Biomol NMR Assign (2009) 3:191–194
DOI 10.1007/s12104-009-9172-6

ARTICLE

^{1}H, ^{13}C, and ^{15}N NMR assignments for the *Bacillus subtilis* yndB START domain

Kelly A. Mercier · Geoffrey A. Mueller · Thomas B. Acton · Rong Xiao · Gaetano T. Montelione · Robert Powers

Received: 10 March 2009 / Accepted: 3 June 2009 / Published online: 24 June 2009
© Springer Science+Business Media B.V. 2009

Abstract The steroidogenic acute regulatory-related lipid transfer (START) domain is found in both eukaryotes and prokaryotes, with putative functions including signal transduction, transcriptional regulation, GTPase activation and thioester hydrolysis. Here we report the near complete ^{1}H, ^{15}N and ^{13}C backbone and side chain NMR resonance assignments for the *Bacillus subtilis* START domain protein yndB.

Keywords *Bacillus subtilis* · Protein yndB · START domain · Lipid binding · NMR · Resonance assignments

Biological context

Bacillus subtilis protein yndB is a 16.1 kDa monomer annotated as a 'conserved hypothetical protein' in major protein databases: SWISS-PROT (YNDB_BACSU, O31806), PIR (C69889) and GENBANK (NP_389656). Yndb is a structural target (NESG_ID SR211) from the Northeast Structural Genomic Consortium (NESG; www.nesg.org).

The solution structures for two related functionally uncharacterized proteins, *Bacillus cereus* protein BC4709 (PDB ID: 1xn6) and *Bacillus halodurans* protein BH1534 (PDB ID: 1xn5), were previously solved by NMR spectroscopy (Liu et al. 2005). The amino-acid sequences for BC4709 and BH1534 are 64 and 57% homologous to yndB, respectively. The BC4709 and BH1534 NMR structures were shown to belong to the steroidogenic acute regulatory-related lipid transfer (START) domain superfamily (CATH: 3.30.530.20). The START domain is found in both eukaryotes and prokaryotes, with putative functions that include signal transduction, transcriptional regulation, GTPase activation and thioester hydrolysis. Currently, there are 16 START domain proteins in the PDB, several of which contain bound lipid ranging in mass from 397.6 to 790.2 Da.

START domains exhibit a range of functions, but share a common fold and ligand-binding mode (Iyer et al. 2001). The START domain superfamily binds a diverse array of ligands, so that the ligands that interact with a specific START domain are a major determinant of the protein's biological function. Sequence modifications contribute to variations in the size, depth and enzymatic activity of the ligand-binding pocket, that has led to an evolutionary divergence in ligand specificity and protein function among the START domain superfamily members. Currently, there is very little information in the literature describing ligand specificity and biological activity for the START domain superfamily. Thus, the experimental structure of *B. subtilis* protein yndB will be an important asset in our efforts to decipher the relationships between structure, function and

K. A. Mercier · R. Powers (✉)
Department of Chemistry, University of Nebraska-Lincoln, 722 Hamilton Hall, Lincoln, NE 68588-0304, USA
e-mail: rpowers3@unl.edu

K. A. Mercier
e-mail: mercierka@niehs.nih.gov

Present Address:
K. A. Mercier · G. A. Mueller
Laboratory of Structural Biology, National Institute of Environmental Health Sciences, 111 T. W. Alexander Drive, Durham, NC 27713, USA

T. B. Acton · R. Xiao · G. T. Montelione
Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Northeast Structural Genomics Consortium, Rutgers University, Piscataway, NJ 08854, USA

ligand specificity within the START domain superfamily. By comparing the three-dimensional structures of available START domain proteins along with the determination of ligand binding profiles, functional classifications and ligand specificity may be inferred for other members of the START domain superfamily. In order to provide a basis for studies of the structure, ligand binding specificity, and function of *B. subtilis* protein yndB, we report the near complete ^{1}H, ^{15}N and ^{13}C backbone and side chain NMR resonance assignments for yndB.

Methods and experiments

Expression and purification of yndB

Uniformly ^{13}C,^{15}N-enriched yndB (152 amino acids) was cloned, expressed and purified following standard protocols used in the NESG consortium (Acton et al. 2005). Briefly, the full length gene yndB or BSU17730 from *Bacillus subtilis* was cloned into a pET21d (Novagen) derivative, yielding the plasmid pSR211-21.2. The resulting yndB open reading frame contains eight nonnative residues at the C terminus (LEHHHHHH) of the protein. *Escherichia coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with pSR211-21.2, and cultured in MJ9 minimal medium containing $(^{15}NH_4)_2SO_4$ and [U-^{13}C]-D-glucose as sole nitrogen and carbon sources (Jansson et al. 1996). Initial growth was carried out at 37°C until the OD_{600} of the culture reached ~0.8 units. The incubation temperature was then decreased to 17°C and protein expression was induced by the addition of IPTG (isopropyl-β-D-thiogalactopyranoside) at a final concentration of 1 mM. Following overnight incubation at 17°C, the cells were harvested by centrifugation and lysed by sonication. [U-^{13}C,^{15}N] yndB was purified in a two-step protocol consisting of Ni-NTA affinity column (Qiagen) and gel filtration column (HiLoad 26/60 Superdex 75 pg, Amersham Biosciences) chromatography. These samples were determined to be largely (>98%) monomeric by analytical gel filtration (20 mM MES, 100 mM NaCl, 5 mM $CaCl_2$, 10 mM DTT, pH 6.5) followed by static light scattering measurements (as described in Acton et al. 2005). The final yield of pure [U-^{13}C,^{15}N] yndB (estimated >97% by SDS–PAGE; 17993.4 Da by MALDI-TOF mass spectrometry) was ~32.4 mg from a 1 l culture.

NMR spectroscopy

Samples of [U-^{13}C,^{15}N] yndB for NMR spectroscopy were prepared at a protein concentration of 1.0 mM in 95% H_2O/5% D_2O solution containing 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3 at pH 6.5 (uncorrected) in a sealed Shigemi tube (Shigemi Inc., Allison Park, PA).

All NMR experiments used for the protein backbone assignments of yndB were collected at 25°C on a five channel 600 MHz Bruker Avance spectrometer equipped with a 5 mm TXI probe. NMR experiments used for the protein side chain resonances were collected at the Pacific Northwest National Laboratory on a four channel 750 MHz Varian INOVA spectrometer equipped with a 5 mm HCN probe and at the National Institute of Environmental Health Sciences on a 600 MHz Varian INOVA spectrometer also equipped with a 5 mm HCN probe. Assignments of the backbone and side chain resonances were obtained from the following spectra: ^{15}N-HSQC, ^{13}C-HSQC, HNCO, HNCA, CBCACONH, CBCANH, HNHA, HBHACONH, C(CO)NH-TOCSY, H(CCO)NH-TOCSY, HCCH-TOCSY, and H(CCH)-COSY (Ferentz and Wagner 2000). The NMR data was processed using NMRpipe (Delaglio et al. 1995) and analyzed using PIPP (Garrett et al. 1991) and NMRViewJ (Johnson 2004).

Assignments and data deposition

Extensive assignments for the backbone and side-chain resonances of yndB were obtained (123/143 HN, 123/144 Cα, 140/168 Hα, 121/132 Cβ, 168/229 Hβ ,88/134 Cγ, 33/75 Cδ, 110/144 C′, excluding the LEHHHHHH purification tag). NMR assignments are based on 3D heteronuclear experiments performed on the ^{13}C/^{15}N-labeled yndB sample using NMRView and NVAssign (Kirby et al. 2004). Figure 1 shows an annotated 2D ^{1}H–^{15}N HSQC spectrum of yndB. The terminal residues were not observed in the ^{1}H–^{15}N HSQC spectrum. Many of the unassigned peaks in the ^{1}H–^{15}N HSQC spectrum are near random-coil chemical shifts and likely residues in the loop regions. Several loops (based on sequence similarity prediction) were difficult to assign because of either missing resonances, degenerate shifts, or a lack on connectivity in the triple-resonance experiments.

The secondary structure determination of yndB was based on characteristic NOE data involving the HN, Hα and Hβ protons from ^{15}N-edited NOESY-HMQC and ^{13}C-edited NOESY-HMQC spectra, $^{3}J_{HN\alpha}$ coupling constants from HNHA, slowly exchanging HN protons and ^{13}Cα and ^{13}Cβ secondary chemical shifts (Marion et al. 1989; Wagner and Wuthrich 1986). The yndB NMR structure is composed of six β strands forming an anti-parallel β-sheet and three α-helices: one long and two short. Residues K37 to M40, H53 to S61 and K106 to E117 yielded limited assignments presumably because of local flexibility. The residues that are unassigned in these loops were: K37, W38, F39, M40, H53, L54, Q55, S56, P57, P60, S61, K106, E107, P108, N109, G113, K114, N116, and E117. Figure 2 shows the

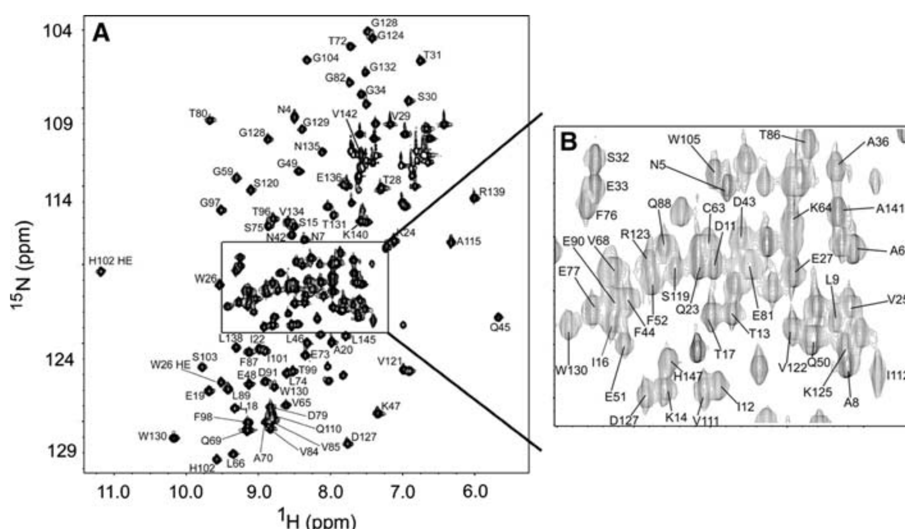

Fig. 1 2D ^{1}H–^{15}N HSQC spectrum (**a**) and zoomed region (**b**) of uniformly ^{13}C, ^{15}N-labeled yndB in 1.0 mM in 95% H_2O/5% D_2O at pH 6.5 (uncorrected). The spectrum was collect on a Bruker Avance 600 MHz spectrometer equipped with a 5 mm TXI probe. The assignments are annotated by the sequence number

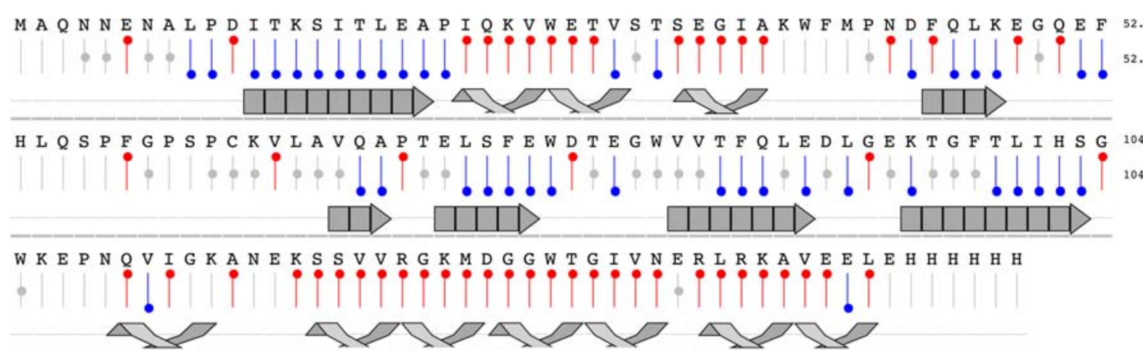

Fig. 2 The ^{15}N, ^{13}Cα, and ^{13}Cβ chemical shift index (Wishart et al. 1991) of yndB, where blue circles imply a β-strand region and red circles suggest an α-helix. The position of each circle along the bar indicates the relative consistency of the chemical shifts with either a β-sheet or α-helix. The lack of a circle indicates a missing NMR assignment. A consensus secondary structure (McGuffin et al. 2000) for the START domain proteins is shown below the chemical shift index, where an arrow indicates a β-strand and a ribbon indicates an α-helix

predicted secondary structure of yndB and the consensus secondary structure of START domain proteins. The unassigned residues correspond to the predicted loop regions of the START domain fold. The identification and addition of a ligand to the yndB NMR sample may improve the quality of the NMR data for these loop regions by reducing the inherent flexibility and inducing a defined conformation. The sequence specific ^{1}H, ^{13}C, and ^{15}N resonance assignments of the yndB were deposited in the BioMagResBank under accession number 16043.

Acknowledgments We thank Prof. Diana Murray for stimulating discussions regarding the NESG START domain project and Dr. John Cort at PNNL for providing the 750 MHz NMR data. This work was supported by a grant from the Protein Structure Initiative of the National Institutes of Health (U54 GM074958 to G.T.M.), Nebraska Tobacco Settlement Biomedical Research Development Fund and by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. Research was performed in facilities renovated with support from the NIH under Grant RR015468-01. NMR spectra were acquired at the US Department of Energy Office of Biological and Environmental Research) located at Pacific Northwest National Laboratory and operated for DOE by Battelle (contract KP130103).

References

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 (2005) Robotic cloning and protein production platform of the northeast structural genomics consortium. Methods Enzymol 394:210–243

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

Ferentz AE, Wagner G (2000) NMR spectroscopy: a multifaceted approach to macromolecular structure. Q Rev Biophys 33:29–65

Garrett DS, Powers R, Gronenborn AM, Clore GM (1991) A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. J Magn Reson 95:214–220

Iyer LM, Koonin EV, Aravind L (2001) Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins 43:134–144

Jansson M, Li YC, Jendeberg L, Anderson S, Montelione BT, Nilsson B (1996) High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in *Escherichia coli*. J Biomol NMR 7:131–141

Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol 278: 313–352

Kirby NI, DeRose EF, London RE, Mueller GA (2004) NvAssign: protein NMR spectral assignment with NMRView. Bioinformatics 20:1201–1203

Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szyperski T (2005) NMR data collection and analysis protocol for high-throughput protein structure determination. Proc Natl Acad Sci USA 102:10487–10492

Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H–^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry 28:6150–6156

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404–405

Wagner G, Wuthrich K (1986) Observation of internal motility of proteins by nuclear magnetic resonance in solution. Methods Enzymol 131:307–326

Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 222:311–333