Published in final edited form as:

J Struct Funct Genomics. 2011 September ; 12(3): 175–179. doi:10.1007/s10969-011-9117-z.

Solution NMR Structure of Dsy0195 Homodimer from *Desulfitobacterium hafniense*: First Structure Representative of the YabP Domain Family of Proteins Involved in Spore Coat Assembly

Yunhuang Yang^{1,2}, Theresa A. Ramelot^{1,2}, John R. Cort^{2,3}, Huang Wang^{2,4}, Colleen Ciccocanti^{2,4}, Mei Jiang^{2,4}, Haleema Janjua^{2,4}, Thomas B. Acton^{2,4}, Rong Xiao^{2,4}, John K. Everett^{2,4}, Gaetano T. Montelione^{2,4,5}, and Michael A. Kennedy^{1,2,*}

¹Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056

²Northeast Structural Genomics Consortium, Piscataway, New Jersey 08854

³Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352

⁴Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854

⁵Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Abstract

Protein domain family YabP (PF07873) is a family of small protein domains that are conserved in a wide range of bacteria and involved in spore coat assembly during the process of sporulation. The 62-residue fragment of Dsy0195 from *Desulfitobacterium hafniense*, which belongs to the YabP family, exists as a homodimer in solution under the conditions used for structure determination using NMR spectroscopy. The structure of the Dsy0195 homodimer contains two identical 62-residue monomeric subunits, each consisting of five anti-parallel beta strands (β 1, 23-29; β 2, 31-38; β 3, 41-46; β 4, 49-59; β 5, 69-80). The tertiary structure of the Dsy0195 monomer adopts a cylindrical fold composed of two beta sheets. The two monomer subunits fold into a homodimer about a single C2 symmetry axis, with the interface composed of two anti-parallel beta strands, β 1- β 1' and β 5b- β 5b', where β 5b refers to the C-terminal half of the bent β 5 strand, without any domain swapping. Potential functional regions of the Dsy0195 structure were predicted based on conserved sequence analysis. The Dsy0195 structure reported here is the first representative structure from the YabP family.

Keywords

PF07873; YabP; Dsy0195; Sporulation Protein; Structural Genomics; NMR

INTRODUCTION

The YabP family (PF07873) is a conserved prokaryotic protein domain family that contains 465 sequences found in 241 bacterial species (Pfam 25.0, <http://pfam.sanger.ac.uk/>, based on

*Correspondence to: Michael A. Kennedy, Department of Chemistry and Biochemistry, Miami University, 701 East High Street, Oxford, OH, 45056. Phone: +1 513 529 8267, Fax: +1 513 529 5715, kennedm4@muohio.edu.

the August 23, 2011 update). Members of the YabP domain family are involved in spore coat assembly during the process of sporulation. Using fusions to green fluorescent protein (GFP), YabP-GFP from *Bacillus subtilis* was observed to assemble into two rings around the forespore before it eventually redistributed to form a shell around the developing spore.¹ As part of the efforts of the Protein Structure Initiative (Phases I and II) to obtain experimental structures of representatives from broadly conserved protein domain sequence families, Dsy0195 from *Desulfitobacterium hafniense*, which belongs to the YabP family, was selected as one of the Northsat Structure Genomics (NESG) consortium targets (UniProtKB/TrEMBL ID, Q251Q8_DESHY; NESG ID, DhR8C).

The truncated *Dsy0195* gene from *Desulfitobacterium hafniense* (strain Y51) encodes for a 62-residue domain (residues 21-82, hereafter referred to as Dsy0195). The NMR structure of the Dsy0195 homodimer was initially solved using conventional methods, in which inter-monomer distance constraints were derived from edited/filtered NOESY data² of samples prepared by mixing ¹³C/¹⁵N-labeled and unlabeled Dsy0195 in a 1:1 molar ratio. Due to poor signal-to-noise of the edited/filtered NOESY and confounding structural symmetry in the homodimer, only 14 unambiguous inter-monomer NOEs across the homodimer interface were assigned manually. This limited number of interchain NOE restraints resulted in a poorly defined interface in the early stages of the automated Dsy0195 homodimer structure determination procedure, which in turn caused some misassignments of both intra- and inter-monomer NOEs.³ The overall fold, however, was similar to the Dsy0195 crystal structure (PDB ID: 3IPF). The backbone RMSD of average solution NMR structures against the crystal structure was 1.51 Å. This solution NMR structure of Dsy0195 was deposited into the Protein Data Bank with PDB ID 2KS0.

To overcome the paucity of inter-monomer restraints available for structure calculations of the homodimer structure, we implemented a novel approach for improving the quality of the solution NMR structure of the homodimer protein that used a combination of paramagnetic resonance enhancement (PRE) and double electron electron resonance (DEER) measurements of spin labeled Dsy0195.³ The Dsy0195 homodimer structure solved by this new approach had higher quality than the structure solved by the conventional approach, especially in the interface region, due to an increased number of accurate intermonomer NOEs assigned during the automatic assignment of NOEs from the NOESY peak lists using the CYANA⁴ program following seeding with long-range interchain distance constraints provided by PRE and DEER measurements. The backbone RMSD of average solution NMR structures of Dsy0195 solved using PRE and DEER restraints compared to the crystal structure was 1.13 Å, which confirmed the better agreement of the overall structure fold with the Dsy0195 crystal structure. The structure of Dsy0195 solved with the aid of PRE and DEER restraints has been deposited into the Protein Data Bank with PDB ID 2KYI. Since the 2KYI structure is of higher quality compared to the 2KS0, it will be presented in this *Structure Report*.

Here, we present the NMR structure of the Dsy0195 homodimer in solution and predict potentially important functional regions of the structure based on conserved sequence analysis. The Dsy0195 monomer adopts a unique cylindrical fold made of five beta strands, a fold that does not match any existing domains in the CATH (<http://www.cathdb.info/>) and SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>) structural classification databases.

METHODS

[U-¹³C, ¹⁵N]- and U-¹⁵N, 5% biosynthetically directed ¹³C (NC5) Dsy0195 samples from *Desulfitobacterium hafniense* (strain Y51) were produced following standard protocols of the Northeast Structural Genomics Consortium (NESG).⁶ Based on analytical gel filtration

chromatography, static light scattering (Supplementary Fig. S1) and one-dimensional ^{15}N T_1 and T_2 relaxation data (Supplementary Fig. S2), Dsy0195 existed as a homodimer in solution under the conditions used in the NMR studies (1.0 mM monomer subunit, 10% v/v D_2O , 20 mM ammonium acetate, 100 mM NaCl, 10 mM DTT, 5 mM CaCl_2 , pH 4.5, 20 °C). Two single cysteine mutants of Dsy0195 (S36C and S52C) were generated by using the QuikChange site-directed mutagenesis kit (Stratagene). Neither mutation, S36C and S52C, perturbed the homodimer structure based on comparison of the 2D ^{15}N -HSQC spectra of mutants with wild type.³ (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate (MTSL) was used to spin-label free cysteine in the S36C and S52C mutants. To obtain inter-monomer distance constraints, samples were prepared for edited/filtered ^{13}C -edited NOESY, PRE and DEER experiments. Backbone ^1H , ^{13}C , and ^{15}N resonance assignments were made by AutoAssign 2.4.0⁷ (Supplementary Fig. S3); side chain assignments were completed manually. NMR resonance assignment validation was performed using the AVS software.⁸ Dihedral angle constraints were computed by TALOS⁹ ($f \pm 30^\circ$; $y \pm 30^\circ$) for ordered residues with confidence scores of 10. Hydrogen bond constraints were derived from AutoStructure 2.1.1¹⁰ and confirmed by deuterium exchange measurements. The Dsy0195 monomer structure was calculated using CYANA, supplied with an unassigned NOESY peak list obtained from 3D ^{13}C -edited NOESY (aliphatic and aromatic regions in ^{13}C dimension), ^{15}N -edited NOESY, and 4D ^{13}C - ^{13}C -HMQC-NOESY-HMQC spectra, together with dihedral angle and hydrogen bond constraints.

CYANA seeded with inter-monomer constraints from the different sources described above together with intra-monomer constraints, dihedral angle and hydrogen bond constraints, led to the automatic assignment of all intra- and inter-monomer NOEs used for structure determination over several rounds of peak list refinement. The final distance constraints were converted into XPLOR format, followed by final structure calculations using XPLOR-NIH¹¹ and refinement in explicit water using CNS. NMR structure quality analyses were performed using the PSVS¹² and RPF¹³ software packages. (The Supplementary Material contains a detailed description of the methods used in this work).

RESULTS AND DISCUSSION

Here we present structural features of 2KYI. As shown in Fig. 1A, the NMR structure of the Dsy0195 homodimer in solution contains two identical 62-residue monomeric subunits, each consisting of five anti-parallel beta strands ($\beta 1$, 23-29; $\beta 2$, 31-38; $\beta 3$, 41-46; $\beta 4$, 49-59; $\beta 5$, 69-80). Among these five beta strands, three relatively short strands $\beta 1$, $\beta 2$ and $\beta 3$ contain 6-8 residues, while the other two strands $\beta 4$ and $\beta 5$ contain 11 and 12 residues, which bend at the middle of each strand. When these two longer beta strands, $\beta 4$ and $\beta 5$, are individually taken as two short parts bending at residue E56 on $\beta 4$ and I76 on $\beta 5$ ($\beta 4a$: 49-56; $\beta 4b$: 57-59; $\beta 5a$: 69-76; $\beta 5b$: 77-80), the structure can be described as being composed of a total of seven β strands forming two anti-parallel beta sheets in the order of $\beta 1$ - $\beta 5a$ - $\beta 4b$ and $\beta 2$ - $\beta 3$ - $\beta 4a$ - $\beta 5b$. The tertiary structure of the Dsy0195 monomer adopts a cylindrical fold composed of these two beta sheets. The two monomer subunits fold into the homodimer with one C2 symmetry axis as shown in Fig. 1A, with the interface composed of two anti-parallel beta strands, $\beta 1$ - $\beta 1'$ and $\beta 5b$ - $\beta 5b'$ without any domain swapping. The secondary structure elements are locally and globally well defined (Fig. 1B). Structural statistics for Dsy0195 homodimer are presented in Table 1.

Fig. 1C shows the overlay of NMR structure (green) against crystal structure (cyan). The secondary NMR structure elements are consistent with the crystal structure (five beta strands in crystal structure: $\beta 1$, 24-28; $\beta 2$, 31-38; $\beta 3$, 41-46; $\beta 4$, 49-64; $\beta 5$, 69-81). The overall fold of NMR and crystal structure are almost identical with the backbone RMSD of 1.13 Å. However, resonances from a five-residue segment ranging from residue 60 to 64 (IKHLD)

are completely absent in the NMR spectra due to conformational exchange broadening. This short segment, which is located at the end of $\beta 4$ in the crystal structure, could not be well determined.

Dsy0195 shows no similarity to domains in either the CATH or the SCOP structural classification databases. PSI-BLAST results indicated very low sequence similarity (< 20%) with any other structures in the PDB. Dali (v.3)⁵ structural alignment analyses using the NMR structure reported here (2KYI) revealed low similarity to known structures with a highest *Z*-score of 3.6 (PDB ID: 2WDA; C ^{α} RMSD, 5.2 Å).

ConSurf¹⁴ analyses of the YabP family using PSI-BLAST (92 sequences in total) revealed two patches containing highly conserved residues. The smaller region, which occurs near the dimer interface, contained five highly conserved residues T29, G30, V31, V34 and Y81. (Fig. 1A, left panel) located at the end of $\beta 1$, the beginning of $\beta 2$, and the loop connecting these two beta strands. The second region, which is oriented away from the dimer interface, is much bigger with twelve highly conserved residues S36, F37, D38, I42, L44, T46, L51, L58, I60, L63, L65 and G68 (Fig. 2A, right panel) mainly located on $\beta 2$, $\beta 3$ and $\beta 4$. PDBemotif analysis revealed one 22-residue segment of L44-(x)₆-L51-(x)₆-L58-(x)₆-L65, showing sequence characteristics of a leucine zipper motif, which is normally present in α -helical dimers (Supplementary Fig. S4, <http://www.ebi.ac.uk/pdbe-site/pdbemotif/>). Seven highly conserved residues including L44, T46, L51, L58, I60, L63 and L65 were identified, which suggests an important role for these residues either for preserving the structure or function of these proteins during the course of the evolution of these species. The side chains of these nonpolar residues, mostly present in $\beta 3$ and $\beta 4$ (Fig. 2B), may help the protein adopt the proper fold.

In summary, we have determined the high quality NMR structure of Dsy0195 homodimer in solution from *Desulfitobacterium hafniense*, the first structure representative of any member from the highly conserved YabP family (PF07873). Though potentially functional regions were identified by Consurf and PDBemotif analyses based on conserved sequence and structure features, elucidation of the exact function of the Dsy0195 homodimer awaits further study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute of General Medical Sciences; Grant Number: U54-GM074958. The majority of the data collection was conducted at the Ohio Biomedicine Center of Excellence in Structural Biology and Metabonomics at Miami University. A portion of the NMR experiments was performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory.

References

1. Ooij CV, Eichenberger P, Losick R. Dynamic patterns of subcellular protein localization during spore coat morphogenesis in *Bacillus subtilis*. *J Bacteriol*. 2004; 186:4441–4448. [PubMed: 15231775]
2. Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE. Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *Journal of the American Chemical Society*. 1997; 119:6711–6721.

3. Yang Y, Ramelot TA, McCarrick RM, Ni S, Feldmann EA, Cort JR, Wang H, Ciccocanti C, Jiang M, Janjua H, Acton TB, Xiao R, Everett JK, Montelione GT, Kennedy MA. Combining NMR and EPR methods for homodimer protein structure determination. *J Am Chem Soc.* 2010; 132:11910–11913. [PubMed: 20698532]
4. Guntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol.* 2004; 278:353–378. [PubMed: 15318003]
5. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 2010; 38:W545–549. [PubMed: 20457744]
6. Acton TB, Xiao R, Anderson S, Aramini J, Buchwald WA, Ciccocanti C, Conover K, Everett J, Hamilton K, Huang YJ, Janjua H, Kornhaber G, Lau J, Lee DY, Liu G, Maglaqui M, Ma L, Mao L, Patel D, Rossi P, Sahdev S, Shastry R, Swapna GV, Tang Y, Tong S, Wang D, Wang H, Zhao L, Montelione GT. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol.* 2010; 493:21–60. [PubMed: 21371586]
7. 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]
8. Moseley HN, Sahota G, Montelione GT. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. *J Biomol NMR.* 2004; 28:341–355. [PubMed: 14872126]
9. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR.* 1999; 13:289–302. [PubMed: 10212987]
10. Huang YJ, Tejero R, Powers R, Montelione GT. A topology-constrained distance network algorithm for protein structure determination from NOESY data. *Proteins.* 2006; 62:587–603. [PubMed: 16374783]
11. Schwieters CD, Kuszewski JJ, Clore GM. Using Xplor-NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy.* 2006; 48:47–62.
12. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. *Proteins.* 2007; 66:778–795. [PubMed: 17186527]
13. 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; 127:1665–1674. [PubMed: 15701001]
14. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 2010; 38:W529–533. [PubMed: 20478830]

ABBREVIATION

DEER	double electron electron resonance
GFP	green fluorescent protein
MTSL	(1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate
NESG	Northeast Structural Genomics Consortium
NMR	nuclear magnetic resonance spectroscopy
NOE	nuclear Overhauser effect
NOESY	nuclear overhauser effect spectroscopy
PRE	paramagnetic relaxation enhancement
RMSD	root mean square deviation

Figure 1.

(A) Stereoview of the lowest energy conformer of the final NMR ensemble of the Dsy0195 homodimer (PDB ID: 2KYI). The C2 symmetry axis of Dsy0195 homodimer is indicated by as red dashed line and labeled C2. (B) Superposition of 20 lowest energy conformers of Dsy0195 homodimer without the disordered C-terminal His-tag tails for clarity. Secondary structure elements are colored in red and loops are colored grey. (C) Overlay of the solution NMR structure of Dsy0195 homodimer (2KYI, green) with the X-ray crystal structure (3IPF, cyan). The backbone RMSD between the NMR and crystal structures was 1.13 Å.

Figure 2.

(A) ConSurf analysis was conducted for YabP family (92 sequences in total using PSI-BLAST). Standard ConSurf residue coloring reflecting the degree of residue conservation over the entire family was used (color scheme: magentas, highly conserved; cyan, variable). Highly conserved residues are labeled. (B) Ribbon drawing of Dsy0195 monomer for identical color, size and orientation in (A, right panel). Side chains of L44, L51, L58 and L65 are shown.

Table 1Summary of NMR and structural statistics for Dsy0195 (2KS0 and 2KYI)^a

PDB ID	2KS0	2KYI
Completeness of resonance assignments ^b		
Backbone (%)	98.7	98.7
Side chain (%)	99.8	99.8
Aromatic (%)	100	100
Stereospecific methyl (%)	100	100
Conformationally-restricting constraints ^c		
Distance constraints		
Total	2490	1894
intra-residue ($i = j$)	638	502
sequential ($ i - j = 1$)	622	416
medium range ($1 < i - j < 5$)	358	152
long range ($ i - j \geq 5$)	872	824
inter-monomer constraints	102	125
Dihedral angle constraints	152	152
Hydrogen bond constraints	80	80
Number of constraints per residue	39.5	33.7
Number of long range constraints per residue	15.0	14.2
Residual constraint violations ^c		
Average number of distance violations per structure		
0.1 – 0.2 Å	19.9	2.45
0.2 – 0.5 Å	6.05	0.35
> 0.5 Å	0	0
Average number of dihedral angle violations per structure		
1 – 10°	6.85	7.75
> 10°	0	0
RMSD from average coordinates (Å) ^{c,d}		
backbone atoms	0.5	0.5
heavy atoms	0.8	0.9
MolProbity Ramachandran statistics ^{c,d}		
most favored regions (%)	86.0	85.4
allowed regions (%)	14.0	14.6
disallowed regions (%)	0.0	0.0
Global quality scores (Raw / Z-score) ^c		
Verify3D	0.33/-2.09	0.36/-1.61
ProsaII	0.22/-1.78	0.19/-1.90
Procheck(phi-psi) ^d	-0.55/-1.85	-0.77/-2.71
Procheck(all) ^d	-0.43/-2.54	-0.47/-2.78
Molprobity clash	30.23/-3.66	21.23/-2.12

PDB ID	2KS0	2KYI
RPF Scores ^e		
Recall / Precision	0.885/0.882	0.934/0.89
F-measure / DP-score	0.914/0.713	0.911/0.79

^aStructural statistics were computed for the ensemble of 20 deposited structures.

^bComputed using AVS software 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 His₆ tag.

^cCalculated using PSVS 1.4 program. Average distance violations were calculated using the sum over r^{-6} .

^dOrdered residue ranges [$S(\phi) + S(\psi) > 1.8$] : 22A-46A,50A-59A,68A-82A,22B-46B,50B-58B,68B-81B.

^eRPF scores reflecting the goodness-of-fit of the final ensemble of structures (including disordered residues) to the NOESY data and resonance assignments.