PROTEIN STRUCTURE REPORTS

Structure of the N-terminal domain of the metalloprotease PrtV from *Vibrio cholerae*

Aaron Edwin,[1,2] Cecilia Persson,[3] Maxim Mayzel,[3] Sun Nyunt Wai,[2,4,5] Anders Öhman,[6] B. Göran Karlsson,[3]* and A. Elisabeth Sauer-Eriksson[1,2]*

[1]Department of Chemistry, Umeå University, Umeå SE-901 87, Sweden
[2]Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå SE-901 87, Sweden
[3]Swedish NMR Centre at the University of Gothenburg, Gothenburg SE-40530, Sweden
[4]Department of Molecular Biology, Umeå University, Umeå SE-901 87, Sweden
[5]The Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå SE-901 87, Sweden
[6]Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå SE-901 85, Sweden

Received 10 August 2015; Accepted 25 September 2015
DOI: 10.1002/pro.2815
Published online 5 October 2015 proteinscience.org

Abstract: The metalloprotease PrtV from *Vibrio cholerae* serves an important function for the ability of bacteria to invade the mammalian host cell. The protein belongs to the family of M6 proteases, with a characteristic zinc ion in the catalytic active site. PrtV constitutes a 918 amino acids (102 kDa) multidomain pre-pro-protein that undergoes several N- and C-terminal modifications to form a catalytically active protease. We report here the NMR structure of the PrtV N-terminal domain (residues 23–103) that contains two short α-helices in a coiled coil motif. The helices are held together by a cluster of hydrophobic residues. Approximately 30 residues at the C-terminal end, which were predicted to form a third helical structure, are disordered. These residues are highly conserved within the genus *Vibrio*, which suggests that they might be functionally important.

Keywords: *Vibrio cholera*; metalloproteases; PrtV; N-terminal domain; NMR

Aaron Edwin, Cecilia Persson, and Maxim Mayzel share equal authorship

Grant sponsors: Kempe Foundation, Swedstruct, Wallenberg Foundation Swedish Research Council; Grant number: K2013-67X-13001-15-3.

*Correspondence to: E. Sauer-Eriksson, Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden. E-mail: elisabeth.sauer-eriksson@umu.se or G Karlsson, Swedish NMR Centre at the University of Gothenburg, SE-40530 Gothenburg, Sweden. E-mail: goran.karlsson@nmr.gu.se

Introduction

Cholera is an infection in the small intestine caused by the motile Gram-negative bacterium *Vibrio cholerae*.[1] During an infection, *V. cholerae* pathogenicity genes code for virulence factors that are directly or indirectly involved in the virulence of the bacteria. These factors include proteases that attack the target cells by breaking down tissue barriers and cellular matrix components.[2–4] One of these proteases is the secreted metalloprotease PrtV that exhibits a very potent cytotoxic effect.[5,6]

PrtV belongs to the M6 peptidase family and is natively expressed as a 102 kDa full-length pre-pro-protein (918 amino acids). In addition to the signal sequence, it consists of four domain types: the N-

terminal domain (residues 23–103), the catalytic active M6 domain (residues 106–749), two polycystic kidney disease domains, PKD1 (residues 755–838) and the PKD2 (residues 839–918).[5,6] M6 constitutes the catalytic metalloprotease domain with the characteristic HexxHxxgxxD Zn^{2+}-binding motif.[7] PKD domains are found in a variety of eukaryotic and prokaryotic proteins and consist of relatively short domains of 80–90 amino acids; they are usually found in the extracellular parts of proteins where they are involved in protein–protein or protein–carbohydrate interactions. It has been suggested that calcium binding at the PKD1 domain controls domain linker flexibility, and plays a regulatory role in the autoproteolytic activity of the 81 kDa pro-protein.[8] Sequence analysis shows that the N-terminal domain is present in many proteases from gammaproteobacteria; however, its specific function is unidentified. Atomic resolution structures of the N-terminal domain are needed to elucidate how the N-terminal domain contributes to PrtV function. In this study, we present the NMR structure of the N-terminal domain.

Results

NMR spectroscopy of the N-terminal domain

The solution structure of the N-terminal domain was solved at the Swedish NMR Centre (www.nmr.gu.se). The ^{1}H–^{15}N heteronuclear single quantum correlation (HSQC) NMR spectrum of the PrtV 83 amino-acid N-terminal domain exhibited characteristics of a partially folded protein. The spectrum was well dispersed, but exhibited a signal clustering characteristic of a disordered protein. The total number of observed signals was fewer than expected from the sequence, either due to overlapping or broadening beyond detection. Similarly, peak shapes showed a distribution from narrow to broad, and intensities varied considerably.

NMR-derived structure of the N-terminal domain

Structure statistics for the 20 lowest energy conformers are summarized in Table I. The first 8 residues from the N-terminal end and the last 32 residues from the C-terminal end of the domain exhibited random-coil chemical shifts with non-uniform intensities suggesting that these regions are disordered. Remaining residues (residues 31–70) exhibited highly dispersed chemical shifts with uniform intensities that indicated a folded core structure [Fig. 1(A)]. A ribbon structure, representative of the final ensemble of NMR-derived structures of the N-terminal domain is shown in Figure 1(B). The main chain structure contains two alpha helices: α1 (residues Val31-Gly45) and α2 (residues Asp53-Ser70). The first helix α1 is bent and the first turn has a 3_{10}-helical geometry. The two helices are connected by a well-defined loop region comprising residues Gln46-Ser52. The C-terminal part of α1 (residues Leu37-Gly45) packs against the C-terminal part of α2 at an angle of about 90°. This is in accordance with the general rule for packing interactions between α helices, that is, the ridges on one helix fit into the grooves of the other, and vice versa. In this area, the two helices are held together by hydrophobic interactions between conserved amino acids [Fig. 2(A,B)]. A space-filling representation of the N-terminal domain reveals a cluster of hydrophobic and basic residues at the N- and C-terminal ends of the folded structure, whereas the loop region of the motif is more negatively charged [Fig. 1(C)].

Table I. *Structural Statistics for the Ensemble of 20 Calculated Structures of PrtV N-Terminal Domain*

Restraints for structure calculation	
NOE restraints (total)	656
Intra ($\|i\text{-}j\|=0$)	222
Medium ($1\leq\|i-j\|\leq4$)	385
Long ($\|i-j\|>4$)	49
Hydrogen bond restraints	52
Dihedral angle constraints	48
Cyana target function	2.2
Ramachandran plot (residues 31–70)	
Most favored regions	95.8%
Additionally allowed regions	4.2%
Generously allowed regions	0.0%
Disallowed regions	0.0%
RMSD of atom position (residues 31–70)	
All backbone atoms:	
α-helical regions	0.7 ± 0.25 Å
All heavy atoms:	
α-helical regions	1.4 ± 0.25 Å

Discussion

The secreted metalloprotease PrtV is a potent virulence protein of *V. cholerae* causing immediate cytotoxic effects during infection.[6] The function of its N-terminal domain is unknown, but its sequence is highly conserved among members of the genus *Vibrio*, and it also exists in proteases from other gammaproteobacteria like Moriella and Marinomonas [Fig. 2(A)]. The solution structure of the N-terminal domain comprises a coiled-coil of two well-defined helices that correspond to two of the three predicted secondary structural elements [Fig. 2(B)]. Interestingly, the predicted third helix that is flexible in the NMR structure comprises more conserved residues than the folded helix α2. Structural similarity searches using the DALI server[12] identified no real hits. The top DALI hits with *z*-scores below 3 and sequence identity less than 10% comprised high mobility group (HMG) proteins [e.g., pdb code 3nm9 and 1j3c, Fig. 2(C)]. Interestingly, these DNA-binding proteins have extended C-terminal helices rich in lysines and arginines that interact with the DNA minor groove. A helical wheel of residues 62–95 in the N-terminal domain shows that the predicted C-terminal helix would be

amphiphilic with aligned lysine residues forming a highly positively charged side [Fig. 2(D)]. This led us to test the DNA-binding ability of the N-terminal domain; however, the domain does not seem to bind unspecifically to the minor groove DNA [Fig. 2(E)]. Possibly the conserved and positively charged lysine residues provide a region involved in membrane association (e.g., anchored to the membrane via lysine snorkeling)—this hypothesis needs to be tested.

Methods

Expression and purification of N-terminal domain

The cloning, expression, and purification of the PrtV N domain construct (residues 23–103) is described elsewhere.[13] In brief, the N-terminal domain was cloned into the pET1a-TrxA vector[14] and overexpressed in *E. coli* Bl21 (DE3) pLysS (Novagen) cells. The protein was purified by affinity chromatography on a Ni-NTA agarose column (Qiagen), cation exchange chromatography on a MonoS 5/5 column (GE Healthcare), followed by a Superdex 200-16/60 size exclusion column (GE Healthcare). The protein is very soluble and can be concentrated to > 100 mg/mL. Pure fractions of the PrtV N domain were pooled and concentrated to 20 mg/mL in 20 m*M* Tris pH 7.5 and 100 m*M* NaCl. From 1 L media ~10 mg of protein was obtained.

Electrophoretic mobility shift assay

DNA–protein interaction was examined by a gel retardation assay using purified N-terminal PrtV protein. As substrates for DNA binding, 400 bp PCR products containing the promoter regions of *vca0107* and *vc0017* genes of *V. cholerae* wild type strain A1552 were used. A known H-NS binding promoter and purified H-NS protein (4 and 12 μ*M*) were used as the positive control in the electrophoretic mobility shift assay. The reaction buffer containing 25 m*M* 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.3), 0.1 m*M* ethylenediaminetetraacetic acid (EDTA), 5 m*M* dithiothreitol (DTT), 10% glycerol, 50 m*M* KCl, and 0.1 μg of poly(dIdC) non-specific competitor DNA was used for the assay. Purified PrtV protein (4 and 12 μ*M*) or H-NS protein (4 and 12 μ*M*) were incubated with 100 ng of PCR products for 20 min at 20°C. Samples were loaded onto a 6% Tris-glycine gel and DNA was visualized by ethidium bromide staining.

Isotope labeling

Expression of PrtV N-terminal domain for NMR analysis was done in 1 L of M9 minimal media supplemented with $^{15}NH_4Cl$ (0.5 g/L media) as the sole nitrogen source, ^{13}C-glucose (1 g/L media) as the sole carbon source, and trace elements as specified.[13] Labeled PrtV N-terminal domain expression was induced and protein purified as described for the unlabeled protein.

This figure also includes an iMolecules 3D interactive version that can be accessed via the link at the bottom of this figure's caption.

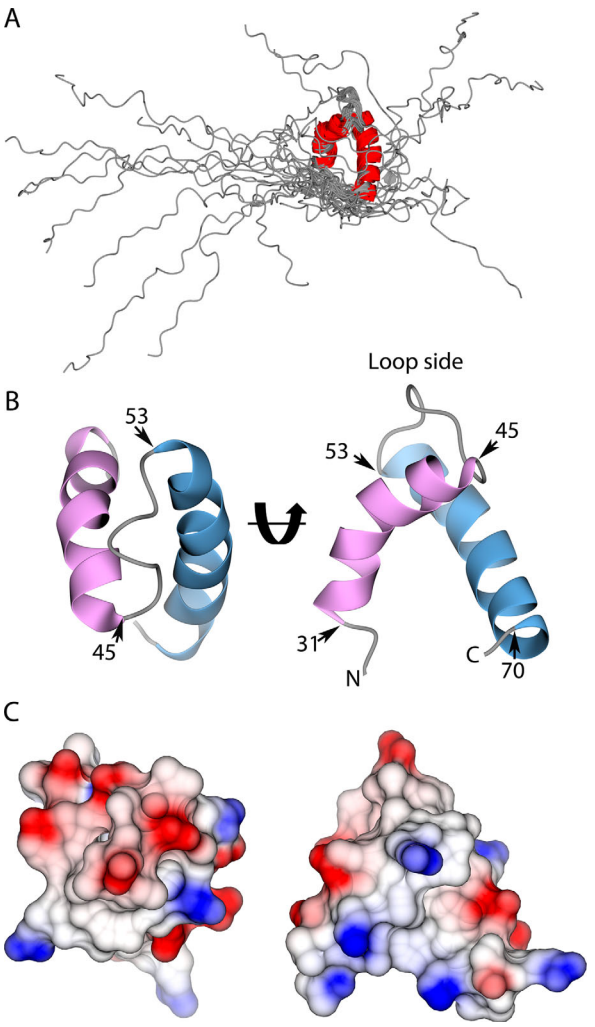

Figure 1. NMR-derived structure of the PrtV N-terminal domain. A: The 20 lowest-energy conformers of the N domain with their backbones (residues 31–70) superimposed. The first 8 N-terminal residues and the last 32 C-terminal residues are not folded in the structure. B: Ribbon representation of the folded N-terminal domain structure closest to the average structure. The structure is shown in two views rotated by 90°. The pink helix includes residues 31–45 (α1) and the blue residues 53–70 (α2). C: Accessibility and electrostatic potential of surfaces of the structures are presented in (B). The loop side is more negatively charged, whereas the N- and C-terminal side surfaces are hydrophobic/positively charged. An interactive view is available in the electronic version of the article.

Structure calculation

Samples of the recombinant PrtV N-terminal domain (0.7 m*M*) were prepared in 90%/10% H_2O/D_2O or

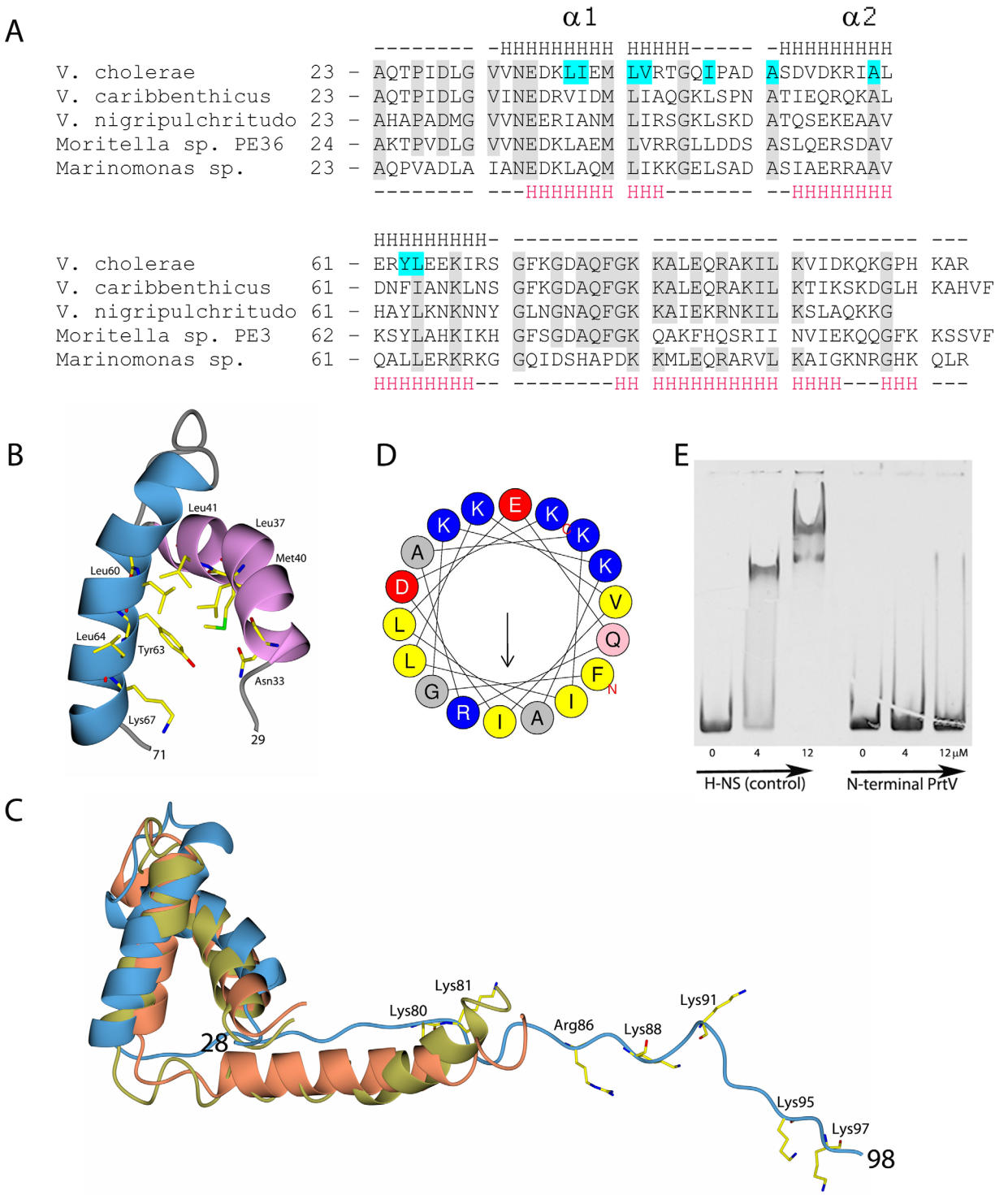

Figure 2. A: Sequence alignment (Blast[9]) of the N-terminal domain of PrtV from *V. cholerae*. AET29433.1; *V. caribbenthicus*, WP_009601447.1; *V. nigripulchritudo*, WP_004405586.1; *Moritella sp*. PE36, WP_006031551.1; *Marinomonas sp.,* WP_009832723.1. The sequence identity to *V. cholera* is 59% (for 83 amino acids), 47% (for 76 amino acids), 54% (for 83 amino acids), 43% (for 81 amino acids), respectively. The secondary structure elements from the current structure are shown in black (H, α-helices). Secondary structure elements predicted for *V. cholerae* with jpred3[10] are shown in red (H, α-helices). Conserved residues that build the hydrophobic core in the *V. cholerae* structure are highlighted in cyan. B: Packing of the N-terminal domain hydrophobic core. C: Superposition of N-terminal model (residues 28–98, light-blue) with HMG proteins from *Drosophila melanogaster* (orange; pdb:3nm9) and *Sus scrofa* (gold; pdb:1j3c). Conserved, positively charged residues in the N-terminal domain predicted to form an α-helix are labeled. D: Helical wheel over residues 78–95 calculated with Heliqust.[11] E: Gel mobility shift assays indicated no unspecific binding of the N-terminal domain to a 400 bp PCR product containing the promoter regions of two genes of a *V. cholerae* wild type strain. As a control, binding of H-NS protein to the DNA is shown.

100% D_2O with 20 mM sodium phosphate (pH 7.0). Sequential assignment was obtained using incremental non-uniform sampling combined with targeted acquisition on a 600 MHz Varian Inova NMR spectrometer with a cold probe, essentially as earlier described.[15] Side chain assignment was obtained from 3D H(CCO)NH and (H)C(CCO)NH experiments. Structure information was obtained from 3D ^{15}N-NOESY-HSQC, 3D ^{13}C-NOESY-HSQC, and 2D NOESY experiments obtained at 800 MHz and 900 MHz Avance III HD spectrometers equipped with 5 mm TCI cryoprobes. ^{15}N-NOE relaxation data were obtained as previously described. All NMR experiments were performed at 10°C. NMR data were processed with NMRPipe[16] and analyzed using the CCPN package.[17]

Structures of PrtV N-terminal domain were calculated with CYANA.[18] NOE distance restraints combined with dihedral restraints and hydrogen bond distance constraints in the alpha helices obtained from the CSI-method[19] were used to calculate ensemble model structures. The 20 lowest energy structures were selected and analyzed.

Molecular graphics were produced using CCP4mg.[20] The NMR-derived ensemble structures of the N-terminal domain were deposited in the Protein Data Bank, with accession code 5abk. The assigned chemical shifts have been deposited within the BMRB database (accession number 25745).

Acknowledgments

Authors thank Drs. Karina Persson, Uwe H. Sauer, Bernt Eric Uhlin, and Tobias Hainzl for valuable ideas and discussions, and Monica Persson for technical assistance. The work was performed at the Umeå Centre for Microbial Research (UCMR).

References

1. Harris JB, LaRocque RC, Qadri F, Ryan ET, Calderwood SB (2012) Cholera. Lancet 379:2466–2476.
2. Miyoshi S, Shinoda S (2000) Microbial metalloproteases and pathogenesis. Microbes Infect 2:91–98.
3. Hase CC, Finkelstein RA (1993) Bacterial extracellular zinc-containing metalloproteases. Microbiol Rev 57: 823–837.
4. Shinoda S, Miyoshi S (2011) Proteases produced by vibrios. Biocontrol Sci 16:1–11.
5. Vaitkevicius K, Lindmark B, Ou G, Song T, Toma C, Iwanaga M, Zhu J, Andersson A, Hammarstrom ML, Tuck S, Wai SN (2006) A Vibrio cholerae protease needed for killing of Caenorhabditis elegans has a role in protection from natural predator grazing. Proc Natl Acad Sci USA 103:9280–9285.
6. Vaitkevicius K, Rompikuntal PK, Lindmark B, Vaitkevicius R, Song T, Wai SN (2008) The metalloprotease PrtV from Vibrio cholerae. FEBS J 275:3167–3177.
7. Kurisu G, Kinoshita T, Sugimoto A, Nagara A, Kai Y, Kasai N, Harada S (1997) Structure of the zinc endoprotease from Streptomyces caespitosus. J Biochem 121:304–308.
8. Edwin A, Rompikuntal P, Bjorn E, Stier G, Wai SN, Sauer-Eriksson AE (2013) Calcium binding by the PKD1 domain regulates interdomain flexibility in Vibrio cholerae metalloprotease PrtV. FEBS Open Biol 3:263–270.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410.
10. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36:W197–W201.
11. Gautier R, Douguet D, Antonny B, Drin G (2008) HELIQUEST: a web server to screen sequences with specific alpha-helical properties. Bioinformatics 24: 2101–2102.
12. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. Bioinformatics 24:2780–2781.
13. Edwin A, Grundstrom C, Wai SN, Ohman A, Stier G, Sauer-Eriksson AE (2014) Domain isolation, expression, purification and proteolytic activity of the metalloprotease PrtV from Vibrio cholerae. Protein Express Purif 96:39–47.
14. Bogomolovas J, Simon B, Sattler M, Stier G (2009) Screening of fusion partners for high yield expression and purification of bioactive viscotoxins. Protein Express Purif 64:16–23.
15. Isaksson L, Mayzel M, Saline M, Pedersen A, Rosenlow J, Brutscher B, Karlsson BG, Orekhov VY (2013) Highly efficient NMR assignment of intrinsically disordered proteins: application to B- and T cell receptor domains. PloS One 8:e62947.
16. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293.
17. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59: 687–696.
18. Guntert P, Mumenthaler C, Wuthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298.
19. Wishart DS, Sykes BD (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 4:171–180.
20. McNicholas S, Potterton E, Wilson KS, Noble ME (2011) Presenting your structures: the CCP4mg molecular-graphics software. Acta Cryst D67:386–394.