NIH Public Access
Author Manuscript
Proteins. Author manuscript; available in PMC 2011 August 15.

Published in final edited form as:
Proteins. 2010 August 15; 78(11): 2563–2568. doi:10.1002/prot.22760.

Solution NMR Structure of Lin0431 Protein from *Listeria innocua* Reveals High Structural Similarity with Domain II of Bacterial Transcription Antitermination Protein NusG

Yuefeng Tang[1], **Rong Xiao**[1], **Colleen Ciccosanti**[1], **Haleema Janjua**[1], **Dong Yup Lee**[1], **John K. Everett**[1], **G.V.T. Swapna**[1], **Thomas B. Acton**[1], **Burkhard Rost**[2], and **Gaetano T. Montelione**[1,3,*]

[1]Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, NJ, 08854, U.S.A., and Northeast Structural Genomics Consortium

[2]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, 10032, U.S.A, and Northeast Structural Genomics Consortium

[3]Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, NJ, 08854, U.S.A., and Northeast Structural Genomics Consortium

Abstract

Lin0431 protein from *Listeria innocua* (UniProtKB/TrEMBL ID Q92EM7/Q92EM7_LISIN) was selected as a target of the Northeast Structural Genomics Consortium (target ID: LkR112). Here, we present the high-quality NMR solution structure of this protein which is the first representative for a member of DUF1312 domain family. Lin0431 protein exhibits a β-sandwich topology. Four anti-parallel β-strands form one face of the sandwich and the other three anti-parallel β-strands together with a short α-helix form the other face of the sandwich. Structure alignment by Dali reveals an unexpected structural similarity with domain II of NusG from *Aquifex aeolicus*. Analyses of the electrostatic protein surface potential and searches for protein surface cavities reveal the conserved basic charged surface cavities of both the Lin0431 and domain II of *Aae*NusG, suggesting they may bind the negatively charged nucleic acids and/or and other binding partners. The high structural similarity and similar surface features, despite the lack of recognizable sequence similarity, between Lin0431 and *Aae*NusG domain II suggest that the domain II of NusG and DUF1312 domains have a homologous relationship and may share similar biochemical functions.

Keywords

structural genomics; Lin0431; NusG

INTRODUCTION

Full length Lin0431 is a 140-residue protein from *Listeria innocua* (UniProtKB/TrEMBL ID Q92EM7/Q92EM7_LISIN) that has no significant sequence similarity with any protein with known structure. It was selected as by the Protein Structure Initiative as a targeted domain family of unknown structure, and assigned to the Northeast Structural Genomics (NESG)

Correspondence should be addressed to G.T.M (guy@cabm.rutgers.edu).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

consortium (NESG Target ID LkR112). Lin0431 belongs to Pfam1 protein family DUF1312 (PF07009), which contains more than 224 members, most of around 120 residues in length. This family has no functional annotation (i.e., DUF – domains of unknown function), though it is represented in at least 71 bacterial species, including metagenomic bacterial species sequenced in human gut metagenomic projects.2,3 In this Structure Note, we report the high-quality solution NMR structure of Lin0431 protein form *L. innocua*. The structure of Lin0431 protein presented here is the first reported three-dimensional (3D) NMR structure for a member of DUF1312 domain family.

N-utilization substance G (NusG) is a highly conserved bacterial protein that plays key roles in transcription elongation, transcription antitermination, and translational regulation.4–7 NusG has been proposed to exert multiple and sometimes opposite effects on transcription elongation by interaction with elongation complexes of RNA polymerase, DNA and RNA.8–11 Some NusG proteins, such as *Aquifex aeolicus* NusG (*Aae*NusG) are comprised of three distinct domains (DI, DII and DIII). The N-terminal domain DI and C-terminal domain DIII, which contain NGN and KOW motifs, respectively, are highly conserved in all NusG proteins. Domain II (DII), on the other hand, is a unique insertion of ~80 residues in the central region of these three-domain NusG proteins. It has been proposed that DI and DIII are involved in different regulatory pathways through different combination of interactions with RNA polymerase, termination factor ρ, and nucleic acid.12–14 However, the functional role and the interaction partners of DII domains, are not yet known. Structure alignment by Dali15 reveals unanticipated high structural similarity between Lin0431 and *Aae*NusG DII, suggesting a previously unrecognized homologous relationship between these two domain families.

MATERIALS AND METHODS

Uniformly ^{13}C, ^{15}N- and 5%- ^{13}C, *U*- ^{15}N-enriched Lin0431 protein from *Listeria innocua* (UniProtKB/TrEMBL ID Q92EM7/Q92EM7_LISIN) were cloned, expressed, and purified following standard protocols of the NESG consortium.16 Briefly, the truncated Q92EM7 (residues 36–140) gene from pET was cloned into a pET 21–23C (Novagen) derivative, yielding the plasmid LkR112-36-127-21.11. The resulting construct codes for the native sequence of residues 36–140 of Lin0431protein, with a additional eight C-terminal nonnative residues (LEHHHHHH). *Escherichia coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with LkR112-36-127-21.11, and cultured in MJ9 minimal medium17 containing $(^{15}NH_4)_2SO_4$ and *U*- ^{13}C-glucose as the sole nitrogen and carbon sources. *U*- ^{13}C, ^{15}N Lin0431 protein was purified in a two step protocol consisting of IMAC (HisTrap HP) and gel filtration (HiLoad 16/60 Sephadex 75, GE Healthcare) chromatography. Briefly, the cells were suspended in binding buffer (50 mM Tris-HCl, pH 7.5, 0.3 M NaCl, 40 mM imidazole, 1 × protease inhibitor cocktail, 1 mM Tris (2-carboxyethyl) phosphine (TCEP)), and disrupted on ice by sonication. The supernatant of the cell lysate was collected by centrifugation at 26,000 × g for 40 min at 4°C and purified by an AKTAxpress system using HisTrap HP affinity column followed by Superdex 75 gel filtration column (GE Healthcare, USA). The protein was eluted from the HisTrap column using Elution Buffer (50 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, 1 mM TCEP, pH 7.5) at 4 ml / min. The proteins are collected into internal storage loops, and then injected onto the Superdex 75 gel filtration column equilibrated with NMR buffer (20 mM MES, 200 mM NaCl, 5 mM $CaCl_2$, 10 mM DTT, 0.02% NaN_3, pH 6.5). The final yield of pure protein (>97% homogeneity by SDS-PAGE; 11.5 kDa by MALDI-TOF mass spectrometry) was ~65 mg/L. In addition, a *U*- ^{15}N and 5% ^{13}C –enriched sample was produced for use in determining stereo-specific assignment of isopropyl methyl groups. Protein samples of *U*- ^{13}C, ^{15}N and *U*- ^{15}N, 5% ^{13}C –enriched Lin0431 protein for NMR spectroscopy were prepared at 0.9 mM protein concentration in 95% H_2O/5% ^2H_2O solution

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

containing 20 mM MES, 200 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3 at pH 6.5. Static light scattering (Suppl. Fig. S3) and ^{15}N T_1/T_2 measurements (Suppl. Fig. S4) demonstrate that the protein sample is monomeric in solution under the conditions used in the NMR studies.

All NMR data were collected at 25 °C on a Bruker AVANCE 800 MHz NMR spectrometer equipped with a cryogenic probe. NMR spectra were processed with NMRPipe and NMRDraw software package.18 The program SPARKY19 was used for data visualization and analysis. Chemical shifts of proton were referenced to internal DSS; ^{13}C and ^{15}N chemical shifts were referenced indirectly based on the proton referencing. Backbone and side chain $^{13}C^{\beta}$ resonance assignments were determined using conventional triple resonance NMR experiments including HNCO, HNCACB and HNcoCACB. AutoAssign20,21 software was used for semi-automated analysis of backbone and side chain $^{13}C^{\beta}$ resonance assignments. Complete side-chain spin system identification was accomplished by using aliphatic HCCH-COSY and CCH-TOCSY. 2D high resolution [^{13}C, ^{1}H]-HSQC spectra were recorded for the 5% biosynthetically directed fractionally ^{13}C-enriched sample in order to obtain stereo-specific assignments for isopropyl groups of valines and leucines. The chemical shift assignments were validated using Assignment Validation Suite (AVS) software22 and deposited in the BioMagResDB (BMRB accession numbers 16563). ^{1}H-^{1}H upper distance constraints for structure calculations were obtained from simultaneous 3D $^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}$ –resolved NOESY spectrum, with mixing time of 100 ms. Backbone dihedral angle constraints were derived from chemical shifts using the program TALOS23 and were used only for residues with TALOS confidence scores of 10. The structure was calculated using CYANA 3.024,25 and refined by restrained molecular dynamics in explicit water using CNS 1.2.26,27 The final coordinates for ensemble of 20 models out of 100 calculated were deposited into the Protein Data Bank (PDB ID, 2kpp). Structural statistics and global structure quality factors, including Verify3D,28 ProsaII,29 PROCHECK,30 and MolProbity31 raw and statistical Z-scores, were computed using the PSVS 1.3 software package.32 Values for the global goodness-of-fit of the final structure ensembles with the NOESY peak list and chemical shift data were determined using the RPF analysis program.33

RESULTS AND DISCUSSION

A multiple sequence alignment of *L. innocua* Lin0431 protein with representative subset of DUF1312 protein domains is shown in Fig. 1A. As construct design bioinformatics analysis indicated intrinsic disorder in the N-terminal 35 residues of Lin0431, the construct used in this study includes only residues 36 to 140. The structure consists of seven β-strands (β_1, residues 42–48; β_2, 51–57; β_3, 64–72; β_4, 75–82; β_5, 86–91; β_6, 110–113 and β_7, 118–124) and a short helix (α_1, residues 97–101) arranged into two sandwiched anti-parallel β-sheets. All secondary elements are locally and globally well-defined. The first 4 N-terminal residues of our construct, as well as residues 125 to 140 plus the 8-residue purification tag (LEHHHHHH) at the C-terminal end, are disordered, consistent with both disorder prediction and heteronuclear ^{15}N NOE data (Suppl. Fig. S2). Stereoimages of the superimposed final ensemble and ribbon diagram of a representative structure are shown in Figs.1B and 1C. The structure has excellent quality assessment scores as listed in Table 1.

In the 3D structure, four anti-parallel β-strands (β1, β2, β6, β7) form one face of the sandwich, while the rest three β-strands (β3, β4, β5) together with a short α-helix (α_1) form the other face of the sandwich. The core of the structure is formed by tightly packed hydrophobic side chains. The patterns of hydrogen bonds and NOEs reveal three β-bulges at positions I53 – R54, G71 – K72 and K89 – E90. The core of the structure is stabilized by strongly conserved hydrophobic residues (A44, I46, I56, L58, M79, V81, I86, I88, I111,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

C113, H116, V118, V120) buried between two sandwiched β-sheets. ConSurf34 analysis (Fig. 1D), generated using the 3D structure of Lin0431determined here along with the seed (a small set of representative members of the family) protein sequences in the DUF1312 pfam family, reveal that these conserved surface residues located between two sandwiched β-sheets are well conserved across the domain family, and may be responsible for their biological functions.

A search of the PDB database for structurally similar protein using DALI15 identified significant hits with DII of NusG from *Aquifex aeolicus*, which has only 20 % sequence identity (Fig. 1A) in the structurally similar region to Lin0431 protein (Fig. 1 E, PDB ID 1m1g with Dali Z score 11.1, RMSD 1.9). The backbone fold of Lin0431 protein and DII of *Aae*NusG are almost identical (Figure 1 E). One obvious difference between Lin0431 and DII of *Aae*NusG is that there is a disulfide bridge present in DII of *Aae*NusG formed between C104 and C119; however, the corresponding residues in Lin0431 are V97 and C113, which are very close in space, can not form disulfide bond. DI and DIII domains of *Aae*NusG are highly conserved through the NusG protein family, and it has been shown that these two domains influence transcription termination and anti-termination by interaction with RNA polymerase, termination factor ρ and nucleic acid.12،13 However, little is known about the functional role of domain DII of *Aae*NusG. The similar basic charged surface distributions of both the Lin0431 (Fig. 1G) and DII of *Aae*NusG (Fig. 1F) suggest they may bind negatively charged nucleic acids and/or and other anionic binding partners. The high structural similarity and similar conserved electrostatic surface features, despite having only 20 % sequence similarity, between Lin0431 and *Aae*NusG DII suggest that the NusG DII and DUF1312 domain families may have diverged from common evolutionary ancestral proteins, and may have similar biochemical functions.

A search of protein surface cavities using ProFunc35 server reveals several cavities on the surface of Lin0431. One of the biggest surface cavities (Fig. 1I) is located at one face of the sandwiched anti-parallel β-sheets, and is rich in basic and hydrophobic residues. This cavity is also present on this conserved surface of *Aae*NusG DII (Fig. 1H). The electrostatic surface potential calculated by Pymol36 indicates a positively charged surface within this cavity. Since protein functional sites are usually located in such large, conserved surface cavities, we suspect that this cavity is important for the protein function; considering its strongly conserved basic properties, it may serve as a nucleic acid binding site.

The NMR structure of Lin0431 protein from *L. innocua* described here is the first 3D NMR structure determined for a DUF1312 Pfam domain family member. Its high structural similarity with the domain II (DII) of NusG from *A. aeolicus*, and similar surface features that appear to be conserved across both the DUF1312 and DII of NusG domain families, suggests a previously unrecognized evolutionary relationship between two protein domain families. Interestingly, five full-length proteins that include a DUF1312 domain have multiple domain architectures that include DUF1312, NusG, and KOW domains, suggesting a possible role for these proteins and the DUF1312 domain in transcription/translation regulating functions like those of NusG proteins. Further structural and functional studies will be required to determine the nucleic acid and/or protein binding partners of the DUF1312 domain family, and the role of such interactions in the biology of this novel domain family.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Acknowledgments

We thank Gaohua Liu and Yuanpeng Janet Huang for helpful discussions. This work was supported by grant U54-GM074958 from the National Institute of General Medical Sciences Protein Structure Initiative.

References

1. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A. Pfam: clans, web tools and services. Nucleic Acids Res. 2006; 34(Database issue):D247–D251. [PubMed: 16381856]
2. Sleator RD, Shortall C, Hill C. Metagenomics. Lett Appl Microbiol. 2008; 47(5):361–366. [PubMed: 19146522]
3. Jones BV, Sun F, Marchesi JR. Comparative metagenomic analysis of plasmid encoded functions in the human gut microbiome. BMC Genomics. 11:46. [PubMed: 20085629]
4. Burova E, Hung SC, Sagitov V, Stitt BL, Gottesman ME. *Escherichia coli* NusG protein stimulates transcription elongation rates in *vivo* and in *vitro*. J Bacteriol. 1995; 177(5):1388–1392. [PubMed: 7868616]
5. Greenblatt J, Mah TF, Legault P, Mogridge J, Li J, Kay LE. Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. Cold Spring Harb Symp Quant Biol. 1998; 63:327–336. [PubMed: 10384297]
6. Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. Science. 2008; 320(5878):935–938. [PubMed: 18487194]
7. Zellars M, Squires CL. Antiterminator-dependent modulation of transcription elongation rates by NusB and NusG. Mol Microbiol. 1999; 32(6):1296–1304. [PubMed: 10383769]
8. Yakhnin AV, Yakhnin H, Babitzke P. Function of the Bacillus subtilis transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the trp leader. Proc Natl Acad Sci U S A. 2008; 105(42):16131–16136. [PubMed: 18852477]
9. Burns CM, Richardson LV, Richardson JP. Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in *Escherichia coli*. J Mol Biol. 1998; 278(2):307–316. [PubMed: 9571053]
10. Zhou Y, Filter JJ, Court DL, Gottesman ME, Friedman DI. Requirement for NusG for transcription antitermination in *vivo* by the λ N protein. J Bacteriol. 2002; 184(12):3416–3418. [PubMed: 12029062]
11. Li J, Mason SW, Greenblatt J. Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. Genes Dev. 1993; 7(1):161–172. [PubMed: 8422985]
12. Steiner T, Kaiser JT, Marinkovic S, Huber R, Wahl MC. Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. EMBO J. 2002; 21(17):4641–4653. [PubMed: 12198166]
13. Knowlton JR, Bubunenko M, Andrykovitch M, Guo W, Routzahn KM, Waugh DS, Court DL, Ji X. A spring-loaded state of NusG in its functional cycle is suggested by X-ray crystallography and supported by site-directed mutants. Biochemistry. 2003; 42(8):2275–2281. [PubMed: 12600194]
14. Mooney RA, Schweimer K, Rosch P, Gottesman M, Landick R. Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. J Mol Biol. 2009; 391(2):341–358. [PubMed: 19500594]
15. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci. 1995; 20(11):478–480. [PubMed: 8578593]
16. Acton TB, Gunsalus KC, Xiao R, Ma LC, Aramini J, Baran MC, Chiang YW, Climent T, Cooper B, Denissova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Rajan PK, Shastry R, Shih LY, Swapna GV, Wilson M, Wu M, Gerstein M, Inouye M, Hunt JF, Montelione GT. Robotic cloning and Protein Production Platform of the Northeast Structural Genomics Consortium. Methods Enzymol. 2005; 394:210–243. [PubMed: 15808222]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

17. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione GT, Nilsson B. High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in *Escherichia coli*. Journal of Biomolecular Nmr. 1996; 7(2):131–141. [PubMed: 8616269]
18. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6(3):277–293. [PubMed: 8520220]
19. Goddard TD, Kneller DG. SPARKY 3 NMR analysis program. 2008.
20. Moseley HN, Monleon D, Montelione GT. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Methods Enzymol. 2001; 339:91–108. [PubMed: 11462827]
21. Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C, Powers R, Montelione GT. Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol. 1997; 269(4):592–610. [PubMed: 9217263]
22. Moseley HN, Sahota G, Montelione GT. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR. 2004; 28(4):341–355. [PubMed: 14872126]
23. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR. 1999; 13(3):289–302. [PubMed: 10212987]
24. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol. 1997; 273(1):283–298. [PubMed: 9367762]
25. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR. 2002; 24(3):171–189. [PubMed: 12522306]
26. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. 1998; 54(Pt 5):905–921. [PubMed: 9757107]
27. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. Proteins. 2003; 50(3):496–506. [PubMed: 12557191]
28. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature. 1992; 356(6364):83–85. [PubMed: 1538787]
29. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins. 1993; 17(4): 355–362. [PubMed: 8108378]
30. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996; 8(4):477–486. [PubMed: 9008363]
31. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins. 2003; 50(3):437–450. [PubMed: 12557186]
32. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins. 2007; 66(4):778–795. [PubMed: 17186527]
33. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc. 2005; 127(6):1665–1674. [PubMed: 15701001]
34. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. 2005; 33(Web Server issue):W299–W302. [PubMed: 15980475]
35. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res. 2005; 33(Web Server issue):W89–W93. [PubMed: 15980588]
36. DeLano, WL. The pymol manual. San Carlos, CA: DeLano Scientific; 2002.
37. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. 1998; 11(9):739–747. [PubMed: 9796821]

38. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph. 1996; 14(1):51–55. 29–32. [PubMed: 8744573]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

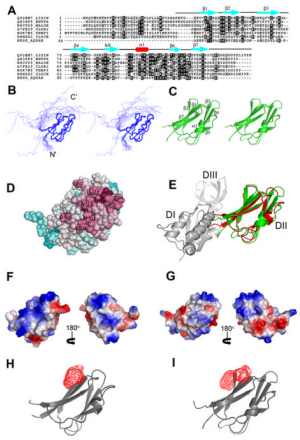

Fig. 1.

(A) Multiple sequence alignment of a representative subset from the entire DUF1312 protein domain family and DII of *Aae* NusG, generated using Clustal X. The alignment includes Lin0431 protein from *Listeria innocua*, plus representatives from *Enterococcus faecalis*, *Halothermothrix orenii*, *Clostridium botulinum*, *Thermoanaerobacter pseudethanolicus*,*Clostridium cellulolyticum* and DII of *Aae* NusG indicated by their Swiss-Prot IDs. Amino acid residues identical or similar in 60% of the entire family are highlighted in black and gray, respectively; conserved residues were colored using the BOXSHADE server. The secondary structural elements found in our solution NMR structure (PDB ID: 2kpp) are illustrated above the alignment. (B) Stereoview of an ensemble of 20 CNS refined structures superimposed on backbone atoms of the regular secondary structure elements for minimal root-mean-square deviation (RMSD). (C) Stereoview of the ribbon drawing of a representative conformer (lowest CNS energy) for the ordered region (residues 39 to 127) of the final solution NMR structure of Lin0431 protein. The secondary elements are labeled. (D) ConSurf34 image of the same representative conformer as shown in (C) based on the multiple sequence alignment of the seed sequences of the DUF1312 protein domain family. Residue coloring reflects the degree of residue conservation with magenta corresponding to highest conservation and cyan to the highest variability. (E) Structure comparison of Lin0431 protein from *L. innocua* (green) and NusG from *A. aeolicus* (PDB ID: 1m1g). Three domains are labeled with DII shown in red and DI and DIII shown in gray. Structure superposition is done using CE server37. (F) Electrostatic potential surface diagrams of the domain II of protein NusG from *A. aeolicus* (residue 50 – 133) and (G) Lin0431 protein from *L. innocua*. For clarity, only the structured residues are shown (residue 39 to 127). The structure shown on the left has the same orientation as in (B–I), while the one on the right is rotated by 180° about the vertical axis. Surface colors represent the electrostatic potential. (H) the conserved basic cavities of the domain II of protein NusG from *A. aeolicus* (residue 50 – 133) and (I) Lin0431 protein from *L. innocua* (residue 39 to 127). Cavities generated by ProFunc35 are shown in red and ribbons are shown in gray. All figures are prepared by MOLMOL38 or PYMOL.36

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Summary of NMR and structural statistics for Lin0431 protein from *L. innocua*[a]

Completeness of resonance assignments[b]	
Backbone (%)	93.8
Side chain (%)	88.6
Aromatic (%)	100.0
Stereospecific methyl (%)	100.0
Conformational-restricting constraints[c]	
Distance constraints	
Total	2542
Intra-residue (i =j)	485
Sequential (|i – j |= 1)	722
Medium range (1 < |i – j| < 5)	314
Long range (|i – j| ≥ 5)	1021
Distance constraints per residue	23.8
Dihedral angle constraints	94
Hydrogen bond constraints	
Total	56
Long range (|i – j| ≥ 5)	41
Number of constraints per residue	25.2
Number of long range constraints per residue	9.9
Residual constraint violations[c]	
Average number of distance violations per structure	
0.1–0.2 Å	1.2
0.2–0.5 Å	0.1
> 0.5 Å	0
Average RMS distance violation/constraint (Å)	0.01
Maximum distance violation (Å)	0.34
Average number of dihedral angle violations per structure	
1–10°	3.35
>10°	0
Average RMS dihedral angle violation/constraint (degree)	0.38
Maximum dihedral angle violation (degree)	4.90
RMSD from average coordinates (Å)[c,d]	
Backbone atoms	0.4
Heavy atoms	0.7
Ramachandran plot statistics[c,d]	
Most favored region (%)	91.6
Additional allowed regions (%)	8.4
Generously allowed (%)	0.0
Disallowed regions (%)	

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Global quality scores[c]	Raw	Z-score
Verify3D	0.32	−2.25
Prosall	0.44	−0.87
Procheck(phi-psi)[d]	−0.46	−1.49
Procheck(all)[d]	−0.28	−1.66
Molprobity clash	15.01	−1.05
RPF Scores[e]		
Recall	0.979	
Precision	0.951	
F-measure	0.965	
DP-score	0.896	

[a] Structural statistics were computed for the 20 final energy-minimized conformers.

[b] Computed using AVS software 22 from the expected number of peaks, excluding: highly exchangeable protons (N-terminal, Lys, and Arg amino groups, hydroxyls of Ser, Thr, Tyr), carboxyls of Asp and Glu, non-protonated aromatic carbons, and the C-terminal tag.

[c] Calculated using PSVS 1.3 program. Average distance violations were calculated using the sum over r^{-6}.

[d] Residues selected based on second structure element: 42–48, 51–57, 64–72, 75–82, 86–91, 97–101, 110–113, 118–124.

[e] RPF scores 33 reflecting the goodness-of-fit of the final ensemble of structures (including disordered residues) to the NMR data.