RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

RSC Advances

RSC Publishing

ARTICLE

High-resolution NMR structure of a Zn^{2+}-containing form of the bacteriophage T5 L-alanyl-D-glutamate peptidase

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Dmitry A. Prokhorov,[a] Galina V. Mikoulinskaia,[b] Nikolai V. Molochkov,[a] Vladimir N. Uversky,[c,d,e,f,*] and Victor P. Kutyshenko[a,*]

This paper represents the spatial solution structure of the Zn^{2+} -containing form of the bacteriophage T5 L-alanyl-D-glutamate peptidase (EndoT5-Zn^{2+}). The core of this α+β protein is formed by three α-helices (residues 7-15, 20-30, and 87-104) and a β-sheet containing three β-strands (residues 35-39, 71-76, and 133-135). The protein has two short loops (residues 16-19 and 31-34), a medium-length loop (residues 77-86) containing a short β-hairpin (residues 77- 82), and two long loops (residues 40-70 and 105-132). The long loops include a stable 3_{10}-helix (residues 66-68) and labile α-helices 46-53 and 113-117. Catalytic Zn^{2+}-binding site is represented by three amino acid residues, His_{66}, Asp_{73}, and His_{133}. The cation-binding His residues are located near the foundations of the long loops, whereas Asp_{73} is positioned in the middle of the core β-sheet. The catalytic center localization contributes to the stabilization of the entire molecule, with Zn^{2+}binding playing a key role in the folding of this protein.

Introduction

Endolysins represent a group of proteins encoded by bacteriophages. The major biological role of endolysins is to destroy the peptidoglycan of the bacterial cell wall during the final stage of the lytic phage development cycle; i.e., lysis of cells to release phage progeny. According to the type of bonds hydrolyzed in the peptidoglycan, endolysins are divided into five classes: 1) lysozyme-like muramidases; 2) lytic transglycosidases; 3) N-acetyl-β-D-glucosaminidases; 4) N-acetylmuramyl-L-alanyl amidases; and 5) peptidases.[1, 2] Currently, due to the emergence of antibiotic-resistant pathogens, the large number of bacteriophage lytic enzymes is considered as an alternative to antibiotics for the treatment and prevention of the infections of bacterial origin.

Endolysins of bacteriophages infecting Gram-positive hosts typically have modular organization, possessing a catalytic domain (usually localized at the N-terminus) and a C-terminal binding domain.[2, 3] It is likely that this modular structure represents the result of co-evolution of phages and their hosts. Peptidoglycans of the cell walls of the Gram-positive bacteria are not protected by an outer membrane as in Gram-negative bacteria and therefore can be subjected to the "lysis from without". It is believed that, due to their fairly narrow substrate specificity, the C-terminal domains of these endolysins provide irreversible binding of the enzyme to the cell-wall after the release of the phage progeny, thereby preventing the participation of the enzyme molecules in the "lysis from without" of other potential hosts of the phage.[3] In contrast, endolysins of phages infecting Gram-negative hosts typically comprise a single domain that combines all the functions of these proteins, such as recognition, substrate binding, and hydrolysis.[3]

Earlier, the endolysin of the virulent coliphage T5 (EndoT5) has been identified and biochemically characterized.[4] This enzyme is a Ca^{2+}-dependent L-alanyl-D-glutamate peptidase that hydrolyzes the link between L-alanine and D-glutamate residues in certain bacterial cell-wall peptidoglycan and belongs to the subfamily C of the family M15 of zinc-containing metallopeptidases.[4, 5] Members of this protein subfamily are found in two moderate phages A118 and A500 that infect Gram-positive bacteria from the *Listeria* genus.[6] In contrast to these proteins, EndoT5 is specific to the cell walls of the Gram-negative microorganisms containing peptidoglycan of the A1γ type[7] and containing no teichoic or teichuronic acids characteristic of the cell-walls of Gram-positive bacteria.[4]

Similar to the vast majority of other endolysins from phages with the Gram-positive hosts, endolysins from phages A118 and A500 are characterized by the modular structure. They contain an N-terminally-located catalytic domain and a C-terminal cell wall-binding domain connected by a short linker. Catalytic domain of the phage A500 endolysin, EAD500, currently is the only L-alanyl-D-glutamate peptidase with known 3D-structure solved by X-ray crystallography.[8] This X-

ray analysis also revealed the presence of a Zn^{2+} cation in the active site of the enzyme.[8]

Figure 1 represents the sequence alignment of the EndoT5 and EAD500 (a catalytic N-terminal domain of the phage A500 endolysin) that reveals moderate but statistically significant similarity 28.0% identity, E = 5×10^{-5}) between these two proteins. This raises an important question on the peculiarities of structural organization of the EndoT5, a sole domain of which is able to simultaneously carry out two different functions, peptidoglycan recognition and its hydrolysis. The answer to this question requires a detailed structural characterization of the EndoT5.

```
Q6QGP7  20   LQKVARRALELSPYDFTIVQGIRTVAQSAQNIANGTS------FLKDPSKSKHITGDAID 73
             ++K+A+ +  L      + QG R+ A+     A G +            +S H  G A+D
Q37979  33   IKKMAKEGIYLC-----VAQGYRSTAEQNALYAQGRTKPGAIVTNAKGGQSNHNYGVAVD 87

Q6QGP7  74   FAPYIN-GK-IDWNDLEAFWAVKKAFEQAGKELGIKLRFGADWNASGDY-HDEI 124
                Y N GK + W    + W   K    A K  G K  +G DW +  DY H E+
Q37979  88   LCLYTNDGKDVIWESTTSRW---KKVVAAMKAEGFK--WGGDWKSFKDYPHFEL 136
```

Figure 1. Sequence alignment of the EndoT5 (UniProt ID: Q6QGP7) and EAD500 (residues 1-167 of the UniProt ID: Q37979) using BLAST2seq tool (http://blast.ncbi.nlm.nih.gov/). Note that although there are 137 and 167 residues in the EndoT5 and EAD500, respectively, only central regions of these proteins (residues 20-124 and 33-136, respectively) were aligned.

It should be noted that relatively few crystal structures of endolysins are known. It is likely that one of the reasons for this is a high intramolecular mobility of these proteins in general and of the modular endolysins in particular. In fact, as a rule, only one domain of modular endolysins can be crystallized.[9] This imposes noticeable restrictions on the structural characterization of endolysins by X-ray crystallography. At the same time, nuclear magnetic resonance provides a unique way to study the structural and conformational behavior of a protein (even very flexible one) in solution and its interactions with the substrates. In this work, we used the high-resolution NMR spectroscopy to analyze the 3D-structure of EndoT5-Zn^{2+} in solution.

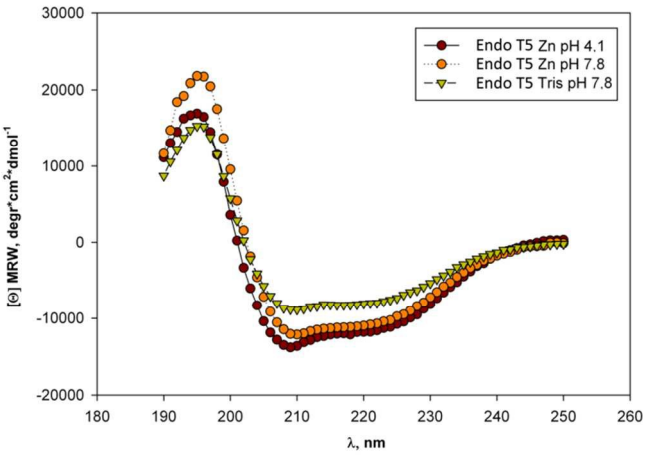

Figure 2. Far-UV CD spectra of the EndoT5 at pH 7.8 (inverse triangles) and EndoT5-Zn^{2+} at pH 7.8 or pH 4.1 (orange and dark red circles, respectively).

Results

Characterization of the EndoT5-Zn^{2+} solution structure by high-resolution NMR

We have solved the 3D-structure of EndoT5 in solution by high-resolution NMR. It should be emphasized here that, based on the spectrophotometric functionality test, the protein preparations used for structural analysis in our study retained full enzymatic activity (see Materials and Methods section for details). By the nature of its structural organization, this peptidoglycan hydrolase can be classified as a LAS metallopeptidase of the Van X type.[10] All LAS peptidases are characterized by the presence of three catalytic residues, two histidines and one aspartic acid. These catalytic residues coordinate a zinc atom and have the same relative localization within the structures of LAS peptidases. Similar to their positioning in other LAS metallopeptidases, these catalytic residues are located within the central part of the EndoT5 molecule comprising the four-stranded antiparallel β-sheet with the conserved topology.

According to the far-UV circular dichroism (CD) analysis, the zinc-depleted apo-form of EndoT5 contains 25% of α-helical and 26% of β-sheet structure (Figure 2). Interaction of the EndoT5 polypeptide with a Zn^{2+} ion leads to the formation of the holo-form, EndoT5-Zn^{2+}, which is accompanied by noticeable changes in the secondary structure content. In fact, the α-helix content rises to 44% and the amount of β-structure is reduced to 10%. Importantly, our far- UV CD analysis also revealed that EndoT5-Zn^{2+} retains its helical structure even under acidic conditions (pH 4.1). Therefore, Zn^{2+} binding induces noticeable folding of EndoT5 and stabilizes folded structure of this protein.

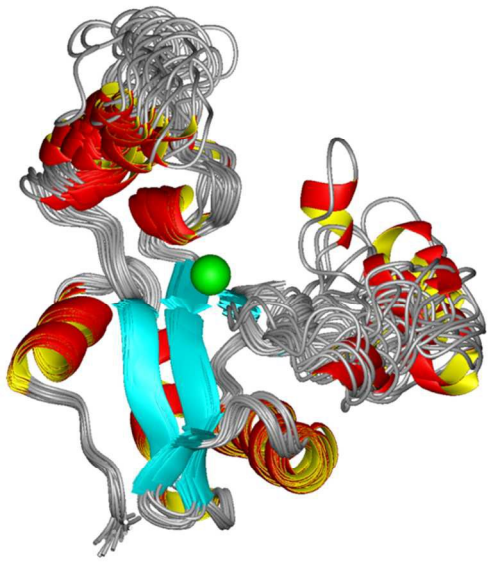

Figure 3. Solution 3D structure of the EndoT5-Zn^{2+}.

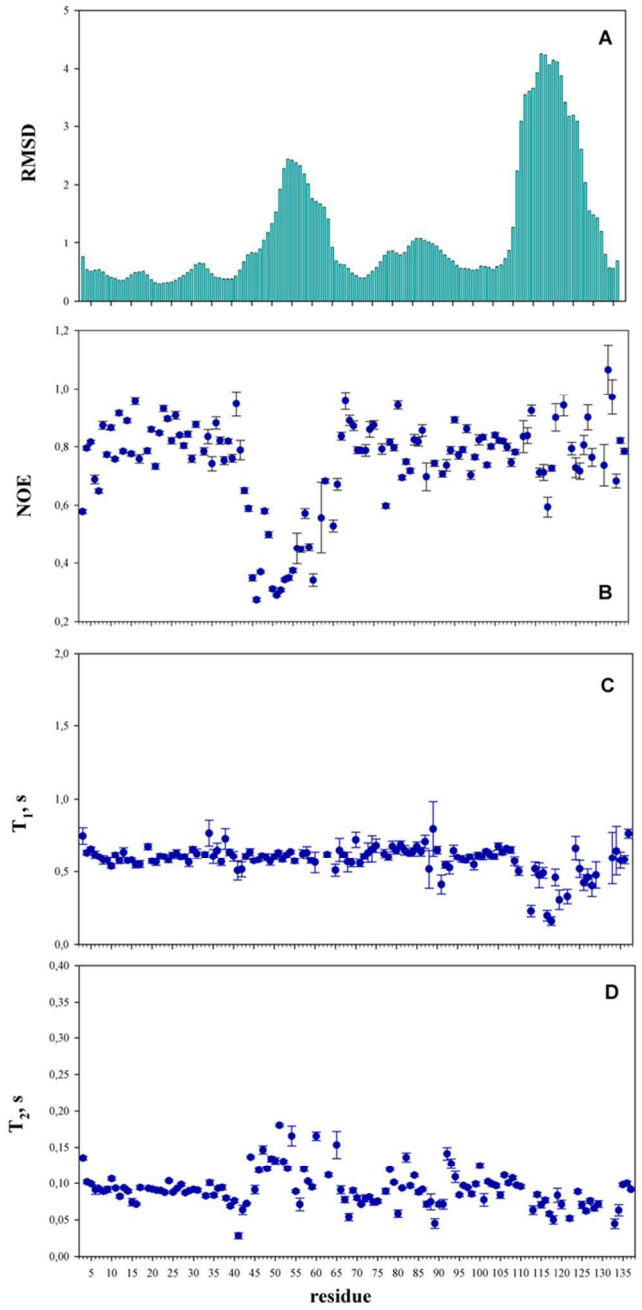

Figure 4. Comparison of the backbone dynamics and backbone RMSD calculated for the ensemble of 20 structures presented in Figure 3 for the Endo T5-Zn^{2+} protein at 298 K and pH 4.1. (A) RMSD; (B) NOE; (C) T1; (D) T2.

These conclusions on the importance of zinc for stabilizing EndoT5 structure are supported by the analysis of the spatial structure of EndoT5-Zn^{2+} by solution NMR (Figure 3). Based on the spatial arrangement of the secondary structure elements, EndoT5-Zn^{2+} belongs to the class of proteins with the α+β fold, which are composed of the α-helices and antiparallel β-strands occurring separately along the backbone. The globular core of this protein is organized by the hydrophobic residues that are involved in the formation of three α-helices (residues 7-15, 20-30, and 87-104) and three β-strands (residues 35-39, 71-76, and 133-135) that form the antiparallel β-sheet. These core secondary structure elements are interconnected with the five loops: two short loops (residues 16-19 and 31-34), one medium-length loop (residues 77-86) that also includes a short β-hairpin (residues 77-82), and two long loops (residues 40-70 and 105-132). Long loops are not completely irregular but contain dynamic α-helices (residues 46-53 and 113-117) and a stable 3_{10}-helix (residues 66-68). The overall contents of α-helical and β-sheet structures are amounts to 40% and 11%, respectively, which is consistent with the results obtained from the far-UV CD analysis. Catalytic Zn^{2+}-binding site is represented by three amino acid residues, His_{66}, Asp_{73}, and His_{133}. His residues are located near the foundations of the long loops, whereas the Asp_{73} residue is positioned in the heart of the β-sheet core.

Obviously, such spatial organization of the catalytic site contributes to the stabilization of the whole molecule. Upon binding of the Zn^{2+} ion, the distant parts of a polypeptide chain are brought together via the zinc coordination. Given the extensive changes in the EndoT5 secondary structure caused by the interaction with zinc, one can conclude that the Zn^{2+} binding is important for the correct folding of the protein and for a simultaneous formation of the catalytic site of the enzyme.

Distribution of the local RMSD values evaluated for the ensemble of the EndoT5-Zn^{2+} structures (Figure 4) indicates a relatively high intramolecular mobility in the areas of the long loops. There is evidence that the extended regions of a polypeptide chain with relatively high internal mobility can destabilize the globular moiety of protein.[11-14] The nature of this influence was investigated using the $^{1}H/^{15}N$ relaxation experiments. To this end, values ^{15}N-T_1; ^{15}N-T_2 and the magnitude of the ^{1}H-^{15}N heteronuclear Overhauser effect were measured for the ^{1}H-^{15}N pairs of the main chain (Figure 4). Figure 4 shows that within the protein region (residues 3-39) located before the first long loop, there are only relatively small ^{15}N-T_2 changes between the adjacent residues along the polypeptide chain. In the spatial structure of the EndoT5, this region corresponds to a quasi-domain consisting of two α-helices (residues 7-15 and 20-30) and one β-strand (residues 35-39) that caps a hydrophobic interface of these helices. By its structural and dynamic properties this region constitutes a cooperative unit and can represent a site from which the protein folding process begins.

The first long loop (residues 40-70) is characterized by a relatively high intramolecular mobility, which is reflected in the corresponding ^{15}N-T_2 and ^{1}H-^{15}N NOE values. The second β-strand (residues 71-76) is included into the globular part of a protein molecule and is fixed within its structure. It is clear that the Zn^{2+}-binding site containing Asp_{73} residue has a significant contribution to the stabilization of this β-strand. The medium-length loop (residues 77-86) containing a short β-hairpin (residues 77-82) has a significant internal mobility. Several residues in this region (such as Ile_{78}, Gly_{80}, and Ile_{82}) can participate in the conformational exchange. The third α-helix

(residues 87-104), being a part of a globular portion of the molecule, is characterized by the larger degree of freedom than those of the N-terminal α-helices (residues 7-15 and 20-30) that are included in the aforementioned quasi-domain. Several residues of this helix (Ala_{89}, Ala_{92}, Val_{93}, and Ala_{100}) are probably involved in the conformational exchange. Apparently, a relative flexibility of this α-helix is determined by the mobile loops flanking this α-helix on both sides.

The short N-terminal region (residues 105-110) of the second long loop (residues 105-132) is largely immobile, whereas the remaining part of this loop has significant intramolecular mobility. In fact, mostly sequential and intra-residual NOE signals can be identified within this part. At the same time, we did not detect NOE signals corresponding to the interactions between the distant residues. It is likely that this fact determines the relatively high RMSD values for the residues 111-132. However, the ^{1}H-^{15}N NOE values determined for these residues are more characteristic to the loop with a relatively low intramolecular mobility. Also, the ^{15}N-T_1 and ^{15}N-T_2 values for these residues are relatively low. Taken together, these data suggest that this loop area is able to loosely bind some paramagnetic ions, which, in the form of minor impurities, are present in the $ZnCl_2$ used in the preparation of the sample.

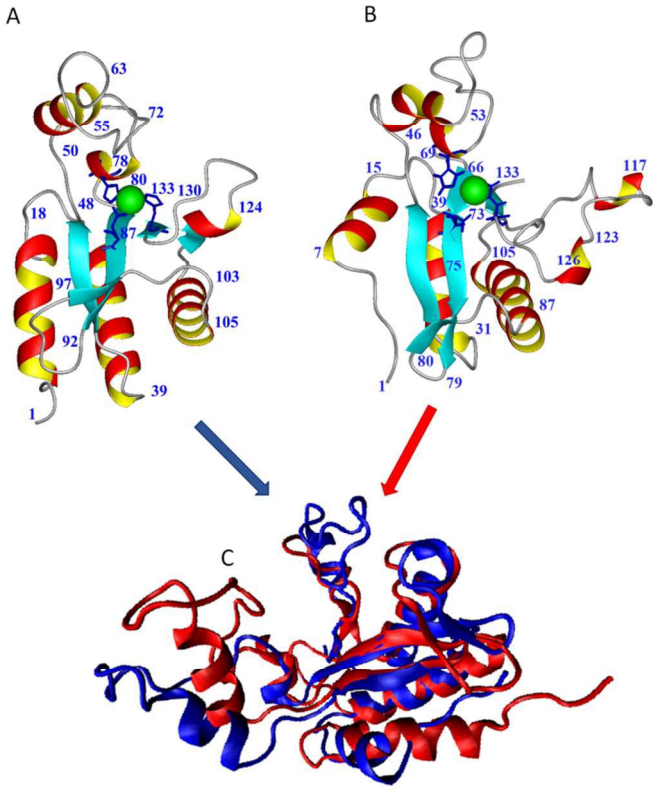

Figure 5. Comparison of the 3D structures of EndoT5-Zn^{2+} (**A**, PDB ID: 2mxz) and a catalytic N-terminal domain of the phage A500 endolysin (**B**, PDB ID: 2vo9). These plots represent details of structures of EndoT5-Zn^{2+} and EAD500 discussed in the text. **C**. Structural alignment of EndoT5-Zn^{2+} (blue structure) and EAD500 (red structure) by MultiProt (http://bioinfo3d.cs.tau.ac.il/MultiProt/).[15] Structure shown in plot **C** were created using the Visual Molecular Dynamics tool VMD 1.9.2.[16]

Fixation of the C-terminal end of the loop is provided by a short β-strands comprising three residues 133-135, one of which, His_{133}, participates in the formation of the catalytic site of the enzyme. Therefore, these observations provide a further support to the aforementioned notion that catalytic center plays a key role in the process of the enzyme folding and stabilizes its spatial structure.

Comparative structural analysis of EndoT5 and enzymatically active domain EAD500 of the endolysin from the bacteriophage A500

The EAD500 catalytic domain of the Listeria bacteriophage A500 was the only L-alanyl-D-glutamate peptidase whose 3D structure was known so far.[8] As aforementioned, the identity of amino acid sequences of EndoT5 and EAD500 is only ~28% ($E=5\times10^{-5}$). Therefore, these enzymes can be considered as distant homologues. On the other hand, the BLAST search of UniProt revealed the presence of a large number of the close orthologs of the EndoT5 (more than 100 proteins with > 40% identity and E value less than 10^{-21}). Unfortunately none of these close orthologs was biochemically characterized.

Despite the fact that the EndoT5 and EAD500 are only distantly homologous, comparative analysis of their 3D-structures revealed a number of significant similarities in their spatial organization (see Figure 5). For example, visual inspection of structures shown in Figure 5A and 5B suggests that these proteins are characterized by very similar folds of their polypeptide chains. This conclusion is further supported by Figure 5C that provides the results of multiple structural alignment of these two proteins performed by the MultiProt server.[15]

Figures 5A and 5B also show that, in general, the regions of the polypeptide chains of these two proteins that form regular secondary structure elements, except for the N-terminal α-helices, retain their relative spatial arrangement. Configurations of the Zn^{2+}-binding sites and the localizations of these sites within the structures of both proteins are also completely identical.

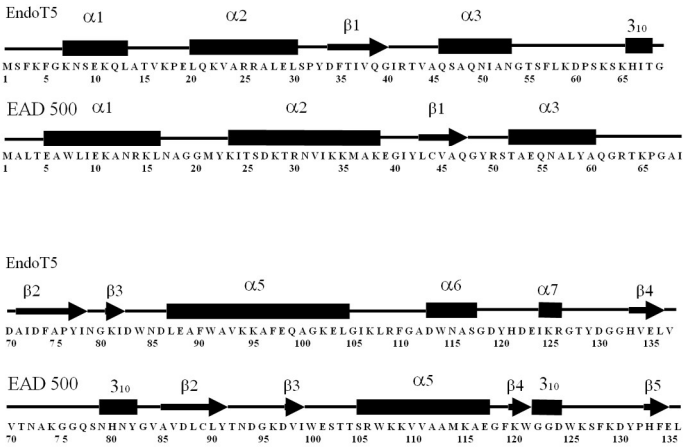

Figure 6. Localization of the secondary structure elements within the sequences of EndoT5-Zn^{2+} and EAD500.

Figure 6 represents the peculiarities of the distribution of the secondary structure elements within the amino acid sequences of the EndoT5 and EAD500 and confirms the presence of significant structural similarity between these two proteins.

On the other hand, Figures 5 and 6 show that there are some noticeable differences between the 3D structures of these two proteins. These differences are most obvious in the loop regions. In fact, the N- terminal α-helix in EndoT5-Zn^{2+} is shorter, and the orientation of its axis relative to the protein's body has changed significantly. The loop containing conserved residue His_{66} involved in the formation of the active center in the EndoT5-Zn^{2+} is slightly shorter (31 residues, Gly_{40}-Asp_{70}) than that of the EAD500 (37 residues, Gly_{48}-Val_{84}). The loop (Thr_{92} - Asp_{97}) connecting β-strands 2 and 3 (Figure 5) in the EAD500,[8] is noticeably shorter in the structure of the EndoT5-Zn^{2+} where it is reduced to an optimized β-turn Asn79-Gly80. The C-terminal 21 residue-long loop in the EndoT5-Zn^{2+} (Gly_{111}-Gly_{131},) is significantly longer than the homologous C-terminal loop in the EAD500 (Gly_{122}-Tyr_{131}, 10 residues). The long C-terminal loop of the EndoT5-Zn^{2+} contains a dynamic α-helix (residues 113-117). It is likely that the lability of this helix is determined by flexible loops flanking this α-helix on both sides.

Evaluating disorder propensities of EndoT5 and EAD500

To investigate the peculiarities of the intrinsic disorder distribution along the protein sequences, algorithms from the PONDR® family have been used. The disorder profiles were obtained by an accurate meta-predictor of intrinsic disorder, PONDR-FIT,[17] and a rather accurate stand-alone disorder predictors PONDR® VSL2,[18] which, based on the comprehensive assessment of in silico predictors of intrinsic disorder, was shown to perform reasonably well.[19, 20] Figure 7 represents the results of this analysis for the EndoT5 (Figure 7A) and the EAD500 catalytic domain of the L-alanyl-D-glutamate peptidase from the Listeria phage A500 (Figure 7B). Despite their low sequence similarity, the disorder propensities of both proteins are rather similar. In fact, N- and C-termini of EndoT5 and EAD500 are predicted to be disordered and both proteins have a central disordered region (centered at residue 60). Positions of predicted disordered regions correlate rather well with the localization of the long loops. Furthermore, the dips in the EndoT5 disorder profile around residues 50 and 110 mark positions of the aforementioned dynamic α-helices (residues 46-53 and 113-117) found within the EndoT5 long loops. Therefore, the results of this analysis are consistent with an important message that some characteristic features seeing in disorder profiles are rather conserved and that such conserved distribution of disorder-related sequence features can have functional implications. Figure 7 also shows that the results of disorder evaluation for these two proteins by two different predictors generally agree. According to PONDR® VSL2 and PONDR-FIT, EndoT5 is predicted to have 32% and 21% disordered residues respectively, whereas for EAD500, the corresponding values are 12% and 16%.

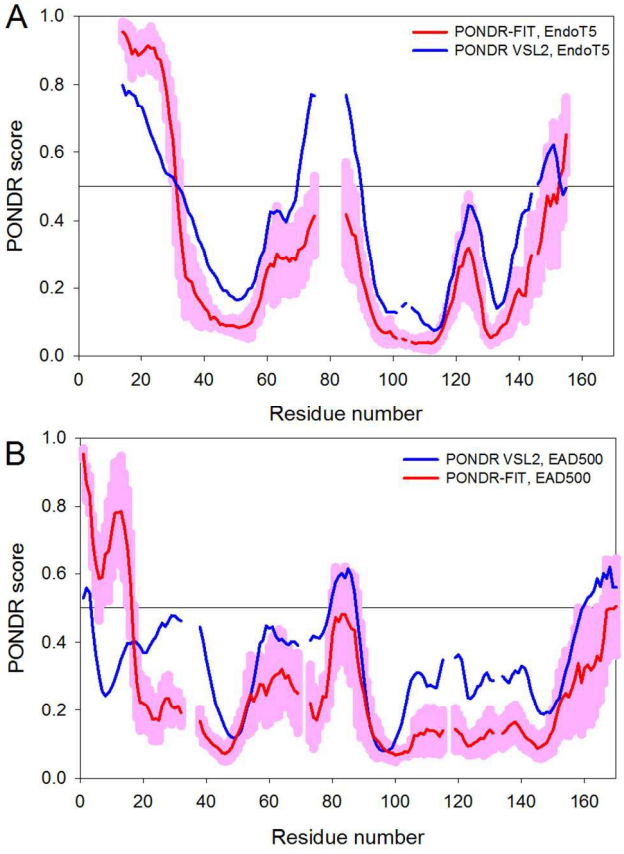

Figure 7. Peculiarities of the intrinsic disorder propensities of the EndoT5-Zn^{2+} (**A**, UniProt ID: Q6QGP7) and EAD500 (**B**, residues 1-167 of the UniProt ID: Q37979). Here, sequences are aligned as in Figure 1, and where the gap in each line corresponds to the gap in the sequence alignment. Disorder propensities were evaluated by PONDR-FIT (red curves with light pink shades) and PONDR® VSL2 (blue lines). A disorder threshold is indicated as a thin line (at score = 0.5). Residues/regions with the disorder scores >0.5 are considered as disordered.

Typically, two arbitrary cutoffs for the levels of intrinsic disorder are used to classify proteins as highly ordered (IDP score<10%), moderately disordered (10%≤IDP score<30%) and highly disordered (IDP score≥30%).[21] Therefore, according to this classification and based on the PONDR® VSL2 analysis, EndoT5 is predicted as highly disordered protein, whereas EAD500 is expected to be moderately disordered.

Peptidoglycan of the Listeria cell walls belongs to the same A1γ type as the peptidoglycan of the Gram-negative bacteria.[7] However, as it was shown earlier, the rate of lysis of the Listeria cells by the EndoT5 enzyme was 4 orders of magnitude lower that the rate of lysis of the *E. coli* cells.[4] Since the structures of EndoT5 and EAD500 were generally similar, except to their loops, it is likely that the substrate specificity of EndoT5 is determined by the structural peculiarities of these loops.

Discussion

Structural similarity of proteins with similar function but weak amino acid homology is a rather common phenomenon. For

example, the 3D-structures of 32 different flavoproteins are known to atomic resolution, and their folding topologies are much more similar to each other than would be expected based on the comparison of their sequences.[22] There are also examples among endolysins, where the CD27L and PlyPSA amidases specific to different types of Clostridia have a high structural similarity with just a 30% similarity between their amino acid sequences.[8] Therefore, the found in our study structural similarity between EndoT5 and EAD500 provide a strong support to the important notion that the convergent evolution of proteins occurs at the levels of secondary and tertiary structure, rather than at the primary structure level, and that this evolution is largely determined by the adjustments of the functional roles of certain protein regions.

On the other hand, there is an important example of the 56-residue-long artificial protein in which only one mutation at position 45 - replacement of leucine by tyrosine - leads to a transition of a molecule from a 3α configuration to the 3β+α fold.[23] This example shows that the presence of high levels of primary structure homology does not always correspond to the high levels of structural similarity or identity at the secondary and tertiary structure levels. Therefore, these data suggest that the information on the protein functional and structural properties is encoded within the protein primary structure in a specific manner. Here, protein structural elements are defined by specific combinations of several residues rather than by a certain amino acid and/or its specific location within the polypeptide chain. Sometimes such different in content combinations of residues can give rise to similar structural elements. However, sometimes replacing just one residue in a particular position can lead to a new and exceptional combination of residues that radically changes the meaning of the entire "sentence" (polypeptide chain), very much like the comma in the well-known example illustrating the "punctuation saves lives" concept ("Let's eat Grandma!" versus "Let's eat, Grandma!").

Functional roles of long loops

Stretches of the polypeptide chain characterized by the relatively high internal mobility are able to destabilize the globular part of the protein molecule.[11-14] According to the far-UV CD analysis (see Figure 2), these destabilizing effects of moving loops in EndoT5 are apparently so strong that the overall fold of a protein is noticeably distorted in the absence of the Zn^{2+} binding, specific coordination of which within the catalytic site "cross-links" the distant parts of the macromolecule.

The aforementioned destabilization of the globular part of the EndoT5 can be regarded as a cost of the implementation of its biological activity which requires the simultaneous presence in the molecule of a tightly packed globular core for the catalytic activity and the loosely packed flexible loops with a large number of internal degrees of freedom and hence a high internal mobility needed for substrate recognition. Obviously, due to the combination of such contradictory requirements, a conventional set of non-covalent intramolecular interactions (Van der Waals, hydrophobic, and electrostatic interactions, hydrogen bonding, etc.) is no longer sufficient for the stabilization of the globular protein. In this case, the specific coordination of the zinc atom acts as a set of staples that helps to compensate for the destabilization of the globular portion of the molecule by its highly mobile loops.

For this scenario to take place, the two longest loops must by functionally significant. This hypothesis is supported by the observations that the first EndoT5 long loop contains conserved residues Arg_{42}, Gln_{46}, and Ser_{64}, which are analogous to the catalytically important Arg_{50}, Gln_{55}, and Ser_{78} in EAD500.[8] The second long loop of the EndoT5 includes conserved residue Asp_{130}, which is analogous to the Asp_{130} residue in EAD500, where it is also important for the catalysis. However, opposite to the catalytic conserved residue Asp_{130}, EAD500 has another conserved residue, Asp_{124}.

Structural analysis of Ca^{2+}-binding proteins revealed that the loops with this mutual position of Asp residues are able to bind Ca^{2+} ions.[24, 25] Interacting with such loops, the Ca^{2+} ion forms the coordination bonds with the oxygen atoms of the carbonyl groups within the polypeptide chain backbone. Free valences of the Ca^{2+} ion are saturated by the coordination bonds with the oxygen atoms of the carboxyl groups of the Asp residues. This type of coordination fully neutralizes the positive charge of the ion. As a result, the loop is cross-linked by the coordinated ion, which leads to a change in structural and dynamic characteristics of this loop. Curiously, the second long loop of EndoT5 contains four Asp residues (at the positions 113, 119, 122, and the aforementioned 130) and therefore has a potential to bind Ca^{2+} ions. In agreement with this hypothesis, it has been shown that the enzymatic activities of EndoT5 and EAD500 are modulated by Ca^{2+} ions.[4, 26] We suggest this could likely be due to the ion binding to the long loops of these proteins.

Materials and Methods

Isolation and purification of endolysin

To obtain labeled EndoT5 E.coli BL21 (DE3) cells carrying plasmid pT5lys (ref.[4]) were grown in the M9 minimal medium containing $^{15}NH_4Cl$ as a nitrogen source, and $^{13}C_6N_{12}O_6$ as a carbon source until optical density at 550 nm reached 0.6 units. Then, the synthesis of the target protein was induced with 0.8 mM IPTG. After the induction, the cells were grown for 3 hours and harvested by centrifugation at 3,000 g for 10 minutes. For protein purification, E.coli BL21 (DE3) cells (1.1 g) were suspended in 10 ml buffer containing 25 mM Tris-HCl (pH = 8.0), 40 mM NaCl and 1 mM EDTA. The cell suspension was sonicated at 75 W for 1 minute (two 30 s pulses), then cleared by centrifugation at 20,000 g for 30 minutes. The supernatant (9.5 ml) was passed through a 10 ml Toyopearl DEAE 650M column, then loaded onto a 5-ml P_{11} phosphocellulose column equilibrated with the same buffer. Proteins were eluted with a linear gradient of sodium chloride (0.04-0.50 M) in a volume of

90 ml. Fractions (2 mL each) were analyzed by the 15% polyacrylamide gel electrophoresis. The target EndoT5 protein was eluted by 0.3 M NaCl. Desalting of the sample was conducted by dialysis against aqueous NH_3 solution, pH = 8.5. The desalted preparation was lyophilized.

Sample preparation

Since due to the relatively slow exchange of amide protons the quality of the NMR spectra is noticeably increased at slightly acidic pH, NMR analysis of EndoT5-Zn^{2+} was performed at pH 4.1. However, the direct substitution of the buffer used for protein isolation (pH 8.0) to the buffer for NMR analysis (pH 4.1) was accompanied by noticeable EndoT5-Zn^{2+} precipitation due to the crossing the protein isoelectric point. Therefore, unfolding by guanidine hydrochloride with subsequent refolding to the NMR analysis buffer was used for the fast and efficient buffer substitution in protein samples without quantitative loss of the material. To this end, lyophilized EndoT5 protein was dissolved in 0.5 ml solvent containing 6 M guanidinium hydrochloride and dialyzed six times against 200 ml of saline solution containing 5 mM $ZnCl_2$ and 0.01% NaN_3. Refolded in the presence of Zn^{2+} protein was lyophilized and dissolved in 50 mM deuterated sodium acetate buffer containing 0.03% NaN_3, pH 4.1. Enzymatic activity was measured spectrophotometrically by following the decrease in the absorbance at 450 nm of the *E.coli* cells, pretreated with chloroform as previously described.[4] Protein samples prepared via this unfolding/refolding cycle were characterized by the enzymatic activity comparable with the activity of the initial protein preparations ($5\text{-}7\times10^3$ U/mg protein).

NMR spectroscopy

NMR samples contained 0.8 mM [$^{15}N/^{13}C$] EndoT5-Zn^{2+} in 50 mM sodium acetate buffer, pH 4.1, containing 0.03% NaN_3, in a $H_2O/^2H_2O$ mixture (9/1). The pH value of 4.1 was chosen to reduce the protein association and prevent it autolysis during the experiment. All NMR experiments were conducted on the AVANCE III 600 spectrometer equipped with a triple resonance probe with pulsed Z-gradient. The experiments were run at a temperature of 298 K.

For the primary data processing and automatic collection of signal positions in the NOESY spectra, the TOPSPIN 2.1 (Bruker Biospin, Karlsruhe, Germany) program was used. The assignments of ^1H, ^{13}C, and ^{15}N signals in the NMR spectra were performed using the CARA algorithm.[27, 28] Semiautomatic assignment polypeptide chain signals to a particular residue in the primary sequence of the protein was carried out based on the data of the 2D-$^1H/^{15}N$-HSQC, 3D-HNCACB and 3D-CBCA(CO)NH spectra using an integrated AutoLinc-9.4 module[29] in the program CARA. At the next stage, HNCACB and CC(CO)NH spectra were used to combine fragments of the semi-automatic assignment and to verify its authenticity.

Furthermore, the CC(CO)NH spectrum was also used to assign ^{13}C resonances of the aliphatic side chain residues. Corresponding proton resonances were established using 3D-^{15}N-TOCSY and 3D – HCCH-TOCSY experiments. The chemical shifts of carbons of the carbonyl groups were determined based on the 3D-HNCO spectrum. ^1H- and ^{13}C-resonances of the side chains of aromatic residues were established based on 2D – CBHD and 3D – HCCH-TOCSY spectra. To classify indole protons of tryptophan residues Trp_{84}, Trp_{91}, and Trp_{114}, the 2D-$^1H/^{15}N$-HSQC, 3D-^{15}N-TOCSY- and 3D-^{15}N-NOESY spectra were used. Chemical shifts of protons are given relative to the signal of the protons of 2,2-dimethyl-2-sylapentan-5-sulfonate, as recommended by the IUPAC.[30-32] Chemical shifts of other nuclei were recalculated from the ratio of the gyromagnetic factors.[33] To study the dynamics of EndoT5-Zn^{2+} parameters ^{15}N-relaxation were measured using the ^{15}N-labeled sample and pulse sequences at the field strength of 14.1 T and a temperature of 298 K.[34] The values of the spin-lattice (T_1) and spin-spin (T_2) relaxation of the ^{15}N nuclei were calculated from a series of 2D-$^1H/^{15}N$-correlation spectra obtained with the following variation of the relaxation delays: 10, 60, 120, 240, 480, 700, 900, 1200, 1500, and 8, 48, 64, 120, 180, 220, 280, and 370 ms for T_1 and T_2, respectively. Values of the heteronuclear $^1H/^{15}N$ NOE were calculated as the ratio of the intensities of the signals in the two-dimensional $^1H/^{15}N$ correlation spectra with and without saturation of the amide protons within 2.7 seconds.

Calculation of the spatial structure

Distance limitations were obtained from the 3D – ^{15}N- and ^{13}C-NOESY spectra with the mixing time of the magnetization components equal to 80 ms for the ^{15}N-NOESY; 160 ms to ^{13}C-NOESY-ali and 100 ms for ^{13}C-NOESY-aro. Editing and integration of the collected signals were performed using CARA.[28] Torsion angles φ and ψ of the polypeptide chain were predicted by the TALOS program based on the chemical shifts of the atomic nuclei of the backbone ^1HN, ^{15}N, ^1H, ^{13}C, and ^{13}CO.[36]

To detect hydrogen bonds 3D-HNCO sequence was used,[37] which was modified for the 2D-version of the registration and optimized for observation of the distant $^{3h}J_{NC'}$ constants realized through hydrogen bonds.[38] We were able to find 9 reliably detected hydrogen bonds.

According to the geometric criteria,[39, 40] the analyzed structure might contain 21 hydrogen bonds. For each hydrogen bond, restrictions were introduced on the distance: two upper (d(O,HN) = 2.0 Å, d(O,N) = 3.0 Å) and two lower d(O,HN) = 1.8 Å, d(O,N) = 2.7 Å). Therefore, in subsequent calculations of final spatial structure of EndoT5, 84 additional distance constraints were used.

The structure was calculated using the CYANA 2.1 algorithm.[41] A standard protocol was used that included seven cycles of automatic NOE assignment and calculation of structures of 100 conformations for each cycle. As a result, the 20 structures were selected with a minimum value of the target function, which were used for the final analysis. The resulting ensemble of structures was consistent with the experimental data and showed a good Ramachandran statistics (Table 1).

Table 1. Statistics of the calculation of the EndoT5-Zn^{2+} structure[a]

Used restrains[b]	
Spatial restraints for the NOE contacts	
Intra-residual (i–j = 0)	393
Sequential (|i–j| = 1)	487
Mean distance (1 < |i–j| ≤ 4)	299
Long distance (|i–j| ≥ 5)	690
Total number of the NOE-based restraints	1869
Restraints on dihedral angles	
Dihedral angles (φ/ψ)	112/113
Restraints of the hydrogen bonds (upper / lower)	42/42
Restrains of the coordination bonds	
With Zn^{2+} ion (upper / lower)	6/6
Violation of the NOE restraints	
at distance > 0.1 Å	9
at distance > 0.2 Å	1
on dihedral angles > 5°	0
RMSD (Å)[c] (residues 3-41; 70-110; 133-137)	
main chain atoms	0.29±0.11
all heavy atoms	0.85±0.07
Ramachandran's plot analysis[d]	
Residues in the most preferable regions, %	81.3
Residues in other allowed regions, %	18.2
Residues in conditionally allowed regions, %	0.5
Residues in forbidden regions, %	0.0

[a] This statistics was obtained for 20 calculated Endo T5-Zn^{2+} structures with the best target function

[b] Spatial restraints retrieved using contacts.

[c] Root mean square deviation (RMSD), calculated via the pair-wise comparison of all the structures in the ensemble with the averaged structure.

[d] Ramachandran's plot for all the residues, including first seven residues with very high mobility was obtained using the program PROCHECK-NMR available at the www.rcsb.org/ site.

Unfortunately, the NMR does not provide direct information on the Zn^{2+} ion localization. However, the preliminary calculations of the EndoT5 structure conducted without considering Zn^{2+} ion revealed close proximity of residues His66, Asp73, and His133. The relative spatial arrangement of these residues corresponded well to the location of residues His80, Asp87, and His133 in the X-ray structure of EAD500, where these residues are involved in the Zn^{2+} ion ligation. These residues are conserved in metallopeptidases from the M15 family. These observations and considerations allowed us to introduce spatial constraints on the ion Zn^{2+} in the calculation of the final structure. The subsequent calculations of the EndoT5-Zn^{2+} final structure where the aforementioned spatial constraints based on the X-ray data and the presence of the Zn^{2+} ion were taken into account did not reveal any steric conflicts. Furthermore, the automatic NOE assignment algorithm revealed the presence of some additional spatial constraints between the HE1-proton of His133 and the amide protons of residues Asp73 and Ile72, and the delta-protons of the Ile72 residue.

For a visual representation of protein structures, the MOLMOL program was used.[42] The experimental NMR constraints and the atomic coordinates of the 20 structures of EndoT5-Zn^{2+} were deposited to the PDB database (access code 2mxz). The assignments of the ^{1}H, ^{13}C, and ^{15}N signals were deposited to the BioMagResBank database (accession number 25437).

Evaluation of the intrinsic disorder propensities of EndoT5 and EAD500

The intrinsic disorder propensities of EndoT5 and EAD500 were evaluated by two disorder predictors of the PONDR family, PONDR® VSL2 (ref.[18]) and PONDR® FIT.[17] Here, scores above 0.5 are considered to correspond to the disordered residues/regions. PONDR® VSL2 is one of the more accurate stand-alone disorder predictors,[18] and based on the comprehensive assessment of in silico predictors of intrinsic disorder,[19, 20] this tool was shown to perform reasonably well. PONDR-FIT represents a metapredictor that combines six individual predictors, which are PONDR® VLXT [43], PONDR® VSL2,[18] PONDR® VL3,[44] FoldIndex,[45] IUPred,[46] and TopIDP.[47] PONDR-FIT is moderately more accurate than each of the component predictors.[17]

Conclusions

We defined the spatial solution structure of the Zn^{2+}-containing form of the bacteriophage T5 L-alanyl-D-glutamate peptidase (EndoT5-Zn^{2+}). This α+β protein has a globular core formed by three α-helices and a β-sheet containing three β-strands. The interesting feature of this protein is the presence of several rather long loops containing a short β-hairpin, a stable 3_{10}-helix and labile α-helices. Zn^{2+} ion in a catalytic site is coordinated by His_{66}, Asp_{73}, and His_{133}. The cation-binding His residues are located near the foundations of the long loops, whereas Asp_{73} is positioned in the middle of the core β-sheet. Therefore, the structural integrity of the entire molecule is controlled by the Zn^{2+} binding that play a key role in the folding and conformational stability of this protein.

Acknowledgements

This work was financially supported by the Government of the Moscow region and RFBR grant № 14-44-03573.

Notes and references

[a]Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia;
[b]Branch of Shemyakin & Ovchinnikov's Institute of Bioorganic Chemistry, Russian Academy of Sciences, Prospekt Nauki 6, Pushchino, Moscow Region 142290, Russia;
[c]Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA;
[d]Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia;
[e]Department of Biology, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Kingdom of Saudi Arabia;
[f]Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, Tikhoretsky Ave. 4, St. Petersburg 194064, Russia
*Authors to whom correspondence should be addressed; VPK, Phone: +7 (4967) 73-92-26; Fax: +7 (4967) 33-05-53; E-mail: kutyshenko@rambler.ru; VNU, Phone: 813-974-5816; Fax: 813-974-7357; E-mail: vuversky@health.usf.edu.

1. J. Borysowski, B. Weber-Dabrowska and A. Gorski, *Experimental Biology and Medicine*, 2006, **231**, 366-377.

2. M. J. Loessner, *Current Opinion in Microbiology*, 2005, **8**, 480-487.
3. V. A. Fischetti, *Bacteriophage*, 2011, **1**, 188-194.
4. G. V. Mikoulinskaia, I. V. Odinokova, A. A. Zimin, V. Y. Lysanskaya, S. A. Feofanov and O. A. Stepnaya, *FEBS J*, 2009, **276**, 7329-7342.
5. G. V. Mikoulinskaia, I. V. Odinokova, A. A. Zimin and O. A. Stepnaya, in *Handbook of Proteolytic Enzymes*, eds. N. D. Rawlings and G. S. Salvesen, Academic Press, Oxford, Editon edn., 2013, pp. 1407-1410.
6. M. J. Loessner, G. Wendlinger and S. Scherer, *Mol Microbiol*, 1995, **16**, 1231-1241.
7. K. H. Schleifer and O. Kandler, *Bacteriol Rev*, 1972, **36**, 407-477.
8. I. P. Korndorfer, A. Kanitz, J. Danzer, M. Zimmer, M. J. Loessner and A. Skerra, *Acta Crystallogr D Biol Crystallogr*, 2008, **64**, 644-650.
9. M. Schmelcher, D. M. Donovan and M. J. Loessner, *Future Microbiol*, 2012, **7**, 1147-1171.
10. M. Firczuk and M. Bochtler, *FEMS Microbiol Rev*, 2007, **31**, 676-691.
11. M. J. Stone, S. Gupta, N. Snyder and L. Regan, *J Am Chem Soc*, 2001, **123**, 185-186.
12. M. J. Stone, *Acc Chem Res*, 2001, **34**, 379-388.
13. L. Spyracopoulos, *Protein Pept Lett*, 2005, **12**, 235-240.
14. V. P. Kutyshenko, D. A. Prokhorov, M. A. Timchenko, Y. A. Kudrevatykh, L. V. Gushchina, V. S. Khristoforov, V. V. Filimonov and V. N. Uversky, *Biochim Biophys Acta*, 2009, **1794**, 1813-1822.
15. M. Shatsky, R. Nussinov and H. J. Wolfson, *Proteins*, 2004, **56**, 143-156.
16. W. Humphrey, A. Dalke and K. Schulten, *J Mol Graph*, 1996, **14**, 33-38, 27-38.
17. B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker and V. N. Uversky, *Biochim Biophys Acta*, 2010, **1804**, 996-1010.
18. K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker and Z. Obradovic, *J Bioinform Comput Biol*, 2005, **3**, 35-60.
19. Z. L. Peng and L. Kurgan, *Curr Protein Pept Sci*, 2012, **13**, 6-18.
20. X. Fan and L. Kurgan, *J Biomol Struct Dyn*, 2014, **32**, 448-464.
21. K. Rajagopalan, S. M. Mooney, N. Parekh, R. H. Getzenberg and P. Kulkarni, *J Cell Biochem*, 2011, **112**, 3256-3267.
22. O. Dym and D. Eisenberg, *Protein Sci*, 2001, **10**, 1712-1728.
23. P. A. Alexander, Y. He, Y. Chen, J. Orban and P. N. Bryan, *Proc Natl Acad Sci U S A*, 2009, **106**, 21149-21154.
24. D. A. Nicoll, M. R. Sawaya, S. Kwon, D. Cascio, K. D. Philipson and J. Abramson, *J Biol Chem*, 2006, **281**, 21577-21581.
25. J. L. Gifford, M. P. Walsh and H. J. Vogel, *Biochem J*, 2007, **405**, 199-221.
26. M. Schmelcher, F. Waldherr and M. J. Loessner, *Appl Microbiol Biotechnol*, 2012, **93**, 633-643.
27. M. Sattler, J. Schleucher and C. Griesinger, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 1999, **34**, 93-158.
28. R. L. J. Keller, *The Computer Aided Resonance Assignment Tutorial.*, Cantina Verlag, Goldau: Switzerland, 2004.
29. J. E. Masse and R. Keller, *J Magn Reson*, 2005, **174**, 133-151.
30. J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright and K. Wuthrich, *Eur J Biochem*, 1998, **256**, 1-15.
31. J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright and K. Wuthrich, *J Biomol NMR*, 1998, **12**, 1-23.
32. J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright and K. Wuthrich, *J Mol Biol*, 1998, **280**, 933-952.
33. D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley and B. D. Sykes, *J Biomol NMR*, 1995, **6**, 135-140.
34. N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay and L. E. Kay, *Biochemistry*, 1994, **33**, 5984-6003.
35. V. Y. Orekhov, D. E. Nolde, A. P. Golovanov, D. M. Korzhnev and A. S. Arseniev, *Applied Magnetic Resonance*, 1995, **9**, 581-588.
36. G. Cornilescu, F. Delaglio and A. Bax, *Journal of Biomolecular Nmr*, 1999, **13**, 289-302.
37. S. Grzesiek and A. Bax, *Journal of Magnetic Resonance*, 1992, **96**, 432-440.
38. Z. O. Shenkarev, T. A. Balashova, Z. A. Yakimenko, T. V. Ovchinnikova and A. S. Arseniev, *Biophysical Journal*, 2004, **86**, 3687-3699.
39. D. F. Stickle, L. G. Presta, K. A. Dill and G. D. Rose, *Journal of Molecular Biology*, 1992, **226**, 1143-1159.
40. E. N. Baker and R. E. Hubbard, *Progress in Biophysics & Molecular Biology*, 1984, **44**, 97-179.
41. P. Guntert, *Methods Mol Biol*, 2004, **278**, 353-378.
42. R. Koradi, M. Billeter and K. Wuthrich, *Journal of Molecular Graphics*, 1996, **14**, 51-&.
43. P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown and A. K. Dunker, *Proteins*, 2001, **42**, 38-48.
44. K. Peng, P. Radivojac, S. Vucetic, A. K. Dunker and Z. Obradovic, *BMC Bioinformatics*, 2006, **7**, 208.
45. J. Prilusky, C. E. Felder, T. Zeev-Ben-Mordehai, E. H. Rydberg, O. Man, J. S. Beckmann, I. Silman and J. L. Sussman, *Bioinformatics*, 2005, **21**, 3435-3438.
46. Z. Dosztanyi, V. Csizmok, P. Tompa and I. Simon, *Bioinformatics*, 2005, **21**, 3433-3434.
47. A. Campen, R. M. Williams, C. J. Brown, J. Meng, V. N. Uversky and A. K. Dunker, *Protein Pept Lett*, 2008, **15**, 956-963.