Initial Molecular Recognition Steps of McjA Precursor during Microcin J25 Lasso Peptide Maturation

Nadine Assrir, Anna Pavelkova, Régine Dazzoni, Rémi Ducasse, Nelly Morellet, Eric Guittet, Sylvie Rebuffat, Séverine Zirah, Yanyan Li, Ewen Lescop

Initial molecular recognition steps of McjA precursor during microcin J25 lasso peptide maturation

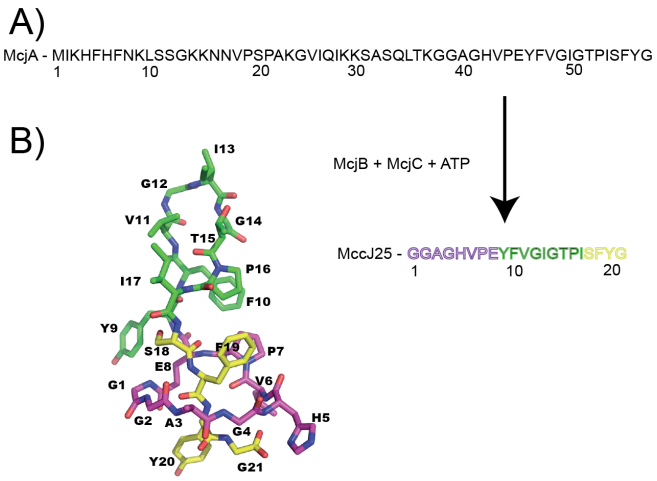
Nadine Assrir[a], Anna Pavelkova[a], Régine Dazzoni[a], Rémi Ducasse[b], Nelly Morellet[a], Eric Guittet[a], Sylvie Rebuffat[b], Séverine Zirah[b], Yanyan Li[b] and Ewen Lescop*[a]

Abstract: Microcin J25 (MccJ25) has emerged as an excellent model to understand the maturation of ribosomal precursor peptides into the entangled lasso fold. MccJ25 biosynthesis relies on the post-translational modification of the precursor McjA by the ATP-dependent protease McjB and the lactam synthetase McjC. Here, using NMR spectroscopy we show that McjA is an intrinsically disordered protein without detectable conformational preference, which emphasizes on the active role of the maturation machinery on the three-dimensional folding of MccJ25. We further show that the N-terminal region of the leader peptide is involved in interaction with both maturation enzymes and identify a predominant interaction of V43-S55 in the core McjA sequence with McjC. Moreover we demonstrate that residues K23-Q34 in the N-terminal McjA leader peptide tend to adopt a helical conformation in presence of membrane mimics, implying a role in directing McjA to the membrane in the vicinity of the lasso synthetase/export machinery. These data provide valuable insights into the initial molecular recognition steps in MccJ25 maturation process.

[a] Institut de Chimie des Substances Naturelles, CNRS UPR2301, Université Paris-Sud, Université Paris-Saclay, 1 avenue de la Terrasse, 91190 Gif-sur-Yvette, France. E-mail: ewen.lescop@cnrs.fr [b] Molécules de Communication et Adaptation des Microorganismes, UMR 7245 CNRS - Muséum National d'Histoire Naturelle, Sorbonne Universités, CP 54, 57 rue Cuvier, 75005 Paris, France

Introduction

Lasso peptides are 15 to 24-residue peptides produced by bacteria and characterized by an extraordinary rotaxane topology[1]. In this highly constrained structure, a macrolactam ring is established by an isopeptide linkage between the *N*-terminal residue of the peptide and the side chain of a glutamate or aspartate at positions 7, 8 or 9. The *C*-terminal tail of the peptide further threads through the *N*-terminal ring and is firmly trapped by steric interactions or/and disulfide bonds[1b, 2]. This unique compact structure confers lasso peptides with high stability against proteolysis or thermal and chemical denaturation. These peptides naturally act as receptor antagonists or enzyme inhibitors, the latter action being sometimes associated with antibacterial and/or antiviral properties. Due to both extraordinary stability and attractive biological activities, lasso peptides constitute a valuable scaffold for drug design. The proof of concept of their usage in drug design has been provided by Marahiel and coll.[3], with the development of integrin antagonists based on a lasso peptide scaffold. However, synthesis of the lasso topology by conventional chemical methods is still an unsolved challenge. It is therefore crucial to understand in details their biosynthesis by producing bacteria to permit engineering novel lasso peptides endowed with desired biological activities.



Scheme 1. (A) Scheme illustrating the maturation process of McjA into the lasso peptide MccJ25 accomplished by the enzymes McjB/McjC in presence of ATP. The primary structures of the precursor McjA and of MccJ25 are shown with their corresponding amino acid numbering. The ring, loop and *C*-terminal regions of MccJ25 are colored magenta, green and yellow, respectively. (B) NMR structure of MccJ25 (PDB Code 1Q71). The amino acids are labeled according to their position in MccJ25 sequence. MccJ25 structure is colored using the same convention as in panel A.

Microcin J25 (MccJ25) is a lasso peptide produced by *Escherichia coli* with potent antibacterial activity against *Enterobacteriaceae* such as certain *Escherichia*, *Salmonella* and *Shigella* species[2, 4]. This activity relies on the inhibition of bacterial RNA polymerase (RNAP) by obstructing its secondary channel[5] and on the production of reactive oxygen species in the respiratory chain[5b, 6]. MccJ25 consists of a 21-residue lasso peptide including a macrolactam ring closed by the G1 *N*-terminal amino group and the E8 side chain (Scheme 1). The rotaxane topology of the peptide is provided by two bulky amino acids of the *C*-terminal tail, F19 and Y20, located on each side of the ring, which lock the tail into the ring[7]. MccJ25 has emerged as an excellent model to understand the maturation of lasso peptides[8]. Its biosynthesis gene cluster encodes the linear precursor McjA, two maturation enzymes McjB and McjC and an ATP-binding cassette (ABC) transporter McjD[9]. McjA is composed of a 37-residue leader peptide at the *N*-terminus and the 21-residue structural core sequence of MccJ25 at the *C*-terminus. The enzymes McjB and McjC are responsible for the post-translational modification of McjA that results in the mature bioactive MccJ25 peptide. McjD is involved in the export of mature MccJ25 in the environment and simultaneously confers immunity to the producing bacteria. We have reconstituted *in vitro* the maturation of MccJ25 from its precursor McjA by the enzymes McjB and McjC in the presence of ATP[8a]. The role of each enzyme has also been delineated[8b]. McjB acts as an ATP-dependent cysteine protease responsible for the cleavage of the leader peptide in McjA (between residues K37 and G38). McjC was demonstrated to be a lactam synthetase and as such

catalyzes the formation of the macrolactam bond. McjB and McjC were proposed to form a complex that was termed lasso synthetase[8b]. The processing of McjA was observed to occur at the bacterial inner membrane[10], which suggests the requirement for McjA and the enzymes McjB/McjC to colocalize near McjD at the cell membrane for efficient maturation and export.

To date, the molecular mechanisms responsible for the acquisition of the three-dimensional fold of lasso peptides still remain poorly understood, in part due to the absence of structural information for the protein partners and the complexes thereof. In this study, we established by NMR spectroscopy that free McjA is largely disordered without detectable transient structural preference. To further explore the molecular interactions of McjA in the vicinity of the lasso synthetase machinery at the bacterial inner membrane, we studied in details the structural and dynamic properties of McjA in presence of SDS micelles or in presence of the maturation enzymes.

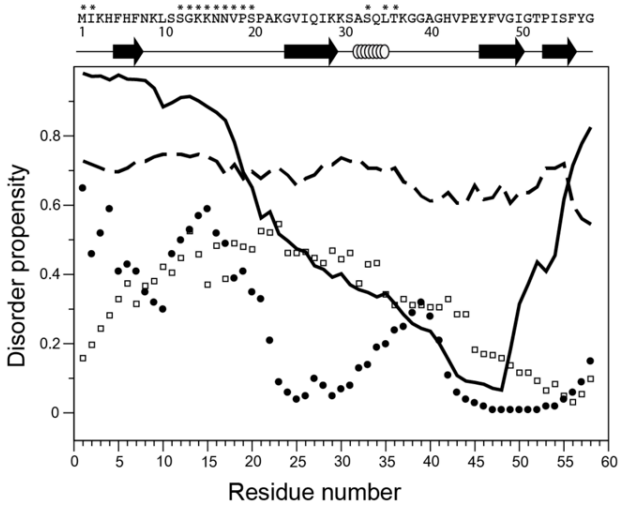


Figure 1: Bioinformatics analysis of the McjA amino acid sequence. The amino acid sequence of McjA is shown at the top of the figure together with the secondary structure elements as predicted by PSIPRED. The disorder propensity was calculated using IUPRED[11] (empty squares), DISPRO[12] (filled circles), META-disorder[13] (dashed line) and PONDR-FIT[14] (continuous line). Residues identified as disordered by the algorithm DISOPRED2[15] are indicated by a star on the protein sequence.

Results

Amino acid sequence analysis of the precursor McjA.

Several bioinformatics tools have been made available in the last years for the study of disordered regions in proteins (see ref[16] for examples). To predict the disordered regions in McjA, we used three predictors IUPRED[11], DISPRO[12], DISOPRED2[15] and the meta-predictors META-disorder (MD)[13] and PONDR-FIT[14] (Figure 1). The different algorithms gave different disorder predictions in the case of McjA: although all of them predicted McjA to be mostly disordered, several identified regions with reduced disorder. For example, according to DISPRO calculations, the regions K23-K30 and H42-I54 have a very low propensity to be disordered. PONDR-FIT predicted a very low disorder propensity in the region A40-I50, but high disorder propensity in the region K23-K30. IUPRED predicted low disorder propensity at the extremities of the protein but relatively high disorder propensity (> 0.4) between residues F5 to L35. In contrast, the meta-predictor META-disorder did not identify regions with low disorder propensity and mostly predicted McjA to be completely disordered. We also used PSIPRED[17] to predict the secondary structure elements of McjA (Figure 1). Several regions (F5-F7, G24-K29, F47-I51 and P53-F56) were shown to harbor β-strand propensity and a short α-helix was predicted in the region A32-L35. Therefore, despite some divergence in the predictions obtained from the different tools, the sequence analysis confirmed that McjA is likely essentially disordered in absence of partners. However bioinformatics tools could not provide a consistent picture of possible transiently formed secondary structures in McjA.

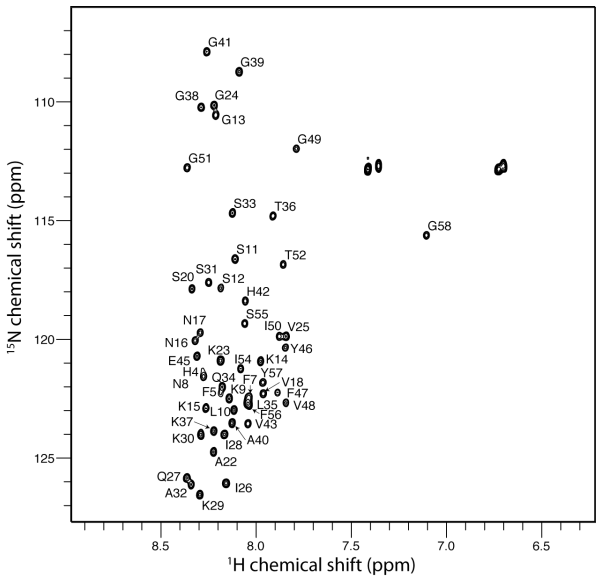


Figure 2: ^{15}N BEST-HSQC spectrum of 200 μM ^{15}N-labeled McjA at 600 MHz in 40 mM MES pH 6 and 282 K. The crosspeaks are labeled according to their amino acid type (one letter code) and numbering in McjA sequence.

The precursor McjA is an intrinsically disordered protein in solution.

The ^{15}N BEST-HSQC of McjA collected at 600 MHz in 40 mM MES buffer (pH 6.0) at 282 K (Figure 2) indicated that the ^{1}H chemical shift dispersion was limited, with most crosspeaks being in the 7.8-8.5 ppm ^{1}H chemical shift range. This is characteristic of a largely unfolded protein. To get deeper insights into the conformational state of McjA in aqueous solution we assigned the ^{1}H/^{15}N resonances using triple resonance experiments collected using a 200 μM ^{15}N/^{13}C-labeled McjA sample. All residues were assigned at the exception of the three N-terminal residues that were not visible on the spectra. The backbone chemical shifts were analyzed using Talos+[18] and δ2d[19] approaches (Figure 3A) to assess the flexibility and the propensity to form secondary structure elements (polyproline helix PPII, β-strand, α-helix, or random coil). The analysis by Talos+ revealed that RCI S^2 values along the McjA sequence were in the 0.3-0.6 range, which indicated the absence of any stable secondary structure. However, a slightly increased rigidity was apparent for some regions such as H4-F5, V18-P21, I26-K29 and A40-I54 that were characterized by larger than averaged S^2 values. We then performed δ2d

analysis from the same set of chemical shifts (Figure 3A). The regions with elevated S^2 values corresponded to regions where δ2d identified propensity to form PPII (H42-E45) or β-strand (G49-T52) or both (V18-P21 and P53-S55). However the measured populations were globally very low (below 13% at the exception of residues I50-G51 with propensities up to 20%). No significant helical propensity was detected for McjA in solution. The analysis of McjA chemical shifts demonstrated that McjA is predominantly disordered in solution and that, if any, secondary structures are not populated more than ~10-20%.

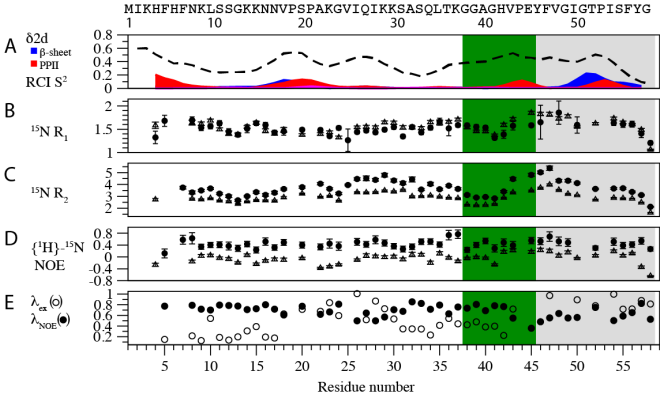


Figure 3: Probing McjA internal dynamics by NMR observables. (A) Chemical shift analysis based on Talos+ RCI algorithm[18] to probe the S^2 squared order parameter (dashed line) and on the δ2d algorithm[19] (continuous line) that estimates the propensity to form β-strand (blue) or polyproline II (red) elements. The populations of α-helix derived from δ2d analysis were very low (below 3%) and are not represented. ^{15}N R_1 (B) and R_2 (C) relaxation parameters and $\{^1H\}$-^{15}N heteronuclear NOE parameters (D) obtained at 600 MHz (triangles) and 950 MHz (circles) ^1H frequencies. (E) λ_{ex} (empty circles) and λ_{NOE} (filled circles) obtained from the HET-SOFAST experiment. All NMR experiments were carried out for 50 μM ^{15}N McjA in 40 mM MES at pH 6 and at 282 K. The green and grey backgrounds represent the McjA amino acid regions corresponding to the residues forming the ring (G1-E8) and the loop and C-terminus (Y9-G21) in MccJ25 sequence, respectively. Errorbars were calculated as described in Experimental Section.

To get more insight into McjA dynamics, we collected ^{15}N R_1 and R_2 relaxation rates and $\{^1H\}$-^{15}N heteronuclear NOE (hetNOE) values that report on H-N bond reorientation properties (Figure 3BCD). At the exception of the *C*-terminal extremity that exhibited shorter relaxation rates and smaller hetNOE values, as expected for highly mobile regions, the ^{15}N R_1 parameter and the hetNOE values were relatively flat over the protein sequence at the two magnetic fields. In particular the low hetNOE averaged values (0.44 at 950 MHz, -0.05 at 600 MHz) were typical of highly flexible molecules. The ^{15}N R_2 profile showed more pronounced variations. The regions encompassing the residues I26-A32 and V43-I50 had slightly larger R_2 values than the rest of the protein sequence, which might reflect slightly increasing rigidity in these regions, as already noticed from RCI S^2 analysis. In contrast, the glycine-rich region G38-G41 showed smaller R_2 values that might reveal increased flexibility. The λ_{ex} and λ_{NOE} parameters obtained from the HET-SOFAST experiment[20] report on the H_2O/H_N exchange properties and on the structural compactness and heterogeneity of polypeptide chains in solution, respectively. Low and high λ_{ex} values indicate rapidly and slowly exchanging amide protons, respectively. Two types of behaviors were observed (Figure 3E): the F5-N17 and K30-H42 regions globally showed rapidly exchanging amide protons (λ_{ex} < 0.6), whereas the V18-K29 and E45-G58 fragments globally had slowly exchanging amide protons. To check whether this behavior could be explained by the residue-type dependent solvent-exchange properties, we predicted the solvent exchange rates for the McjA sequence using the Sphere server[21] (data not shown). The Sphere-predicted exchange rates had a globally similar profile as compared to the observed λ_{ex} values, indicating that the measured λ_{ex} can be well explained by the amino acid-type dependent solvent exchange properties and more importantly, that amide protons are poorly protected from solvent-exchange all along the sequence, in agreement with a largely unfolded protein and limited local ordering. The λ_{NOE} parameter reports on ^1H-^1H dipolar interactions that depend on ^1H local spin density and dynamics. High λ_{NOE} values (> 0.75) were observed for most of the protein sequence, indicating highly mobile regions. Nevertheless, the I26-K29 and V43-F56 regions showed lower λ_{NOE} values (< 0.65) that are usually indicative of slightly more restricted mobility. Taken together this NMR study revealed that McjA is a largely intrinsically disordered protein with no significant formation of residual secondary structure elements, although two regions (I26-K29 and V43-F56) show consistent slightly increased rigidity compared to the rest of the protein.

Characterization of SDS micelles-bound McjA

Since the processing of McjA might occur near the bacterial membrane, we next envisaged that McjA might fold partially or completely upon membrane interaction. To support this, a previous circular dichroism (CD) study has shown that the helical contents of His_6-tagged McjA increased up to 10-15% upon interaction with the negatively-charged SDS micelles[8a] that are considered as good mimics of bacterial membranes[22]. However the localization of the newly formed helical region(s) could not be identified by this technique. Due to their small size, SDS micelles are compatible with solution-state NMR techniques. Therefore the SDS micelles-bound McjA was characterized by NMR spectroscopy using a ^{15}N McjA sample without N-terminal tag, to avoid potential artifacts. Successive additions of 2 mM SDS to a 50 μM ^{15}N-labeled McjA were performed to optimize SDS concentration. The complete disappearance of McjA signals in the ^{15}N SOFAST-HMQC experiment was observed in presence of 2 mM SDS (data not shown), while SDS at 4 mM induced the reappearance of McjA signals (Figure , top right panel). At 4mM SDS, about half resonances were strong and narrow, but the remaining signals were weak and broad, suggesting residual line-broadening. We therefore increased SDS concentration up to 8 mM which led to high quality spectra (Figure , lower panel). Under these conditions, SDS formed micelles and under the assumption of 80 SDS molecules per micelle, about 0.5 McjA molecule was interacting with one micelle. The NMR spectrum of McjA dramatically changed upon SDS interaction suggesting a major conformational and dynamic change in the SDS-bound state. The NMR resonances were assigned using a set of triple resonance experiments (Figure , lower panel) and the resulting chemical shifts were analyzed using the δ2d algorithm (Figure 5A). A significant propensity (40-60%) to form an α-helical conformation was observed in the K23-Q34 region and to a lower propensity (~15%) in the small segment around E45-Y46. Very similar results were obtained from Talos+ analysis (see Figure 5A). Taken together, the δ2d algorithm identified an

average value of 13% for helical propensity over the protein sequence, which nicely correlated with the previous CD measurements carried out at the same SDS concentration[8a].

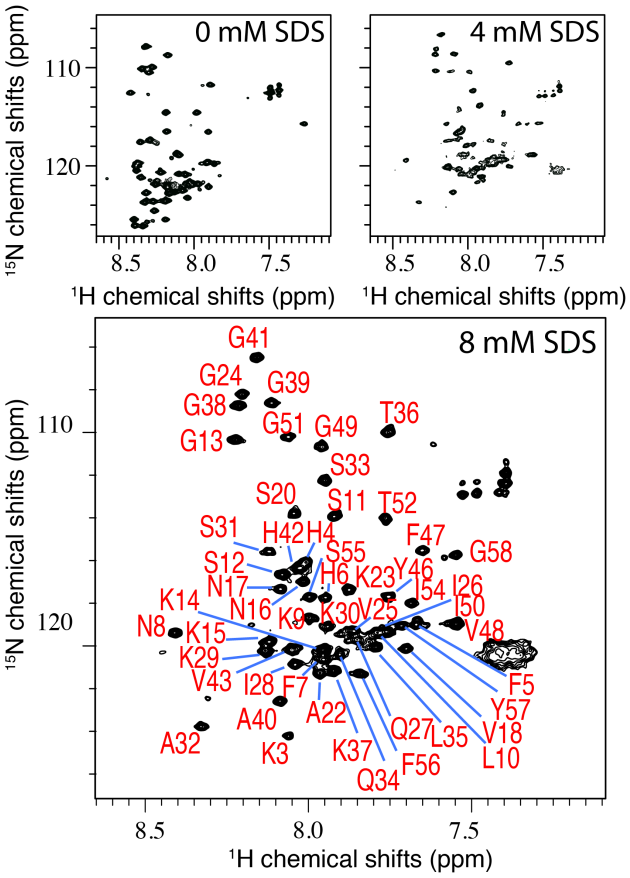


Figure 4: ^{15}N SOFAST-HMQC spectra collected on a sample containing 100 μM ^{15}N-^{13}C labeled McjA in 40 mM MES pH 6 at 298 K and 800 MHz without SDS and in presence of 4 and 8 mM SDS concentrations. The assignment of the SDS-bound form of McjA is shown at 8 mM SDS concentration.

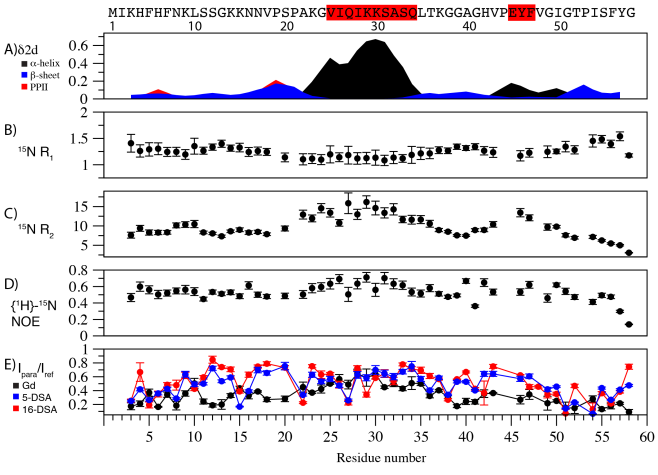


Figure 5: Interaction between McjA and SDS. The McjA amino acid sequence is shown on top. Residues found by Talos+ analysis to adopt a helical conformation are coloured on a red background. The secondary structure propensity as obtained from δ2d-based analysis of chemical shifts is shown in panel A. The R_1 and R_2 ^{15}N relaxation and {^{1}H}-^{15}N heteronuclear NOE parameters are shown in panel B, C and D respectively. Panel E represents the paramagnetic effects as judged from the intensity change (I_{para}/I_{ref}) upon addition of 5 mM Gd(DPTA-BMA) (black circles) used as a paramagnetic agent to probe the solvent accessibility. 600 μM of 5-DSA (blue circles) and 16-DSA (red circles) were used to probe the micelle surface and core, respectively. All NMR experiments were carried out for 100 μM ^{15}N McjA and 8 mM SDS in 40 mM MES at pH 6, T = 298 K and 800 MHz ^{1}H frequency. Errorbars were calculated as described in Experimental Section.

We then measured the ^{15}N R_1 and R_2 relaxation and {^{1}H}-^{15}N heteronuclear NOE parameters to probe McjA dynamics (Figure 5BCD). The relaxation parameters were homogeneous over the protein sequence at the exception of the two regions showing partial folding as α-helices and of the C-terminal extremity. The fragments identified to probe helical conformations were characterized by lower R_1 values and elevated R_2 and {^{1}H}-^{15}N heteronuclear NOE values, in agreement with more restricted motions. In contrast, the *C*-terminal extremity showed elevated R_1 values and lower R_2 and {^{1}H}-^{15}N heteronuclear NOE that are collectively indicative of significant mobility at the ps-ns timescale.

We next questioned the relative positioning of McjA with respect to the SDS micelles using paramagnetic probes. The water soluble neutral Gd(DPTA-BMA)[23] and the hydrophobic stearic acid probes were used to identify the solvent-accessibility and membrane-binding regions, respectively. In the latter case, we used modified stearic acids in which the paramagnetic doxyl groups were inserted near the headgroup at position 5 (5-DSA) or at the extremity at position 16 (16-DSA), respectively[24]. When adding 5 mM Gd(DPTA-BMA) to the McjA solution, all amino acid residues showed reduced intensity in the ^{15}N HSQC spectrum (Figure 5E, black squares). The less affected residues were located in the region K23-Q34 that corresponded to the main helix in McjA, thus suggesting that this helix is less accessible to the solvent compared to other McjA regions. In contrast, the residues mostly affected by Gd(DPTA-BMA) were located at the *N*-terminal (residues K3, H4, H6, N8, S11-G13) and at the *C*-terminal (G51, T52, S55-F56 and G58) extremities, which suggests the high solvent accessibility of these residues. The 5-DSA and 16-DSA paramagnetic agents are expected to induce line-broadening, and therefore reduction in NMR signal intensity, for residues near the surface or near the center of the micelle, respectively[25]. The addition of 600 μM 5-DSA (Figure 5E, blue squares) also led to significant intensity changes. The less affected amino acids were K9-K14, N17-S20, K23-I26, I28-L35, K37, G39-A40 and H42-V48, suggesting that they are not in direct contact with the SDS micelles, which correlates well with Gd(DPTA-BMA) data. In contrast, the residues K3, F5, N8, K15, A22, Q27, G38, G51-I54 were affected upon addition of 5-DSA, revealing their close proximity to the micelle. Surprisingly the addition of 16-DSA, which is supposed to probe a different depth within the micelle compared to 5-DSA, gave very similar intensity changes as compared to 5-DSA (Figure 5E). Such an observation is not unprecedented[25b, 26] and may be explained by an increase of the flexibility of the spin-label in the micelles[26] or by specific interaction between the peptide and the spin-label[25b]. 5-DSA and 16-DSA therefore only revealed the tendency of residues to interact with the fatty interior of the micelles.

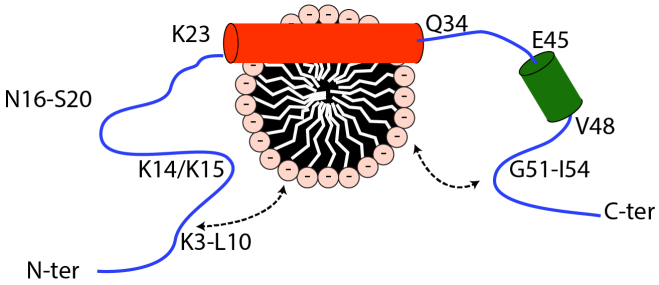


Figure 6: Model of the interaction between McjA and SDS micelles. The interior of the SDS micelle is shown in black and the SDS surface-exposed negative charges are indicated. The helical fragments identified from ^{13}C chemical shift analysis are shown as tubes. The helical central region K23-Q34 would be able to interact with the negatively charged surface of the SDS micelle via its three positively charged residues K23, K29 and K30 that are located on one face of the helix and the hydrophobic residues A22-V25-I26-I28-A32 would be oriented towards the interior of the micelle. The arrows illustrate that the regions K3-L10 and G51-I54 would transiently interact either with the bulk water or the micelles.

The Gd(DPTA-BMA) and doxylstearate paramagnetic relaxation agents helped revealing the global positioning of McjA with respect to the SDS micelles and we propose a model of the McjA/SDS micelles interaction (Figure 6). The N-terminal extremity (from K3 to L10) and the C-terminal G51-I54 stretch were significantly affected by both types of paramagnetic agents, which suggests that they are exposed to both bulk solvent and micellar environments (arrows in Figure 6). This can be explained by their mixed charged (K3, H4, H6, K9) and hydrophobic (F5, F7, P53, I54) property. The regions S12-K14, N16-S20 and G39-V48 are highly solvent-exposed, as expected from their amino acid compositions. K15 interacts more with the micellar environment, possibly with the negatively charged SDS headgroups. Residue A22 strongly interacts with the membrane. The K23-Q34 region that shows helical propensity is protected from both the solvent and the core of the micelle. The helix-micelles interactions are therefore stabilized by the hydrophobic surface formed by A22-V25-I26-I28-A32 that is more deeply embedded in the micelle, whereas the positively charged lysines K23-K29-K30 may create salt bridges with the surface exposed SDS negative charges.

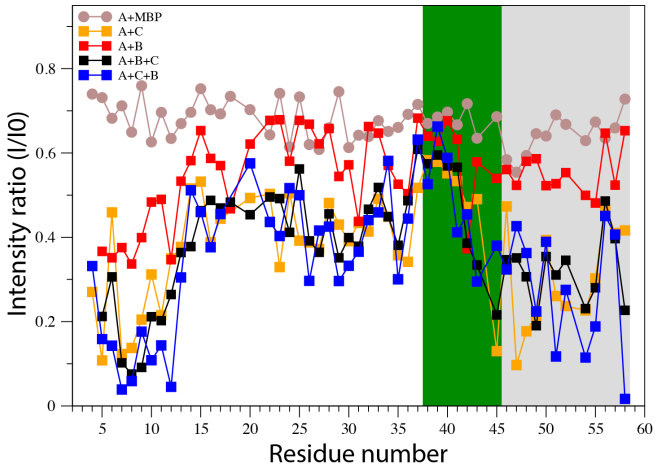


Figure 5: Interactions of McjA with McjB and McjC enzymes. ^{15}N SOFAST-HMQC spectra were collected for a 25 μM ^{15}N-labeled McjA solution in absence of in presence of MBP, McjB and McjC in 40 mM MES pH 6, 0.02% DOC, 200 mM NaCl. The figure shows the intensity drop (I/I_0) for each McjA residue upon addition of either MBP (brown circles) or of MBP-McjB (red, "A+B") followed by the addition of McjC (black, "A+B+C") and for the reverse protocol: addition of McjC (orange, "A+C") followed by the addition of McjB (blue, "A+C+B"). The green and grey backgrounds represent the McjA amino acid regions corresponding to the residues forming the ring (G1-E8) and the loop and *C*-terminus (Y9-G21) in MccJ25 sequence, respectively.

Interaction of McjA with McjB and McjC

To explore the molecular interaction of McjA with maturation enzymes, N-terminal MBP-fused McjB and McjC were used because the His-tagged proteins could not be purified to homogeneity. Attempts to isolate McjA/McjB or McjA/McjC complexes using gel filtration were unsuccessful, suggesting that the peptide-protein interaction is weak. In this regard, NMR spectroscopy is efficient to probe weak molecular interactions. When 25 μM ^{15}N-labeled McjA was mixed with stoichiometric McjB or McjC, ^{15}N SOFAST-HMQC signals of McjA disappeared completely, in agreement with significant intermolecular interactions between McjA and each of the two enzymes. In an attempt to identify the regions of McjA susceptible to show more specific contacts with the enzymes, ^{15}N labeled McjA (25 μM) was mixed with substoichiometric concentrations (15 μM) of McjB or of McjC and the residue-specific peak intensity changes in ^{15}N SOFAST-HMQC spectra were monitored upon addition of the proteins (Figure 5). MBP was also included as a control in the set of experiments to assess the possible effect of the MBP tag on the interaction. In presence of 15 μM MBP, a ~30% global decrease in NMR signal intensity was observed for McjA which exceeded by far the dilution factor (1%) (Figure 5). This indicated that McjA tended to interact with MBP despite the high ionic strength (200 mM NaCl) and the presence of a mild detergent (0.02% sodium deoxycholate (DOC)). The intensity drop was homogeneous along the protein sequence, which suggested non-specific interaction between McjA and MBP. Upon addition of McjB, or McjC, or of both, we observed a stronger decrease in intensity for McjA compared to the control MBP-addition experiment (Figure 5), in particular at the *N*-terminal extremity, demonstrating that McjB and McjC both interact with McjA within the MBP-fused context. In contrast to the control experiment, the intensity drop for McjA signals was highly sequence-dependent for the interaction with McjB ("A+B" experiment) or with McjC ("A+C" experiment). When McjB or McjC was added to McjA, the *N*-terminal extremity (residues F5 to S12) underwent a stronger decrease in intensity than the rest of the McjA sequence, which suggests contacts between this region and McjB or McjC. Regarding the rest of the McjA sequence (residues 13 to 58), a rather homogeneous profile of intensity change was observed upon McjB addition, whereas the addition of McjC led to large sequence-dependent changes. More precisely the region V43-S55 showed the most prevalent intensity drop (down to ~0.25), in agreement with stronger contacts between this region of McjA and McjC. In contrast, the region K37-A40 that contains the K37/G38 cleavage site, exhibited the minimal intensity change (~0.6), which reveals the weakest contacts between this region and McjC. Finally, the central region K14-T36 had intermediate ratios. We next compared the spectra obtained upon the addition of the two proteins onto McjA in different orders. The sequence of addition of the two proteins did not affect the final spectrum, as indicated by the very similar intensity ratio profiles for the "A+B+C" and "A+C+B" experiments. These profiles were very also similar to

that obtained with the addition of McjC only ("A+C" experiment), suggesting that McjC dominates the molecular interaction of McjA in the McjA/McjB/McjC mixture. Of note, the MBP-McjB form used here was active. Although MBP-McjC was not active, in contrast to His-McjC which was active[8a] but could not be purified to homogeneity for structural studies, our NMR data clearly evidence that McjC has kept its ability to bind McjA in the MBP fused context, thus providing a relevant insight into the enzyme mechanisms involved in the lasso formation. Presumably, MBP would perturb the McjB/McjC interactions due to steric clashes, thus precluding the total reaction to occur. Taken together, these binding experiments demonstrate that under the current conditions, McjA preferentially binds McjC rather than McjB. This interaction relies largely on the residues F5-S12 in the leader sequence and V43-S55 in the core sequence of McjA, whereas the residues K37-A40 have the minimal contacts with McjB or McjC.

Discussions

MccJ25 maturation is a complex process and involves multiple peptide-protein and protein-protein interactions. Some intriguing questions remain about the role of the leader peptide, the order of the reactions as well as how the lasso topology is acquired. Different roles have been assigned to the leader peptide in various ribosomal peptide systems. For example, it has been proposed[27] to act as a secretion signal, as a chaperone to assist in the folding of the precursor peptide, as a recognition motif for the enzymes during post-translational modifications, to protect the peptide from degradation, or to keep it inactive during processing, but also to promote efficiency and ensure order and fidelity of the post-translational modification reactions[28]. In the case of the MccJ25 lasso peptide maturation, we show here that the leader sequence of the precursor McjA does not induce the prefolding of the precursor peptide. This emphasizes on the crucial role of the maturation machinery on the acquisition of the lassoed shape from the precursor. The study of the interactions established between the precursor and the enzymes carried out here indicates that in the leader peptide, only the *N*-terminal extremity (F5-S12) is prone to interact with the enzymes McjC and McjB. This is consistent with a recent report that McjB-like proteins harbor leader recognition elements[29]. It is therefore tempting to predict that the region F5-S12 may anchor the precursor on the maturation enzymes. To this regard, an *in vivo* study[30] has shown that removal of the first 15 *N*-terminal amino acids of McjA only resulted in a 5-fold decrease in MccJ25 production suggesting that the McjA/enzyme interaction mediated by the leader peptide is dispensable for maturation/export of the lasso peptide. Therefore, the leader peptide / enzyme interaction is not crucial for the reaction but may contribute to accelerate MccJ25 maturation. The most striking structural feature of the leader peptide observed in our study is the ability of the central region K23-Q34, which is located in close vicinity to the K37-G38 McjB cleavage site, to fold into an α-helix in contact with SDS micelles used here as a surrogate of bacterial membranes. This sequence has a limited propensity to fold in a helix as a result of the bioinformatics and solution-state analysis, thus highlighting the role of the membrane mimics to drive peptide helical folding. McjA maturation has been proposed to occur near the bacterial membranes[10] in agreement with the idea that the mature lasso peptide must be rapidly exported after production via the ABC transporter McjD to protect the host bacteria. Therefore, the leader peptide of McjA may function as directing the precursor to the bacterial inner membrane to increase its local concentration near the maturation/export machinery. Our hypothesis is further supported by an *in vivo* study[30] that demonstrated that the region Q27-S31 of McjA, which forms the center of the helix observed in presence of SDS micelles, is absolutely required for MccJ25 production in the culture medium. One cannot exclude, however, that the helical region formed in the leader peptide in contact to an hydrophobic environment might be also involved in additional protein-protein interactions during maturation, as already suggested for microcin B17[31]. To this regard, we did not observe any significant interaction between residues Q27-S31 and the enzymes under our conditions, although such contacts might exist in later steps.

Lasso peptide maturation involves the McjB-mediated proteolytic cleavage of McjA between residues K37 and G38 as well as the McjC-mediated macrolactam ring formation between residues G38 and E45 (G1 and E8 in MccJ25 sequence numbering). The peptide however must adopt the correct conformation (pre-folding) as a β-hairpin structure prior to the closure reaction[7] to avoid the formation of branched-cycling peptide. Interestingly our NMR study demonstrates that the region V43-S55 in McjA interacts with McjC and to a much lesser extent to McjB. This suggests that McjC binding to McjA acts as the initial peptide-protein recognition step. To this regard, the region V43-S55 that largely defines the β-hairpin conformation in the final lassoed conformation might already undergo a conformational change upon McjC binding to facilitate the progress along the reaction coordinates. Relevant to MccJ25 biosynthesis, ATP-dependent activation of the carboxyl side-chain of E45 by McjC likely constitutes the first event of the maturation process. This initial interaction and E45 adenylation may subsequently drive the complete pre-folding of McjA. Unexpectedly, minimal interaction was observed between the maturation enzymes and the K37-A40 region of McjA that encompasses the proteolytic cleavage site. We then speculate that these residues must remain solvent-exposed in the McjA/McjC complex for optimal accessibility to McjB for the cleavage reaction.

Conclusions

Our work addresses the molecular recognition events that occur during the initial steps of McjA maturation into MccJ25. We propose that the central part of the leader peptide adopts an α-helical structure which acts as a membrane anchoring region to localize McjA at the membrane and that McjC initially recognizes McjA through its *C*-terminus. However, additional work is needed for a deeper characterization of the maturation pathway and in particular, through the availability of the long-awaited high-resolution structures of the enzymes and complexes thereof.

Experimental Section

Cloning, expression and purification of ^{15}N-labelled McjA

E. coli BL21 Star (DE3) cells (Thermo Fisher) were transformed with pET31-mcjA encoding McjA fused with a ketosteroid isomerase (KSI) tag

(KSI-McjA). This recombinant strain was grown at 37°C until OD600 reached 0.8-1 in minimal medium (M9) containing $^{15}NH_4Cl$ and, when needed, ^{13}C-enriched glucose as sole nitrogen and carbon sources respectively. Induction of protein expression was performed by isopropyl-β-D-thiogalactopyranoside (IPTG) at 1 mM (37°C, 4 hours). Cell pellets were resuspended in 20 mL resuspension buffer (20 mM Tris-HCl, pH 8, 100 mM NaCl) supplemented with 5 µg/mL of DNAase I and lysed through a cell disruptor (Constant Systems LTD). After centrifugation at 22 000 g for 30 min at 4°C, the pellet containing insoluble KSI-McjA was successively washed with washing buffer (2 × 10 mL : 1 mg/mL Na-deoxycholate, 1 mM EDTA, 20 mM β-mercaptoethanol) and resuspension buffer (2 × 10mL). The inclusion bodies were solubilized in 10 mL buffer containing 20 mM Tris-HCl (pH 8) and 6 M guanidine hydrochloride. Upon dilution by adding 40 mL deionized water, KSI-McjA was precipitated out and collected by centrifugation at 22000 g for 15 min at 4°C. To cleave McjA from the KSI tag, 3.5 mL of 70% trifluoroacetic acid and 0.5 mL of 5 M CNBr in acetonitrile were added to the precipitate, and the reaction was carried out at room temperature in the dark for 16 hours. After removal of CNBr by SpeedVac, the peptide McjA was resuspended in 40% acetonitrile and further purified by RP-HPLC on a CapCell Pak C18 column (4.6 × 150 mm; 120 Å, 5µm- Shiseido).

Cloning, expression and purification of MBP-McjB / MBP-McjC

The genes *mcjB* and *mcjC* were cloned into a modified plasmid pMAL-c5e to be fused at the *C*-terminus of an *N*-terminal His6 tagged maltose binding protein (MBP), generating plasmids pMAL-mcjB and pMAL-mcjC. *E. coli* BL21(DE3) cells transformed with these plasmids were grown in LB medium at 37°C until OD_{600} reached 0.7-1. Protein expression was induced by addition of 0.1 mM IPTG and cells continued to grow at 20°C for 20 h. Cells were resuspended in buffer A (50 mM Tris-HCl, 300 mM NaCl, 2 mM tris(2-carboxyethyl)phosphine (TCEP), pH 8) supplemented with protease inhibitors cocktail (Roche). Lysis was performed with a cell disruptor. The cleared supernatant was loaded on a Histrap FF crude column (GE Healthcare life sciences). MBP-McjB and MBP-McjC were eluted with 250 mM imidazole and further purified on a Superdex 200 10/300 GL (GE Healthcare life sciences) equilibrated with NMR buffer (40 mM MES, pH 6).

NMR Spectroscopy

NMR samples of 50 µM ^{15}N and 200 µM ^{15}N-^{13}C labeled McjA were prepared in 40 mM MES buffer (pH 6) in 90% H_2O / 10% D_2O. NMR experiments were collected on a Bruker AVANCE I 600 MHz or a Bruker AVANCE III 800 MHz or 950 MHz equipped with TCI cryoprobes, unless specified. ^1H-^{15}N correlation experiments were collected using the BEST-HSQC or SOFAST-HMQC[32] pulse sequences. The backbone resonance assignment of McjA was obtained using the BEST-version of 3D HNCO, HN(CO)CA, HNCA, HN(CO)CACB, HNCACB and HN(CA)N triple resonance experiments[33]. ^{15}N R_1 and R_2 relaxation rates were collected in an interleaved manner as a pseudo-3D experiment with an interscan delay of 6 s. For optimal peak dispersion in the ^{15}N dimension, the spectral resolution was increased using optimized ^{15}N spectral width obtained by the ASCOM method[34]. 16 and 10 relaxation delays were typically used for the ^{15}N R_1 and R_2 measurements respectively. The intensities extracted for each residue using NMRPipe[36] tools were then fitted to a two-parameter exponential function and errorbars were estimated from Monte-Carlo analysis. $\{^1H\}$-^{15}N heteronuclear NOE values were obtained from the ratio in intensities extracted from ^1H-^{15}N correlation maps with and without a 5 s ^1H_N presaturation period achieved by a train of 120° pulses. Errorbars were calculated from peak intensities and the estimated noise level from the two spectra using NMRPipe[36] tools. For the HET-SOFAST experiment, three ^{15}N SOFAST-HMQC spectra were collected without ^1H RF irradiation as a reference and under water or aliphatic proton selective saturation. The aliphatic-proton selective saturation was achieved by a 1.5 ms 180° REBURP shaped pulse[35] centered at 1.5 ppm. The water selective saturation was achieved using a 10 ms REBURP shaped pulse centered at 4.7 ppm. Saturation was obtained by the rapid scan repetition (d_1 = 0.1 s). The λ_{ex} and λ_{NOE} parameters were extracted respectively from the ratios of intensities I_{ex}/I_{ref} and I_{NOE}/I_{ref}, where I_{ex}, I_{NOE} and I_{ref} are the cross-peak intensities observed in the spectra under water-selective saturation (I_{ex}) or aliphatic proton-selective saturation (I_{NOE}) conditions and in absence of irradiation (I_{ref}). To limit radiation damping effects, the HET-SOFAST experiment were collected with a room-temperature probe. Data processing and analysis were performed using NMRPipe[36] and CCPNMR[37] softwares.

Acknowledgements

This work was financed by the ANR grant (no. BLAN_NT09_692063). We thank Philippe Durand (ICSN) for his help on the CNBr-based KSI cleavage reaction. Financial support from the TGIR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged.

Keywords: lasso peptide, NMR, intrinsically disordered protein, SDS micelles, biosynthesis

[1] a) J. D. Hegemann, M. Zimmermann, X. Xie, M. A. Marahiel, *Acc. Chem. Res.* **2015**, *48*, 1909-1919; b) M. O. Maksimov, S. J. Pan, A. J. Link, *Nat. Prod. Rep.* **2012**, *29*, 996-1006; c) Y. Li, S. Zirah, S. Rebuffat, *Lasso peptides: bacterial strategies to make and maintain bioactive entangled scaffolds*, Springer New York, **2015**.

[2] S. Rebuffat, A. Blond, D. Destoumieux-Garzón, C. Goulard, J. Peduzzi, *Curr. Protein Pept. Sci.* **2004**, *5*, 383-391.

[3] a) T. A. Knappe, F. Manzenrieder, C. Mas-Moruno, U. Linne, F. Sasse, H. Kessler, X. Xie, M. A. Marahiel, *Angew. Chem. Int. Ed. Engl* **2011**, *50*, 8714-8717; b) J. D. Hegemann, M. De Simone, M. Zimmermann, T. A. Knappe, X. Xie, F. S. Di Leva, L. Marinelli, E. Novellino, S. Zahler, H. Kessler, M. A. Marahiel, *J. Med. Chem.* **2014**, *57*, 5829-5834.

[4] a) R. A. Salomon, R. N. Farias, *J. Bacteriol.* **1992**, *174*, 7428-7435; b) P. A. Vincent, R. D. Morero, *Curr. Med. Chem.* **2009**, *16*, 538-549.

[5] a) K. Adelman, J. Yuzenkova, A. La Porta, N. Zenkin, J. Lee, J. T. Lis, S. Borukhov, M. D. Wang, K. Severinov, *Mol. Cell.* **2004**, *14*, 753-762; b) J. Mukhopadhyay, E. Sineva, J. Knight, R. M. Levy, R. H. Ebright, *Mol. Cell.* **2004**, *14*, 739-751.

[6] A. Bellomio, P. A. Vincent, B. F. de Arcuri, R. N. Farias, R. D. Morero, *J. Bacteriol.* **2007**, *189*, 4180-4186.

[7] a) M. J. Bayro, J. Mukhopadhyay, G. V. Swapna, J. Y. Huang, L. C. Ma, E. Sineva, P. E. Dawson, G. T. Montelione, R. H. Ebright, *J. Am. Chem. Soc.* **2003**, *125*, 12382-12383; b) K. J. Rosengren, R. J. Clark, N. L. Daly, U. Goransson, A. Jones, D. J. Craik, *J. Am. Chem. Soc.* **2003**, *125*, 12464-12474; c) K. A. Wilson, M. Kalkum, J. Ottesen, J. Yuzenkova, B. T. Chait, R. Landick, T. Muir, K. Severinov, S. A. Darst, *J. Am. Chem. Soc.* **2003**, *125*, 12475-12483.

[8] a) S. Duquesne, D. Destoumieux-Garzón, S. Zirah, C. Goulard, J. Peduzzi, S. Rebuffat, *Chem. Biol.* **2007**, *14*, 793-803; b) K. P. Yan, Y. Li, S. Zirah, C. Goulard, T. A. Knappe, M. A. Marahiel, S. Rebuffat, *Chembiochem* **2012**, *13*, 1046-1052.

[9] J. O. Solbiati, M. Ciaccio, R. N. Farias, J. E. Gonzalez-Pastor, F. Moreno, R. A. Salomon, *J. Bacteriol.* **1999**, *181*, 2659-2662.

[10] D. J. Clarke, D. J. Campopiano, *Org. Biomol. Chem.* **2007**, *5*, 2564-2566.

[11] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, *Bioinformatics* **2005**, *21*, 3433-3434.

[12] J. Cheng, M. J. Sweredoski, P. Baldi, *Data Mining Knowl. Disc* **2005**, *11*, 213–222.

[13] A. Schlessinger, M. Punta, G. Yachdav, L. Kajan, B. Rost, *PLoS One* **2009**, *4*, e4433.

[14] B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, V. N. Uversky, *Biochim. Biophys. Acta* **2010**, *1804*, 996-1010.

[15] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, D. T. Jones, *J. Mol. Biol.* **2004**, *337*, 635-645.

[16] a) B. He, K. Wang, Y. Liu, B. Xue, V. N. Uversky, A. K. Dunker, *Cell. Res.* **2009**, *19*, 929-949; b) Z. Dosztanyi, P. Tompa, *Methods Mol. Biol.* **2008**, *426*, 103-115.

[17] D. T. Jones, *J. Mol. Biol.* **1999**, *292*, 195-202.

[18] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213-223.

[19] C. Camilloni, A. De Simone, W. F. Vranken, M. Vendruscolo, *Biochemistry* **2012**, *51*, 2224-2231.

[20] P. Schanda, V. Forge, B. Brutscher, *Magn. Reson. Chem.* **2006**, *44*, S177-184.

[21] a) Y. Bai, J. S. Milne, L. Mayne, S. W. Englander, *Proteins* **1993**, *17*, 75-86; b) G. P. Connelly, Y. Bai, M. F. Jeng, S. W. Englander, *Proteins* **1993**, *17*, 87-92; c) Y.-Z. Zhang, University of Pennsylvania, PA, USA. **1995**.

[22] a) E. Strandberg, A. S. Ulrich, *Concepts Magn. Reson.* **2004**, *23A*, 89–120; b) S. Bourbigot, E. Dodd, C. Horwood, N. Cumby, L. Fardy, W. H. Welch, Z. Ramjan, S. Sharma, A. J. Waring, M. R. Yeaman, V. Booth, *Biopolymers* **2009**, *91*, 1-13.

[23] G. Pintacuda, G. Otting, *J. Am. Chem. Soc.* **2002**, *124*, 372-373.

[24] a) L. R. Brown, C. Bosch, K. Wüthrich, *Biochim. Biophys. Acta* **1981**, *642*, 296-312; b) P. Damberg, J. Jarvet, A. Gräslund, *Methods Enzymol.* **2001**, *339*, 271-285.

[25] a) C. H. Papavoine, R. N. Konings, C. W. Hilbers, F. J. van de Ven, *Biochemistry* **1994**, *33*, 12990-12997; b) J. Jarvet, J. Zdunek, P. Damberg, A. Gräslund, *Biochemistry* **1997**, *36*, 8153-8163.

[26] L. A. Sommer, J. J. Janke, W. F. Bennett, J. Burck, A. S. Ulrich, D. P. Tieleman, S. A. Dames, *J Phys Chem B* **2014**, *118*, 4817-4831.

[27] T. J. Oman, W. A. van der Donk, *Nat. Chem. Biol.* **2010**, *6*, 9-18.

[28] C. J. Thibodeaux, J. Wagoner, Y. Yu, W. A. van der Donk, *J. Am. Chem. Soc.* **2016**, *138*, 6436-6444.

[29] B. J. Burkhart, G. A. Hudson, K. L. Dunbar, D. A. Mitchell, *Nat. Chem. Biol.* **2015**, *11*, 564-570.

[30] W. L. Cheung, S. J. Pan, A. J. Link, *J. Am. Chem. Soc.* **2010**, *132*, 2514-2515.

[31] R. S. Roy, S. Kim, J. D. Baleja, C. T. Walsh, *Chem. Biol.* **1998**, *5*, 217-228.

[32] P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **2005**, *127*, 8014-8015.

[33] a) E. Lescop, P. Schanda, B. Brutscher, *J. Magn. Reson.* **2007**, *187*, 163-169; b) P. Schanda, H. Van Melckebeke, B. Brutscher, *J. Am. Chem. Soc.* **2006**, *128*, 9042-9043.

[34] E. Lescop, P. Schanda, R. Rasia, B. Brutscher, *J. Am. Chem. Soc.* **2007**, *129*, 2756-2757.

[35] H. Geen, R. Freeman, *J. Magn. Reson.* **1991**, *93*, 93-141.

[36] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277-293.

[37] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins* **2005**, *59*, 687-696.