structure, are observed towards the end of the simulation. However, many other contacts occur transiently that are not present in this model structure.

The above observations regarding the hydrogen-bonding properties of Asp75 and the persistent Trp76–Tyr80 hydrophobic contact are illustrated graphically in Fig. 7 based on the first 140 ns of simulation **A** (region of conformational space closest to the urea-denatured model structure; Figs. 2, 3, 4). The superimposed structures show the presence of two successive α-turns, initiated on their more rigid N-terminal part by Asp75 and extending to Tyr80 on their more flexible C-terminal part. The side chain of Asp75 folds onto the side of the helix, forming persistent hydrogen bonds with Ala77 or Ser78. The close spatial proximity of residues Trp76 and Tyr80 enables the observed hydrophobic contact, in spite of the important mobility of the two side chains. The superposition of structures sampled from the entire trajectory (Fig. 7c) confirms the presence of a persistent helical turn region (residues Asp75-Tyr80) throughout simulation **A**. A similar though significantly less clearly defined turn region can also be identified for simulation **B** (Fig. 7d).

Violations with respect to experimentally derived NOE distance upper bounds are reported numerically in Table 1. In simulation **A**, 11 distance upper bounds out of a total of 31 show positive violations, indicating a discrepancy with experimental data, with five distance upper bounds violated by more than 0.1 nm. Among these five violations, two correspond to NOEs involving Ile73 together with Trp76 or Ala77, one involves the pair Asn74–Ala77 and two the pair Trp76–Ile79. In simulation **B**, 19 distance upper bounds out of a total of 31 show positive violations, with 13 distance upper bounds violated by more than 0.1 nm. Among these 13 violations, 9 involve the residues Gly81 and Val82, 3 involve the residue Trp76, and the remaining 1 involves Ile73 and Ala77.

The time evolution of the instantaneously fulfilled NOE distance upper bounds is displayed in Fig. 8 for the two simulations. Four classes of NOE distance upper bounds may be distinguished. First, bounds 1–7 correspond to

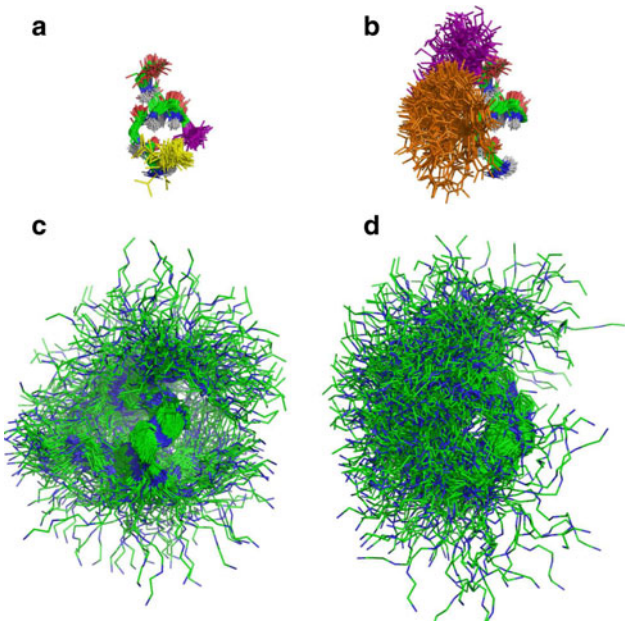

Fig. 7 **a**, **b** Overlay of 140 trajectory frames sampled at 1 ns intervals during the initial 140 ns of simulation **A**. Backbone atoms (C′, C^{α}, H, N, O) are represented together with the sidechain atoms of Asp75 and Ser78 (**a**) shown in *yellow* and *magenta*, respectively, and the sidechain atoms of Trp76 and Tyr80 (**b**) shown in *orange* and *purple*, respectively. **c**, **d** Overlay of 400 trajectory frames sampled at 1 ns intervals along simulations **A** and **B**, respectively. Backbone atoms (C′, C^{α}, N) are represented. All structures were superimposed onto the urea-denatured model structure (Fig. 1c) based on the backbone trace atoms (C′, C^{α}, N) of residues 75–80

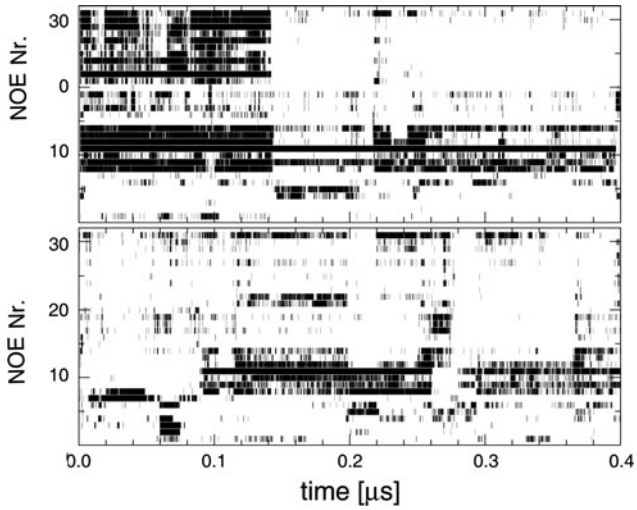

Fig. 8 Time evolution of the instantaneously fulfilled ($\Delta d \leq 0$) NOE distance upper bounds (Table 1) based on 4,000 trajectory frames sampled at 0.1 ns intervals along simulations **A** (*top*) and **B** (*bottom*)

proton pairs involving residues Arg72 to Asn74 together with residues Asp75 to Ile79. These bounds are seldom fulfilled in simulation **A** and, to a lesser extent, in simulation **B**, resulting in significant average violations over the entire trajectories (Table 1). Transient fulfillment of these bounds generally correlates with the formation of secondary structure involving residues Tyr71 to Asn74 (Fig. 4). With the exception of bound 6, which is fulfilled on average in both simulations, these bounds correspond to residue pairs n to $n + 3$ or n to $n + 4$, compatible with helical structure in that region found in the urea-denatured model structure but not in the simulations. Second, bounds 8–14 correspond to proton pairs involving residues Asp75 and Trp76 together with residues Ser78 and Ile79, except for bound 8, which corresponds to the pair Asn74–Ile79. The fulfillment of these bounds is generally correlated with the presence of the α-turn between Asp75 and Ile79 (Fig. 4). Third, bounds 15–20 correspond to proton pairs involving residues Trp76 together with Ile79. Some of these bounds appear to be incompatible with the formation of the α-turn, as apparent in the initial 140 ns of simulation **A**. Fourth, bounds 21–31 correspond to proton pairs involving the residues Ala77 to Ile79 and Tyr80 to Val82. These are generally satisfied when the α-helical structure extends from Asp75 up to Val82, which occurs during the first 140 ns of simulation **A** and almost never in simulation **B** (Fig. 4). Correspondingly, large positive average violations over the entire trajectory for the last set of bounds are encountered for simulation **B** only (Fig. 8). Violations were also monitored for the combined trajectories of simulations **A** and **B**, as well as for a subset consisting of 1,400 frames corresponding to the first 140 ns of simulation **A** (data not shown). The corresponding results were found to be very similar to those based on simulation **A** (Fig. 8).

Conclusions

In the present study, two long-timescale (0.4 μs) unrestrained explicit-solvent simulations of a tetradecapeptide fragment of the protein OmpX under strongly denaturing conditions (8 M aqueous urea) were reported and analyzed. These two simulations were initiated either from an NMR-model structure for the ordered state of the corresponding segment in the urea-denatured protein (Tafer et al. 2004) (simulation **A**) or from an entirely extended configuration (simulation **B**). The main observations can be summarized as follows:

1. In spite of the long timescale of 0.4 μs reached in the present simulations, the generated ensembles do not sample the entire configurational space accessible to the peptide. In particular, major transitions between different conformational basins occur on the 0.1 μs timescale and the influence of the selected starting configuration remains visible throughout the two trajectories (note that the small temperature difference of 22 K

between the two simulations may also represent a secondary factor).

2. Both simulations show in multiple events the reversible formation and disruption of α-helical secondary structure [similar to the NMR model structure for the ordered state of the corresponding segment in the urea-denatured protein (Tafer et al. 2004)], and β-turn secondary structure [similar to the secondary structure of this segment in the folded membrane-bound protein (Fernández et al. 2004)].
3. The formation of an α-helical secondary structure is consistent with the occurrence of a previously postulated (Tafer et al. 2004) (but still experimentally uncharacterized) hydrophobic contact between Trp76 and Tyr80. Such a contact is observed in both simulations, its occurrence being strongly correlated with the formation of α-helical structure.
4. The α-helix, whenever formed, predominantly involves a single turn. It is sometimes extended by one, two, or three residues at its C-terminus. In contrast, the N-terminal residue is nearly always Asp75, which appears to both promote helix formation in the N → C direction and block it in the C → N direction. Events of helix formation are correlated with the formation of hydrogen bonds between the side chains of Asp75 and Ser78, and the presence of a stable hydrophobic contact between Trp76 and Tyr80. These features are compatible with the NMR model structure for the ordered state of the segment in the urea-denatured protein (Tafer et al. 2004) and may thus contribute to the stability of this particular fold.
5. The β-turn, whenever formed, involves a wide variety of possible β-bridges enclosing residues Asp75 to Ser78 as a loop. These features are qualitatively consistent with the structure of the segment in the folded membrane-bound protein (Fernández et al. 2004), although they suggest a considerably larger conformational variability of accessible β-turn structures at equilibrium. This is not surprising in view of the different environmental conditions considered: the simulations involve a tetradecapeptide fragment under denaturing conditions of 8 M urea in aqueous solution, while the native structure accounts for the conformation of the corresponding segment in the entire native protein, itself embedded within a lipid membrane.
6. The agreement between the simulation results and experimental NOE-derived interproton distance upper bounds is qualitative but not quantitative. The predominant cause for this quantitative discrepancy is probably an insufficient conformational sampling in the simulations, with possible force field inaccuracies representing a secondary factor. Nearly all distance bounds are satisfied at different points along the two simulations, but never all of them simultaneously. Due to the inverse-sixth power distance averaging, the presence of a limited subset of configurations characterized by a short interproton distance is sufficient to give rise to the detection of an NOE cross-peak (Bürgi et al. 2001). Two consequences are that (1) the satisfaction of NOE-derived upper-bounds requires a more complete sampling of the configurational space compared to that achieved in the present simulations, and (2) these bounds may be satisfied independently by different subsets of configurations without ever being simultaneously fulfilled in one single set. In particular, we note that (1) the 8 bounds involving residues Arg72 to Asn74 are only satisfied in a small fraction of the sampled configurations in either of the two simulations and the configurations responsible for these 8 bounds may thus not have been sampled, and (2) 4 out of the 9 bounds involving proton pairs in residues Trp76 together with Ile79 are violated in most configurations, even those involving an α-helical structure for residues Asp75 to Val82 and the satisfaction of the 19 other bounds involving these residues (i.e., the configurations responsible for these 4 bounds may actually differ from the α-helical ones; the hydrophobic contact involving Trp76 and Ile79 is also not found in the simulations, even though this contact would be compatible with an α-helical geometry in that region).

In summary, even the long (0.4 μs) simulation timescale involved in the present study is still insufficient to provide a complete sampling of the conformational ensemble accessible to the tetradecapeptide on the experimental timescale (point 1). A number of important regions of this space appear to be lacking, resulting in significant NOE violations (point 6). However, this simulation timescale is sufficient to encompass conformations representative for the corresponding segment in both the native and urea-denatured OmpX protein (point 2). Because simulations provide information inaccessible to experiment (distributions rather than averages; atomic level and femtosecond resolution), the analysis of the sampled configurations provides further insight into the detailed structural and dynamic features of the considered peptide (points 3, 4, and 5; which are compatible with the experimental results but could not have been inferred from these experiments alone). In particular, the NMR experiments detected a 25% population of a nonnative α-helical ordered state, with the remaining 75% population corresponding to species producing no NOE cross-peaks. The MD simulations suggest that the corresponding conformational ensemble may also encompass a significant population of native-like (although significantly less ordered) β-turn structures, with an interconversion timescale on the order of microseconds or longer.

This information may in turn be used as a guide in the design of new experiments to elucidate the molecular mechanisms involved in the folding of the OmpX membrane protein.

Acknowledgments S.H. was supported by grants to Kurt Wüthrich from the Swiss National Science Foundation and the ETH Zürich, through the NCCR Structural Biology. Kurt Wüthrich is kindly acknowledged for his interest in this work and for his continuous support.

References

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, pp 331–342

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

Bürgi R, Pitera J, van Gunsteren WF (2001) Assessing the effect of conformational averaging on the measured values of observables. J Biomol NMR 19:305–320

Chandrasekhar I, Kastenholz M, Lins RD, Oostenbrink C, Schuler LD, Tieleman DP, van Gunsteren WF (2003) A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field. Eur Biophys J 32:67–77

Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE (1999) Peptide folding: when simulation meets experiment. Angew Chem Int Ed 38:236–240

DeLano WL (2002) The PyMOL molecular graphics system. http://www.pymol.org/

Fennen J, Torda AE, van Gunsteren WF (1995) Structure refinement with molecular dynamics and a Boltzmann-weighted ensemble. J Biomol NMR 6:163–170

Fernández C, Hilty C, Bonjour S, Adeishvili K, Pervushin K, Wüthrich K (2001) Solution NMR studies of the integral membrane proteins OmpX and OmpA from *Escherichia coli*. FEBS Lett 504:173–178

Fernández C, Hilty C, Wider G, Güntert P, Wüthrich K (2004) NMR structure of the integral membrane protein OmpX. J Mol Biol 336:1211–1221

Fletcher CM, Jones DNM, Diamond R, Neuhaus D (1996) Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. J Biomol NMR 8:292–310

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298

Hiller S, Wasmer C, Wider G, Wüthrich K (2007) Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. J Am Chem Soc 129:10823–10828

Hiller S, Wider G, Imbach LL, Wüthrich K (2008) Interactions with hydrophobic clusters in the urea-unfolded membrane protein OmpX. Angew Chem Int Ed 47:977–981

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Kleinschmidt JH, Tamm LK (2002) Secondary and tertiary structure formation of the β-barrel membrane protein OmpA is synchronized and depends on membrane thickness. J Mol Biol 324:319–330

Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. Science 295:1719–1722

Kräutler V, Kastenholz M, Hünenberger PH (2005) The esra molecular mechanics analysis package. http://sourceforge.net/projects/esra/

Leimkuhler B, Reich S (2004) Simulating hamiltonian dynamics. Cambridge University Press, Cambridge

Lins RD, Hünenberger PH (2005) A new GROMOS force field for hexopyranose-based carbohydrates. J Comput Chem 26:1400–1412

McCarney ER, Kohn JE, Plaxco KW (2005) Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. Crit Rev Biochem Mol Biol 40:181–189

Montelione GT, Zheng D, Huang YJ, Gunsalus KC, Szyperski T (2000) Protein NMR spectroscopy in structural genomics. Nat Struct Biol 7:S982–S985

Mori H, Ito K (2001) The Sec protein-translocation pathway. Trends Microbiol 9:494–500

Nanzer AP, van Gunsteren WF, Torda AE (1995) Parametrisation of time-averaged distance restraints in MD simulations. J Biomol NMR 6:313–320

Pautsch A, Vogt J, Model K, Siebold C, Schulz GE (1999) Strategy for membrane protein crystallization exemplified with OmpA and OmpX. Proteins 34:167–172

Radford SE, Dobson CM, Evans PA (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature 358:302–307

Reif MM, Kräutler V, Kastenholz MA, Daura X, Hünenberger PH (2009) Molecular dynamics simulations of a reversibly-folding beta-heptapeptide in methanol: influence of the treatment of long-range electrostatic interactions. J Phys Chem B 113:3112–3128

Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341

Schuler L, Daura X, van Gunsteren WF (2001) An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. J Comput Chem 22:1205–1218

Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, van Gunsteren WF (1999) The GROMOS biomolecular simulation program package. J Phys Chem A 103:3596–3607

Shortle D (1993) Denatured states of proteins and their roles in folding and stability. Curr Opin Struct Biol 3:66–74

Shortle D (1996) The denatured state (the other half of the folding equation) and its role in protein stability. FASEB J 10:27–34

Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. Science 293:487–489

Smith LJ, Berendsen HJC, van Gunsteren WF (2004) Computer simulation of urea-water mixtures: a test of force field parameters for use in biomolecular simulation. J Phys Chem B 108:1065–1071

Smith LJ, Jones RM, van Gunsteren WF (2005) Characterization of the denaturation of human α-lactalbumin in urea by molecular dynamics simulations. Proteins 58:439–449

Soares TA, Hünenberger PH, Kastenholz MA, Kräutler V, Lenz T, Lins RD, Oostenbrink C, van Gunsteren WF (2005) An improved nucleic acid parameter set for the GROMOS force field. J Comput Chem 26:725–737

Tafer H, Hiller S, Hilty C, Fernández C, Wüthrich K (2004) Non-random structure in the urea-unfolded *Escherichia coli* outer membrane protein X (OmpX). Biochemistry 43:860–869

Tamm LK, Arora A, Kleinschmidt JH (2001) Structure and assembly of β-barrel membrane proteins. J Biol Chem 276:32399–32402

Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. J Chem Phys 102:5451–5459

van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott W, Tironi IG (1996) The GROMOS96 manual and user guide. Verlag der Fachvereine, Zürich

Vogt J, Schulz GE (1999) The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. Structure 7:1301–1309

Weber W, Hünenberger PH, McCammon JA (2000) Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: influence of artificial periodicity on peptide conformation. J Phys Chem B 104:3668–3675

Wider G, Wüthrich K (1999) NMR spectroscopy of large molecules and multimolecular assemblies in solution. Curr Opin Struct Biol 9:594–601

Wimley WC (2003) The versatile β-barrel membrane protein. Curr Opin Struct Biol 13:404–411

Wüthrich K (1995) NMR—this other method for protein and nucleic acid structure determination. Acta Crystallogr D 51:249–270

Wüthrich K (2003) NMR studies of structure and function of biological macromolecules (Nobel lecture). Angew Chem Int Ed 42:3340–3363