NIH Public Access
Author Manuscript

Structure of YidB Protein from *Shigella flexneri* Shows a New Fold with Homeodomain Motif

Jerzy Osipiuk, **Natalia Maltseva**, **Irina Dementieva**, **Shonda Clancy**, **Frank Collart**, and **Andrzej Joachimiak**[*]

Biosciences Division, Midwest Center for Structural Genomics and Structural Biology Center, Argonne National Laboratory, Argonne, Illinois

Introduction

The crystal structure of an uncharacterized conserved protein S4005, coded by *yidB* gene of *Shigella flexneri* (gi:30043267),[1] has been determined by single wavelength anomalous diffraction (SAD) method and refined to 1.45 Å. The YidB structure is the first representative of COG3753 and Pfam06078 medium size families of bacterial proteins of unknown function (Fig. 1). The *yidB* gene of *S. flexneri*, as well as *yidB* gene of *Escherichia coli*, is located between *yidA* and *gyrB* genes which are involved in DNA processing. Biochemical function of the *yidA* product is unknown, but it is predicted to have hydrolase/phosphotase activity.[2,3] The other neighbor, *gyrB*, codes subunit B of DNA gyrase type II topoisomerase which controls DNA supercoiling and DNA-relaxing.[4] It is often found that genes in bacteria are clustered according to their products functions.[5] Thus, it is possible that the YidB protein can have functions associated with DNA. YidB is found in number of pathogenic species including *Escherichia, Bordetella, Burkholderia*, and *Shigella* species (Fig. 1).

Here we report the crystal structure of YidB protein at 1.45 Å resolution. The structure represents a new protein fold and shows distant structural similarity to eukaryotic homeodomain proteins.

Materials and Methods

Protein cloning, expression, and purification

The y*idB* gene was cloned in pMCSG7 vector[6] and overexpressed in *E. coli* BL21 (DE3) - Gold (Stratagene) harboring an extra plasmid encoding three rare tRNAs (AGG and AGA for Arg, ATA for Ile). The pMCSG7 vector bearing a TEV protease cleavage site creates a construct with cleavable His_6-tag fused into N-terminus of the target protein and adds three artificial residues (Ser–Asn–Ala) on that end. The cells were grown using selenomethionine (SeMet) containing enriched M9 medium and conditions known to inhibit methionine biosynthesis.[7,8] The cells were grown at 37°C to an OD_{600} of ~0.6 and protein expression induced with 1 mM IPTG. After induction, the cells were grown overnight with shaking at 20°C. The harvested cells were resuspended in five volumes of lysis buffer (50 mM HEPES, pH

8.0, 500 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, and 5% v/v glycerol) and stored at −20°C.

The thawed cells were lysed by sonication after the addition of inhibitor proteases (Sigma, P8849) and 1 mg/mL lysozyme. The lysate was clarified by centrifugation at 30,000 × g (RC5C-Plus centrifuge, Sorval) for 20 min, followed by filtration through 0.45 μm and 0.22 μm inline filters (Gelman).

The standard purification protocol is thoroughly described previously.[9] Immobilized metal affinity chromatography (IMAC-I) using a 5-mL HiTrap Chelating HP column charged with Ni^{+2} ions and buffer-exchange chromatography on a HiPrep 26/10 desalting column (both Amersham Biosciences) were performed using AKTA EXPLORER 3D (Amersham Biosciences). His_6-tag was cleaved using the recombinant TEV protease expressed from the vector pRK508.[10] The protease was added to the target protein in a ratio of 1:30 and the mixture was incubated at 4°C for 48 h. The YidB protein was then purified using a 1-mL HiTrap Chelating column (Amersham Biosciences) charged with Ni^{+2} ions. The protein was dialyzed in 20 mM Tris-HCl, (pH 7.1), 50 mM NaCl, 2 mM DTT, and concentrated using a Centricon Plus-20 Centrifugal Concentrator (Millipore).

Protein crystallization

Initially protein crystals were obtained in sitting drops by using Index (Hampton Research) and Wizard I and II (Emerald Biostructures) crystallization screens with the help of a HoneyBee crystallization workstation (Cartesian Technologies). The first crystals appeared after several days in Index #18, #48, and Wizard I #19, #23 crystallization conditions. In approximately 2 wk, a large crystal conglomerate was formed in Index condition #4. Further optimization of this condition was done manually using the hanging drop technique. The protein solution (1 μL, 44 mg/mL) was mixed with 1 μL of 0.1 M Bis-Tris (pH 7.2) and 2.0 M ammonium sulfate, and equilibrated over 1 mL of crystallization solution at 23°C. Quality crystals, which appeared in 2 wk, were flash-frozen in liquid nitrogen with crystallization solution complemented with 15% (v/v) glycerol as cryoprotectant prior to data collection.

Data collection

Diffraction data were collected at 100 K temperature at the 19BM beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. The single wavelength anomalous dispersion (SAD) data at 0.9793 Å (peak: 12.6603 keV) up to 1.45 Å were collected from a single (0.1 × 0.02 × 0.05 mm) Se–Met labeled protein crystal at 100 K. The space group was C2 with cell dimension of a = 57.48 Å, b = 40.48 Å, c = 48.33 Å, α =90.00°, β =93.78δ, and γ = 90.00°. There is one protein molecule in the asymmetric unit. All data were processed and scaled with HKL2000 suite[11] (Table I).

Structure determination, refinement, and deposition

The YidB structure was determined by SAD phasing using HKL2000_PH (W. Minor University of Virginia, personal communication) and RESOLVE[12] and refined to 1.45 Å using REFMAC 5.2[13] in CCP4 suite.[14] The initial model was completed by using ARP/wARP[15] and manual fitting using COOT[16] and O[17] programs. The Structure Analysis server (STAN) was used to run the WASP program[18] for identification of sodium ion in the structure. The stereochemistry of the structure was checked with PRO-CHECK.[19] Atomic coordinates and experimental structure factors of YidB have been deposited with the PDB and are accessible under the code 1Z67.

Results and Discussion

The YidB structure is composed of eight α-helices connected by short β-turns (H1~H8, Fig. 2), where helices (H2–H7) form a compact six helix bundle with a well defined hydrophobic core. A number of hydrophobic residues that contribute to the core are conserved (Leu35, Trp51, Leu70, Leu93, Leu97). Two N-and C-terminal helices, H1 and H8, project out from the protein body and make few contacts with the main protein body (particularly H1). Their orientation is maintained by crystal packing contacts, therefore these helices could assume different orientations and may serve as interaction surfaces. Both helices have well defined hydrophobic/hydrophilic surfaces, but most residues are not conserved. The exception is N-terminal sequence motif MGL(L/F)D (MG are not visible in our structure) and Gly9 and Gly14 in H1 that are very strongly conserved. The loop connecting H1 and H2 helices is very short (residues 15–16), but it is flanked at both ends by glycine residues (G14 and G17) which may allow H1 helix to move. The loop connecting H7 and H8 helices is formed by 12 amino-acid residues. Two internal residues of the loop, Ser111 and Ala112, are not visible in the structure. No putative dimer interface was identified by visual inspection of protein contacts inside the crystal and PQS[20] search predicted a monomeric form for this protein.

A structural homology search using DALI server[21] showed some very distant structural homologs. The closest match was the NMR structure of Mouse Homeodomain-Only Protein HOP (1UHS.pdb) with Z-score and RMSD equal to 3.9 and 3.7, respectively. The next two matches were other homeodomain proteins in complexes with DNA, yeast MATa1/MATa2 homedomain heterodimer (1AKH.pdb)[22] and *Drosophila* Engrailed Homeodomain (2HDD.pdb)[23] with Z-scores equal to 3.7 and 3.5 and RMSD - 2.3 and 2.5, respectively. The superposition of structural homologs onto YidB structure showed that similarity is limited to only three of eight helices of our structure, H2, H6, and H7 (Fig. 3). These helices contribute to the hydrophobic core and contain several highly conserved residues. The homeodomain structural motif of three successive helices is characteristic of eukaryotic homeodomains which are one of the key DNA-binding domains used in gene regulation.[24] The helix H7 of YidB corresponds to the DNA-binding helix 3 of homeodomains[22] (data not shown). The H7 helix has several hydrophilic and charged residues that are solvent exposed and may interact with nucleic acid. However, the exact superposition with 1AKH results in collision of the H4 - H5 helices region of our structure with MATa1/MATa2 homedomain heterodimer DNA complex. Therefore, YidB protein could interact with nucleic acid only after undergoing significant conformational change or if its mode of interaction is very different from homeodomains. Nevertheless the presence of such a configuration in bacterial proteins suggests that this motif was invented very early in protein evolution. However, taking into account only partial homology of YidB structure to homeodomains (25% of sequence) we believe that our structure represents a unique structure and a new protein fold.

We have searched a number of databases including PQS,[20] BLAST,[25] ProFunc,[26] DALI,[21] and ISREC-TMpred[27] servers to assign more detailed protein function. Sequence comparisons showed 201 matching sequences found by PSI-BLAST with nearly all of them being conserved proteins of unknown function. The ISREC-TMpred Server[27] did not find any trans-membrane helices. Enzyme template search with ProFunc[26] identified Glu28 and Asp77 in S4005 as a part of horse lysozyme active site template but these residues are not well conserved across the YidB family suggesting that YidB is unlikely an enzyme. Interestingly, at this site the sodium ion was found in the structure coordinated by Ser76, Asp77, Gln80 and several nearby carbonyls. Therefore, there are some indications that YidB may be a nucleic acid binding protein but this hypothesis requires further investigation and experimental verification.

Acknowledgments

References

1. Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, Mayhew GF, et al. Complete genome sequence and comparative genomics of *Shigella flexneri s*erotype 2a strain 2457T. Infect Immun 2003;71:2775–2786. [PubMed: 12704152]

2. Krieger CJ, Zhang P, Mueller LA, Wang A, Paley S, Arnaud M, Pick J, et al. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res 2004;32:438–442.

3. Burland V, Plunkett G 3rd, Daniels DL, Blattner FR. DNA sequence and analysis of 136 kilobases of the *Escherichia coli* genome: organizational symmetry around the origin of replication. Genomics 1993;3:551–561. [PubMed: 7686882]

4. Adachi T, Mizuuchi M, Robinson EA, Appela E, O'Dea MH, Gellert M, Mizuuchi K. DNA sequence of the E. coli *gyrB* gene: application of a new sequencing strategy. Nucl Acids Res 1987;15:771–784. [PubMed: 3029692]

5. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci USA 1999;96:2896–2901. [PubMed: 10077608]

6. Stols L, Gu M, Dieckman L, Raffen R, Collart FR, Donnelly MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus virus protease cleavage site. Protein Expr Purif 2002;25:8–15. [PubMed: 12071693]

7. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J Mol Biol 1993;229:105–124. [PubMed: 7678431]

8. Walsh MA, Dementieva I, Evans G, Sanishvili R, Joachimiak A. Taking MAD to the extreme: ultrafast protein structure determination. Acta Crystallogr D Biol Crystallogr 1999;55:1168–1173. [PubMed: 10329779]

9. Kim Y, Dementieva I, Zhou M, Wu R, Lezondra L, Quartey P, Joachimiak G, et al. Automation of protein purification for structural genomics. J Struct Funct Genomics 2004;5:111–118. [PubMed: 15263850]

10. Kapust RB, Waugh DS. Controlled intracellular processing of fusion proteins by TEV protease. Protein Expr Purif 2000;19:312–318. [PubMed: 10873547]

11. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.

12. Terwilliger TC. Automated main-chain model-building by template-matching and iterative fragment extension. Acta Crystal-logr D Biol Crystallogr 2003;59:38–44.

13. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240–255. [PubMed: 15299926]

14. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763. [PubMed: 15299374]

15. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol 1999;6:458–463. [PubMed: 10331874]

16. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126–2132. [PubMed: 15572765]

17. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119. [PubMed: 2025413]

18. Nayal M, Di Cera E. Valence screening of water in protein crystals reveals potential Na+ binding sites. J Mol Biol 1996;256:228–234. [PubMed: 8594192]
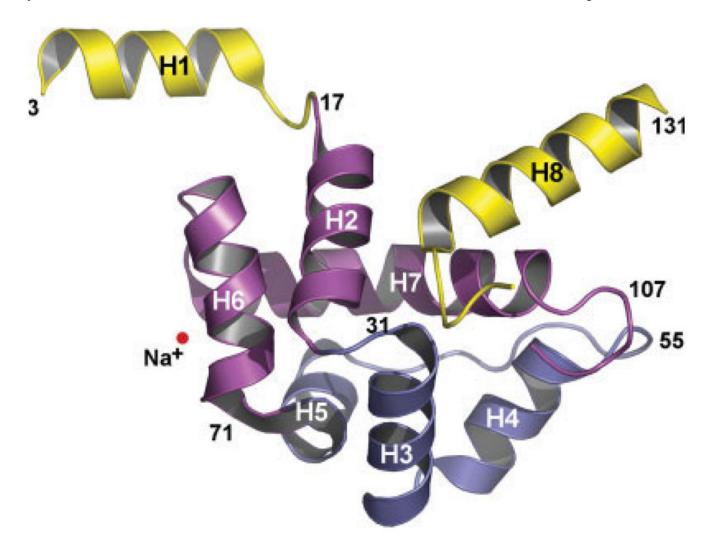
19. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PRO-CHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.

20. Henrick K, Thornton SM. PQS: a protein quaternary structure file server. Trends Biochem Sci 1998;23:358–361. [PubMed: 9787643]

21. Holm L, Sander C. Touring protein fold space with Dali/FSSP. Nucleic Acids Res 1998;26:316–319. [PubMed: 9399863]

22. Li T, Jin Y, Vershon AK, Wolberger C. Crystal structure of the MATa1/MATalpha2 homeodomain heterodimer in complex with DNA containing an A-tract. Nucleic Acids Res 1998;26:5707–5718. [PubMed: 9838003]

23. Tucker-Kellogg L, Rould MA, Chambers KA, Ades SE, Sauer RT, Pabo CO. Engrailed (Gln50→Lys) homeodomain-DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity. Structure 1997;5:1047–1054. [PubMed: 9309220]

24. Gehring WJ, Affolter M, Burglin T. Homeodomain proteins. Annu Rev Biochem 1994;63:487–526. [PubMed: 7979246]

25. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 1997;25:3389–3402. [PubMed: 9254694]

26. Laskowski RA, Watson JD, Thornton JM. From protein structure to biochemical function. J Funct Genomics 2003;4:167–177.

27. Hofmann K, Stoffel W. TMbase—A database of membrane spanning proteins segments. Biol Chem Hoppe-Seyler 1993;374:166.

28. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. Trends Biochem Sci 1998;23:403. [PubMed: 9810230]

29. DeLano, WL. The PyMOL molecular graphics system. http://www.pymol.org

Fig. 1.

Multiple sequence alignment of YidB protein from *Shigella flexneri* (S4005 protein, gi: 30043267) and its homologs gi|75177733| (*Shigella boydii*), gi|67662828| (*Burkholderia cenocepacia*), gi|67534931| (*B. vietnamiensis*), gi|74019150| (*B. mallei*), gi|75258238| (*Escherichia coli* E22), gi|75242512| (*E. coli* F11), gi|38704197| (*E. coli* O157), gi|33597705| (*Bordetella parapertussis*), gi|33592815| (*B. pertussis*), gi|68004308| (*Geobacter metallireducens*). Some of these proteins have been identified as members of COG3753. The alignment was generated by the ClustalX program.[28] The secondary structure elements in YidB (S4005) protein are superimposed on its primary sequence. This diagram was generated based on secondary structure analysis by the PDBsum server at http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/.

Fig. 2.
Crystal structure of YidB protein. Ribbon diagram of protein structure. Terminal helices, H1 and H8, which are not a part of the protein core, are shown in yellow. The helices of the protein core are shown in slate and magenta. Helices H2, H6, and H7 shown in magenta correspond to homeodomain motif based on superposition with 1AKH.pdb structure. Location of the sodium ion is marked as a red sphere. Residue numbers are described in black. The picture was generated using the Pymol program.[29]

Fig. 3.
Superposition of S4005 protein (1Z67.pdb, green) and MATa1/MATa2 homedomain heterodimer (1AKH.pdb, magenta).

TABLE I

Summary of the YidB crystallographic data

Data collection	
Experiment	SAD
Space group	C2
Unit cell dimensions	
a, b, c (Å)	57.48, 40.48, 48.33
α, β, γ (deg)	90.00, 93.78, 90.00
Protein molecules/a.u.[a]	1
Energy (keV)	0.9793
Data processing	
Resolution range (Å)	40–1.45 (1.48–1.45)[b]
Unique reflections	17111
Redundancy	4.1 (1.6)[b]
Completeness (%)	86.5 (28.6)[b]
I/σI	39.2 (2.14)[b]
R_{merge}[c] **(%)**	4.0 (33.0)[b]
Refinement	
Resolution range (Å)	40.0–1.45 (1.49–1.45)[b]
R_{work} (%)	16.2 (18.3)[b]
R_{free} (%)	20.8 (29.0)[b]
RMSD of	Bond lengths = 0.017 Å Bond angles = 1.427°
Protein residues	127 of 135[d]
Water	156
Na^+	1
Mean *B* factor	22.9
Ramachandran plot statistics	
Residues in most favored regions (%)	97.2
Residues in additional allowed regions (%)	2.8
Residues in disallowed region (%)	0.0

[a] a.u. = asymmetric unit.

[b] In parenthesis are statistics for the highest resolution shell.

[c] $R_{merge} = (\Sigma|I_{hkl} - <I>I)/\Sigma|I_{hkl}|$, where the average intensity $<I>$ is taken over all symmetry equivalent measurements and I_{hkl} is the measured intensity for any given reflection. $R_{factor} = \Sigma|F_{obs} - F_{calc}|/\Sigma|F_{obs}|$.

[d] Including three residues at protein N-terminus added as a cloning artifact.