Nonrandom Structure in the Urea-Unfolded \textit{Escherichia coli} Outer Membrane Protein X (OmpX)\textsuperscript{†}

Hakim Tafer, Sebastian Hiller,* Christian Hilty, César Fernández,\textsuperscript{1} and Kurt Wüthrich

\textit{Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland}

Received September 15, 2003; Revised Manuscript Received November 20, 2003

ABSTRACT: On the basis of sequence-specific resonance assignments for the complete polypeptide backbone and most of the amino acid side chains by heteronuclear nuclear magnetic resonance (NMR) spectroscopy, the urea-unfolded form of the outer membrane protein X (OmpX) from \textit{Escherichia coli} has been structurally characterized. \textsuperscript{1}H–\textsuperscript{1}H nuclear Overhauser effects (NOEs), dispersion of the chemical shifts, amide proton chemical shift temperature coefficients, amide proton exchange rates, and \textsuperscript{15}N–\textsuperscript{1}H NOEs show that OmpX in 8 M urea at pH 6.5 is globally unfolded, but adopts local nonrandom conformations in the polypeptide segments of residues 73–82 and 137–145. For these two regions, numerous medium-range and longer-range NOEs were observed, which were used as the input for structure calculations of these polypeptide segments with the program DYANA. The segment 73–82 forms a quite regular helical structure, with only loosely constrained amino acid side chains. In the segment 137–145, the tryptophan residue 140 forms the core of a small hydrophobic cluster. Both nonrandom structures are present with an abundance of about 25% of the protein molecules. The sequence-specific NMR assignment and the physicochemical characterization of urea-denatured OmpX presented in this paper are currently used as a platform for investigations of the folding mechanism of this integral membrane protein.

Protein folding into the unique three-dimensional structure required for biological activity remains one of the central challenges in structural biology. It can be assumed that detailed insight into any protein folding mechanism will require characterization of all important states along the folding pathway, including the “denatured state”, which may possibly include local residual nonrandom structure and limited internal dynamics. In this context, studies under denaturing conditions are of interest, since these can enable the observation of nonrandom conformations that are not sufficiently accumulated for NMR\textsuperscript{3} studies under solution conditions that favor folding, although they may appear early in the folding pathway. Residual nonrandom structure and local intramolecular interactions found under strongly denaturing solution conditions may thus be indicative of nucleation sites for the folding process under physiological conditions (1–5). For membrane-associated proteins, structural characterization of non-native states is further of interest because of their potential role in the transport of proteins across the membranes and in cellular processes such as signal transduction. Added interest in such studies comes from the demonstration that numerous proteins are intrinsically unstructured in their biologically functional states, and only become structured upon binding to their biological target (4, 6).

Sequence-specific resonance assignments (7) form the basis for a detailed analysis of NMR parameters such as chemical shifts, NOEs, spin–spin couplings, spin relaxation rates, and amide proton exchange data in folded as well as partially or fully unfolded proteins. NMR assignments have been reported for a limited number of denatured proteins, several of which contain significant residual nonrandom structure under the denaturing conditions used (3, 8–13). Although for some of these proteins the assignments for the unfolded state have been obtained by magnetization transfer from the folded state (14), triple-resonance NMR with the uniformly \textsuperscript{13}C,\textsuperscript{15}N-labeled polypeptides (15, 16) provides a more generally applicable method for obtaining resonance assignments of proteins in non-native states. Thereby, more use of \textsuperscript{15}N and \textsuperscript{13}C resonances than in folded proteins is advisable, since these two atom positions are significantly influenced by the residue type as well as the amino acid sequence, and are therefore quite well-dispersed even in unfolded states (17). This paper describes NMR assignments of the \textsuperscript{13}C,\textsuperscript{15}N-labeled unfolded form of OmpX in 8 M aqueous urea solution. The unfolded state of OmpX was then characterized using \textsuperscript{1}H–\textsuperscript{1}H NOEs, and supplementary NMR experiments bearing on local conformational equilibria and variable internal mobility along the polypeptide chain. Finally, a structure calculation was performed for two segments of the polypeptide chain exhibiting nonrandom behavior.

MATERIALS AND METHODS

Production of Uniformly \textsuperscript{13}C,\textsuperscript{15}N-labeled OmpX and NMR Sample Preparation. Uniformly \textsuperscript{13}C,\textsuperscript{15}N-labeled OmpX was...
expressed in Escherichia coli as reported in ref 18 for studies in DHPc micelles. For the NMR studies in 8 M urea, the protein was further purified by anion-exchange chromatography. A HiTrap Q sepharose HP column with a volume of 5 mL used on an Äkta Prime chromatographic system (Amersham Pharmacia Biotech) was equilibrated with buffer A (8 M urea, 20 mM Tris-HCl, pH 8.5). The protein was loaded onto the column, washed with buffer A, and eluted with a 60 mL gradient of 0–100% of buffer B (8 M urea, 20 mM Tris-HCl, 1 M NaCl, pH 8.5) at a flow rate of 1 mL/min. The fractions containing OmpX were transferred to a Centricon ultrafiltration device with 3 kDa molecular weight cut-off (Millipore), and the buffer was exchanged against the NMR buffer (8 M urea, 20 mM phosphate, 0.1 mM NaCl) over a period of 34 h at 4 °C. The chemical shift deviations from random coil values in urea-unfolded OmpX were determined from a series of 200 2D [15 N, 1 H]-HSQC spectra recorded in D2O. The rate constants were obtained from a least-squares fit of a single-exponential function to the peak volumes. The protection factors were calculated according to Bai et al. (26).

Optimization of NMR Sample Conditions. To investigate the influence of the sample conditions on the quality of the NMR spectra, a series of 2D [15 N, 1 H]-HSQC spectra were recorded. The parameters varied were the pH (5.8–6.5), the temperature (15–30 °C), and the denaturant (6 M guanidinium hydrochloride or 8 M urea). Over the range of conditions tested, only minor effects of pH and temperature on the signal-to-noise ratio and the spectral resolution could be seen in the NMR spectra. Urea as a denaturant proved to be a better choice than guanidinium hydrochloride, both for the purification with ion-exchange chromatography and for NMR spectroscopy, primarily because it is nonionic.

NMR Spectroscopy for Sequential Resonance Assignments. All the experiments used to obtain the sequential backbone and side chain resonance assignments were measured on Bruker DRX-600, DRX-750, or DRX-800 spectrometers equipped with triple resonance pulsed field gradient probes with actively shielded z-gradients. Unless otherwise indicated, all experiments were performed at a temperature of 15 °C. The 1H carrier was centered on the water resonance, the 15N carrier was set to 118 ppm, and the 13C carrier to 176 ppm for carbonyl carbons, to 53 ppm for α-carbons, to 43 ppm when both 13Cα and 13Cβ resonances were to be excited, and to 125 ppm for aromatic carbons. The proton chemical shifts were referenced to internal DSS (19), and those for nitrogen-15 and carbon-13 were indirectly referenced (20). All experiments used for obtaining the resonance assignments are listed in Table 1. The spectra were processed with the program PROSA (21) and analyzed with XEASY (22).

Table 1: Acquisition Parameters Used for NMR Experiments with OmpX in 8 M Aqueous Urea

<table>
<thead>
<tr>
<th>Experiment (mixing time; 1H frequency) (ref)</th>
<th>t1,max</th>
<th>t2,max</th>
<th>t3,max</th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
</tr>
</thead>
<tbody>
<tr>
<td>[15N,1H]-HSQC (750 MHz) (49)</td>
<td>152 (N)</td>
<td>97 (H)</td>
<td>300</td>
<td>1024</td>
<td>1024</td>
<td></td>
</tr>
<tr>
<td>(H)N(CO)CA/NH (750 MHz) (50)</td>
<td>22 (N)</td>
<td>26 (N)</td>
<td>105 (H)</td>
<td>43</td>
<td>54</td>
<td>1024</td>
</tr>
<tr>
<td>HN(CA)CO (750 MHz) (57)</td>
<td>18 (C)</td>
<td>20 (N)</td>
<td>114 (H)</td>
<td>45</td>
<td>50</td>
<td>1024</td>
</tr>
<tr>
<td>HNCO (750 MHz) (52)</td>
<td>24 (C)</td>
<td>27 (N)</td>
<td>114 (H)</td>
<td>75</td>
<td>54</td>
<td>1024</td>
</tr>
<tr>
<td>ct-HNCA (750 MHz) (53)</td>
<td>26 (C)</td>
<td>28 (N)</td>
<td>105 (H)</td>
<td>150</td>
<td>55</td>
<td>1024</td>
</tr>
<tr>
<td>HN(CO)CA (600 MHz) (54)</td>
<td>13 (C)</td>
<td>23 (N)</td>
<td>114 (H)</td>
<td>50</td>
<td>45</td>
<td>1024</td>
</tr>
<tr>
<td>HNCACB (600 MHz) (55)</td>
<td>11 (C)</td>
<td>15 (N)</td>
<td>114 (H)</td>
<td>91</td>
<td>30</td>
<td>1024</td>
</tr>
<tr>
<td>15N-resolved [13C,1H]-TOCSY (600 MHz) (49, 56)</td>
<td>26 (H)</td>
<td>28 (N)</td>
<td>107 (H)</td>
<td>190</td>
<td>45</td>
<td>1024</td>
</tr>
<tr>
<td>15N-resolved [13C,1H]-N-NOESY (t= 100 ms; 800 MHz) (57)</td>
<td>21 (H)</td>
<td>21 (N)</td>
<td>107 (H)</td>
<td>200</td>
<td>50</td>
<td>1024</td>
</tr>
<tr>
<td>ct-[15C,1H]-HSQC for aromatics (800 MHz) (58)</td>
<td>25 (C)</td>
<td>92 (H)</td>
<td>415</td>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H(CO)NH-TOCSY (800 MHz) (59)</td>
<td>6 (C)</td>
<td>29 (N)</td>
<td>76 (H)</td>
<td>65</td>
<td>59</td>
<td>1024</td>
</tr>
<tr>
<td>13C-resolved [13C,1H]-N-NOESY for aromatics (t= 120 ms; 800 MHz) (59)</td>
<td>19 (H)</td>
<td>16 (N)</td>
<td>107 (H)</td>
<td>90</td>
<td>40</td>
<td>1024</td>
</tr>
<tr>
<td>ct-[13C,1H]-HSQC for aromatics (800 MHz) (58)</td>
<td>33 (C)</td>
<td>92 (H)</td>
<td>200</td>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2D 1H-TOCSY-relayed ct-[13C,1H]-HSQC for aromatics (800 MHz) (61)</td>
<td>52 (C)</td>
<td>92 (H)</td>
<td>830</td>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13C-resolved [13C,1H]-N-NOESY for aromatics (t= 120 ms; 800 MHz) (60)</td>
<td>20 (H)</td>
<td>10 (C)</td>
<td>107 (H)</td>
<td>165</td>
<td>60</td>
<td>1024</td>
</tr>
</tbody>
</table>

a (N), (C), and (H) identify the time axes for 15N, 13C, and 1H, respectively. b Four scans per increment of the indirect dimensions were accumulated for all experiments except HN(CA)CO, ct-[13C,1H]-HSQC for aromatics and 2D 1H-TOCSY-relayed ct-[13C,1H]-HSQC for aromatics, which used eight scans.
as input, structure calculations were performed independently for the two polypeptide segments 37-82 and 137-145, using the program DYANA (29). To limit chain-end effects, two extra residues were included at the N- and C-termini of these polypeptide segments. DYANA calculations started from 500 conformers with random torsion angle values, and 8000 torsion angle dynamics steps were performed using the standard simulated annealing procedure. Finally, the 20 conformers with lowest final target function values were subjected to energy minimization with the program OPALp (30, 31) and used to represent the NMR solution structures.

RESULTS

NMR Sample Preparation. Following the procedures described in Materials and Methods, u-13C,15N-labeled OmpX was expressed in E. coli, isolated from inclusion bodies, and purified by anion exchange chromatography. The yield of protein after solubilization in 8 M urea and after the anion exchange chromatography step was about 50 mg and 38 mg per liter of minimal medium, respectively. High purity NMR samples were obtained, as judged by SDS-PAGE and NMR spectroscopy. A 2D [15N,1H]-HSQC spectrum of doubly labeled OmpX in 8 M urea recorded under the conditions chosen for this study (3 mM u-13C,15N-labeled OmpX in 8 M urea, 20 mM sodium phosphate buffer at pH 6.5, T = 15 °C) is shown in Figure 1. The protein stability was monitored by periodic recording of [15N,1H]-HSQC spectra. Under the conditions used, no additional peaks and no significant variations of the relative peak intensities were observed during this study.

Sequence-Specific Resonance Assignments. In urea-unfolded OmpX, the chemical shift dispersion of the 1H N protons is less than 0.8 ppm, which contrasts with a dispersion of about 3.2 ppm in the folded form in DHPC micelles (18). Small chemical shift dispersion was also observed for 13Cα and 13Cβ. However, the dispersion of the backbone 15N chemical shifts (21 ppm) and 13C′ chemical shifts (7 ppm) was large enough to yield spectra with well-resolved resonances for the majority of the residues in the unfolded protein. Therefore, although supporting information was obtained from a variety of other NMR spectra (Table 1), the sequential resonance assignment was achieved mainly

Figure 1: Contour plots of a 2D [15N,1H]-HSQC spectrum of uniformly 13C,15N-labeled OmpX in 8 M urea in 95% H2O/5% D2O (1H-frequency 750 MHz, protein concentration = 3 mM, pH = 6.5, T = 15 °C). The cross-peaks are labeled with the one-letter amino acid symbol and the sequence position. Three resonances corresponding to a minor protein species lacking the residue Ala1 are identified with a prime (′) (see text). Side chain resonances of Asn and Gln residues are labeled with “NH2 N,Q”. The cross-peaks marked ‘N–H ‘R’ belong to Arg side chain N–H groups. Indole N1–H1 resonances of Trp residues are shown in the inset. The arrow indicates the cross-peak corresponding to Gly64 (see text).

Figure 2: (a) 3D (H)N(COCA)NH experiment with OmpX in 8 M urea (same sample as in Figure 1; see Table 1 for the NMR parameters used). The strips were taken at the 15N chemical shifts indicated at the top, and are centered about the corresponding 1H chemical shifts. At the top of each strip, the sequence-specific assignment is indicated by the one-letter amino acid symbol and the sequence position. Vertical and horizontal broken lines connect the intraresidual and sequential connectivities, and thus outline the sequential pathway for the segment comprising the amino acid residues Ile40 to Ser49. Cross-peaks with negative intensity are represented with broken contour lines. (b) 3D HN(CA)CO experiment, same presentation as in panel a.
using the 3D (H)N(COCA)NH experiment (Figure 2a), and the combination of the 3D HN(CO)CA (Figure 2b) and 3D HNCO experiments. The two independent pathways for obtaining sequential connectivities complemented each other quite well in the sense that ambiguity in one of them could often be resolved with the help of the second pathway. These three experiments yielded a complete set of sequential connectivities between the backbone HN, 15N, and 13C′ resonances of all pairs of neighboring residues. 3D 15N-resolved [1H,1H]-TOCSY and the 3D 15N-resolved [1H,1H]-NOESY were used to confirm these sequential assignments, and as usual these experiments were frequently consulted to help resolve ambiguities during the assignment procedure. Although a large number of sequential d_{HN}, d_{NN}, and d_{NN} NOEs could be identified (Figure 3), an attempt at obtaining complete sequential assignments via 1H−1H NOE connectivities alone would have been successful, due to the limited 1H, 1H, and 1H chemical shift dispersion. The 3D ct-HNCA, HN(CO)CA, and HNCACB experiments were then used to extend the resonance assignment to 13Cα and 13Cβ for a preliminary residue-type identification. Thereby, the superior resolution of the ct-HNCA spectrum, combined with the information from the HN(CO)CA experiment, yielded most of the 13Cα assignments. The 3D HNCACB experiment yielded identifications of numerous 13Cα−13Cβ fragments, which provided the starting platform for complete identification of the amino acid side chain spin systems.

The 3D 15N-resolved [1H,1H]-TOCSY spectrum together with 3D H(CO)NH-TOCSY constituted the basis for the aliphatic side chain proton assignments. To obtain the side chain 13C chemical shifts, a 3D (H)C(CO)NH-TOCSY spectrum was used, and the results obtained were verified by correspondence with the peaks in a 2D ct-[13C,1H]-HSQC spectrum. With this procedure, all aliphatic 13CH groups of OmpX in 8 M urea were completely assigned. The labile side chain protons of Arg, Asn, and Gln could not be individually assigned, due to chemical shift degeneracy of the peripheral side chain proton resonances.

To assign the aromatic side chain resonances, a 2D 1H-TOCSY-relayed ct-[13C,1H]-HSQC, a 2D ct-[13C,1H]-HSQC, and a 3D 13C-resolved [1H,1H]-NOESY experiment were used. Due to extensive signal overlap in the aromatic region
of the \(^{13}\text{C},^{1}\text{H}\)-HSQC spectrum and to chemical shift degeneracy of the \(^{13}\text{C}\)-\(^{1}\text{H}\) resonances of the aromatic side chains, only the side chains of Trp76 and Trp140, and the \(^{1}\text{H}\) resonances of His100 were unambiguously assigned.

In the NMR sample, about 25% of the protein lacked the first amino acid, Ala1, probably due to proteolytic digestion. This minor protein species is manifested in the NMR spectra by separate resonance lines for the residues 3, 4, and 5 (Figure 1).

**Structural Characterization of Urea-Unfolded OmpX by NMR.** The \(^{15}\text{N}\)-resolved \(^{1}\text{H},^{1}\text{H}\)-NOESY spectrum (mixing time \(\approx 100\) ms) of OmpX in 8 M urea showed 783 unambiguously assigned NOEs, among which 95% were intraresidual, sequential, and medium-range \(d(i,i+2)\) NOEs (Figure 3). This coincides with model considerations that predict that sequential \(d_{\text{NN}}\), \(d_{\text{RN}}\), and \(d_{\text{NN}}\) NOEs should have detectable intensities for all residues in “random coil” polypeptide chains \((32)\). In contrast to this homogeneous distribution of sequential and \(d(i,i+2)\) NOEs throughout the sequence, the presence of longer-range NOEs is limited to two short segments of the OmpX sequence, which comprise the residues 73–82 (I) and 137–145 (II) (Figures 3 and 4). This indicates that the segments I and II adopt sizable populations of nonrandom local conformations in 8 M urea solution. Residues with large \(^{1}\text{H}\) and \(^{1}\text{H}\) chemical shift differences from the random coil values are mostly located in the segments I and II of the polypeptide sequence (Figures 3 and 5a,b). As an illustration, the Figure 6 compares the statistical \(^{1}\text{H}\) chemical shifts of two residue types with observed values in a random coil and a structured polypeptide segment of OmpX. For Gly145, which is located in segment II, two well-resolved \(^{1}\text{H}\) and \(^{1}\text{H}\) resonances were observed (Figure 6a). Ile79 in segment I shows a significant chemical shift deviation from the random coil values for the side chain resonances (Figure 6b). Of the three residues with temperature coefficients of the amide proton chemical shifts smaller than 4 ppb/K (Figure 5c), two are located in the region I, again indicating local nonrandom structure \((33, 34)\). The amide proton exchange protection factors vary between 2.8 and 5.8 along the sequence, with an average value of 4.4, and thus provide no evidence for the existence of stable hydrogen bonds involving amide protons. Only Thr2 and...
Ser3 showed negative heteronuclear $^{15}$N-$^1$H-NOE values, indicating increased dynamics on the subnanosecond time scale. All other residues have $^{15}$N-$^1$H-NOEs close to the average value of 0.33 observed for OmpX in 8 M urea, with no significant clustering along the sequence of values deviating significantly from the mean. In contrast to the N-terminus, the C-terminal tripeptide did not show negative $^1$H-$^1$H-NOE values, indicating that these residues adjacent to the structured segment II have similar mobility as the residues 4–146.

For the two polypeptide segments I and II (Figure 3), we performed structure calculations with the program DYANA. For the segment I, a total of 31 $d(i,i+3)$, $d(i,i+4)$, and $d(i,i+5)$ NOE cross-peaks were observed in 3D $^{15}$N- and $^{13}$C-resolved $^{1}$H, $^{1}$H-NOESY spectra (Figures 3 and 7). Figure 7 shows that with this input the polypeptide backbone of residues 73–81 is well defined, but all amino acid side chains are conformationally disordered. This is reflected by RMSD values of 1.06 ± 0.28 Å for the backbone atoms, and 1.79 ± 0.37 Å for all heavy atoms. The residual DYANA target function value and the residual constraint violations were very small, indicating that the input data represent a self-consistent set, but do not represent tight constraints on the conformation. The hydrophilic side chains of Asn74, Asp75, and Ser78 point toward the solvent. The side chains of Ile73, Trp76, and Ile79 are located close to each other as evidenced by some NOEs between them (Figure 7b), and by the observation that the $^1$H resonances of Ile79 are shifted upfield by ring current effects from Trp76 (Figure 6b). There were also NOEs from the Trp76 side chain to other protons in the aromatic chemical shift region. These other aromatic proton shifts most likely belong to Tyr80 but could not be unambiguously assigned and were thus not considered for the structure calculations. The side chains of Trp76 and Tyr80 are therefore not well defined in the structure, although there might be a hydrophobic interaction between them. The polypeptide backbone is well defined in a 3α-helical conformation. In 17 of the 20 conformers, at least one of the two backbone hydrogen bonds Trp76NH–O Ile73 and Ala77 NH–O Asn74 is present, and in nine of the conformers both of them are formed. The formation of these backbone hydrogen bonds is also supported by the observation of low-temperature coefficients for the residues 76 and 81 (34). The picture emerges that the hydrophobic Ile73–Trp76 interactions as well as for technical reasons unidentified Trp76–Tyr80 interactions drive the backbone to adopt its helical conformation.

For the segment II, a total of 29 $d(i,i+3)$, $d(i,i+4)$, and $d(i,i+5)$ NOE cross-peaks were observed (Figures 3 and 8). Figure 8 shows that the polypeptide backbone of residues 137–145 and the orientation of the side chains of Trp140, Ile141, Ala142, and Val144 are well defined, whereas the side chains of Val137 and Thr139 are disordered. The RMSD values are 0.50 ± 0.08 Å for the backbone atoms, and 0.75 ± 0.07 Å for all heavy atoms. Close packing contacts are observed on both sides of the indole ring of Trp140, with Val137 covering the indole ring on one side, and the backbone atoms of Gly143, Val144, and Gly145 covering the other side. The methyl groups of Ile141, Ala142, and Val144 are all clustered against each other. The upfield proton chemical shifts for residues 141, 142, 144, and 145 (Figures 3, 5, and 6a) seem to be a consequence of the packing against the indole ring of Trp140.

Outside of the regions I and II, the tripeptide segment of Tyr62, Tyr63, and Gly64 stands out because Gly64 has the smallest temperature coefficient and the largest amide proton chemical shift deviation from the random coil value of all residues in OmpX in 8 M urea (Figures 1 and 5b,c). The upfield shift and the small temperature coefficient of the amide proton of Gly64 probably arise from interactions with the aromatic ring of Tyr62, for which the $^1$H NMR lines could not be individually assigned. Aromatic ring–amide proton interactions of this type have been observed in a number of peptides (35–38) and denatured proteins (11), and are thought to be a general feature in peptides and proteins containing the sequence Tyr–Xxx–Gly (39). Here, this presumed local nonrandom structure could not be further characterized on the basis of NOE constraints.

DISCUSSION

The major issue left to be treated in this section is the extent to which the local structures of Figures 7 and 8 are populated in OmpX solutions in 8 M urea. A starting point for this discussion is that a single set of NMR lines is observed for the OmpX polypeptide chain, which would be
compatible either with the presence of only a single conformation, or with rapid interconversion, on the chemical shift time scale, of a manifold of different conformers (3).

For the calculations of the structures in Figures 7 and 8, we used exclusively NOE distance constraints as the input, and assumed that all observable long-range NOEs originate from the same OmpX molecules (see below). The observed NOE intensities are an average over the contributions from all rapidly interconverting random and nonrandom OmpX conformations (3). In fully randomized unfolded proteins, intense sequential $d_{SN}$ and $d_{NN}$ NOEs, as well as weaker medium-range $d_{SN}(i,i+2)$ and $d_{NN}(i,i+2)$ NOEs would therefore be expected (7, 32). Because of the $1/r_{ij}^6$ dependence of the NOE, where $r_{ij}$ is the distance between the two protons related by the NOE, $i$ and $j$, nonrandom conformers containing short distances between discrete pairs of hydrogen atoms can make a dominant contribution to the NOEs between these proton pairs even if they represent only a small fraction of the structural ensemble. With the assumption that the contributions from fully randomized molecules to longer-range NOEs, with $|i-j| \geq 3$, are negligibly small, the input for the structure calculations in Figures 7 and 8 included only these longer-range NOEs. Since for both nonrandom polypeptide segments (Figure 3), all observed upper-limit NOE distance constraints with $|i-j| \geq 3$ were simultaneously satisfied by a single, quite well-defined structure (Figures 7 and 8), we concluded that all other conformations present in 8 M urea did indeed not significantly contribute to the NOEs used as input for the structure calculations. This then also implied that no other nonrandom conformations with similar short $^1H-^1H$ distances are significantly populated in these solutions.

Two different experimental observations enabled us to obtain an estimate of the population of the nonrandom conformations. First, the contribution from a given molecular species to the observed average $^1H$ chemical shifts is weighted simply by its population in solution. The different averaging of NOE intensities and chemical shifts is clearly manifested if the shift deviations from the random coil values (Figures 5a,b and 6) are compared with the predictions from ring current calculations for the aliphatic protons in the structures of Figures 7 and 8. With the implicit assumption

**Figure 7:** NMR structure of the polypeptide segment I (Figure 3) in urea-unfolded OmpX calculated from the NOE upper distance constraints with $|i-j| \geq 3$. (a) Bundle of the 20 best energy-minimized conformers superimposed for minimal RMSD of the backbone atoms of residues 73–82, showing the polypeptide backbone. (b) Ball-and-stick presentation of the conformer with the lowest residual DYANA target function value in the same orientation as panel a. The experimental upper-limit NOE distance constraints used as input for the structure calculations are shown as yellow lines. (c) Stereoview of the all-heavy-atom presentation of the bundle in panel a, where the backbone is drawn with a thicker line than the side chains. The figures have been prepared with the program MOLMOL (62).
of 100% population of these structures, these calculations predict larger deviations from the random-coil shifts than those observed, leading to estimates for the populations of the local structures I and II of 30 and 25%, respectively.

Second, a statistical analysis of the relative intensities of short-range and longer-range NOE cross-peaks opened a second, independent avenue for estimating populations in the ensemble of conformers represented by the NMR spectrum. The simple fact that the $d(i,i+4)$ and $d(i,i+5)$ NOEs have readily detectable intensity shows that the population of the nonrandom structures must be not less than about 10% (3). A more precise value for this lower bound can be obtained by accounting for the relative intensities of all observed NOEs in the polypeptide segments with and without residual nonrandom structure. In this novel approach, it is assumed that the observed intensities of the $d(i,i+3)$, $d(i,i+4)$, and $d(i,i+5)$ NOEs arise entirely from the nonrandom conformation, whereas the observed intensity of the intraregular, sequential, and medium-range $d(i,i+2)$ NOEs arises from the entire population of protein molecules. Since the intensities of the intraregular, sequential, and medium-range $d(i,i+2)$ NOEs in the structured segments (Figures 7 and 8) can be calculated from the corresponding $^1\text{H}^-^1\text{H}$ distances in the NMR structures, their relative populations can be assessed from the observed signal intensities for these NOEs. Pairs of protons with scalar coupling were excluded from this analysis, to exclude possible errors that could arise from contributions to the NOE intensities from ZQ transitions (40, 41). The $d(i,i+3)$, $d(i,i+4)$, and $d(i,i+5)$ NOEs were used for the calibration of the NOE cross-peak intensity-$^1\text{H}^-^1\text{H}$ distance ratio. Using the statistical F-test, lower bounds for the populations of the nonrandom conformations I and II (Figures 3, 7, and 8) were thus obtained to be ≥30 and ≥20%, respectively, which is in close agreement with the corresponding data from the comparison of NOE and chemical shift averaging.

One of the referees expressed that he would like us to discuss our assumption of cooperative folding of the local nonrandom structures in some detail. For nonglobular polypeptide chains with lowly populated local nonrandom conformations, one cannot a priori distinguish between the following two limiting situations, or combinations thereof: (i) The individual NOE-observable long-range $^1\text{H}^-^1\text{H}$ contacts are statistically distributed among the entire population of polypeptide chains. (ii) There is cooperative folding of local nonrandom structure, and all the NOE-observable long-range $^1\text{H}^-^1\text{H}$ contacts are located in the same polypeptide chains. As mentioned at the outset of this discussion, we
arbitrarily based the structural interpretation of the NMR data (Figures 7 and 8) on the situation (ii). The results obtained support that this working hypothesis is reasonable for the treatment of OmpX in 8 M urea, since within the precision of the experimental measurements there were no inconsistencies between the different NMR data. In particular, the ensemble of all NOE distance measurements and of all chemical shift data, which follow different rules for ensemble averaging, are compatible with the unique bundles of conformers in the Figures 7 and 8. Additional support for the assumption of cooperative folding in the OmpX polypeptide chain in 8 M urea appears to come from considerations on the lifetime of the local nonrandom structures. The long-range NOE distance constraints used as input for the structure calculations derive from the observation of negative NOEs, i.e., from $^1$H→$^1$H NOEs related to effective correlation times longer than one nanosecond, which would then also be a lower limit on the lifetime of the local structures defined by the NOEs. It seems unlikely that such long-lived local structures would prevail in the absence of cooperative effects including a sizable number of simultaneous close interatomic contacts (Figures 7 and 8). In contrast, since the local nonrandom structures I and II (Figure 3) have been calculated independently, no statements can be made on whether there is a cooperative effect, with the two structures predominantly present either only in the same molecules or only in different molecules, or rather a statistical distribution of the two local structures among all OmpX molecules.

Comparison of the structure of denatured OmpX (Figures 3, 7, and 8) with the structure of the folded protein in DHPC micelles (18, 42, 43) and in crystals with n-octyldiployoxyethylene (C₈POE) (44) shows that the residual structures in 8 M urea do not coincide with the spatial arrangement of the corresponding peptide segments in the folded protein. The residues 73–77 in the structure I (Figure 7) correspond to the turn connecting the $\beta$-strands $\beta_4$ and $\beta_5$, and all other residues in the structures I and II are in $\beta$-strands in the folded protein (Figure 3). In the urea-unfolded OmpX, the nonrandom structure II appears to be stabilized by hydrophobic interactions in a small cluster of amino acid side chains formed around Trp 140 (Figures 7 and 8). These hydrophobic contacts would sterically not be possible in the folded OmpX, but hydrophobic clustering around Trp residues has been reported previously in studies of unfolded proteins and short peptides (8, 11, 45).

It will be interesting to further investigate possible roles of the local nonrandom structures seen in the urea-unfolded form of OmpX for OmpX folding and insertion into ordered lipid structures. Non-native hydrophobic clusters have previously been hypothesized to function as local “hot spots”, serving as anchoring points to the membrane prior to folding into the native structure and insertion into the lipid arrays (46–48). The presently described resonance assignments for the urea-unfolded form of OmpX, combined with the knowledge of the chemical shifts and the three-dimensional structure of folded OmpX in DHPC micelles (18, 42, 43), should provide a platform for designing and evaluating additional NMR experiments for studies of the folding pathway of OmpX and related membrane proteins.

ACKNOWLEDGMENT

We thank Prof. Gerhard Wider for useful discussions.

REFERENCES


