Published in final edited form as:

J Struct Funct Genomics. 2012 September ; 13(3): 163–170. doi:10.1007/s10969-012-9134-6.

Three structural representatives of the PF06855 protein domain family from *Staphylococcus aureus* and *Bacillus subtilis* have SAM domain-like folds and different functions

G.V.T. Swapna,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and the Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, and Robert Wood Johnson Medical School, UMDNJ, Piscataway, NJ 08854, USA

Paolo Rossi,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and the Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, and Robert Wood Johnson Medical School, UMDNJ, Piscataway, NJ 08854, USA

Alexander F. Montelione,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and the Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, and Robert Wood Johnson Medical School, UMDNJ, Piscataway, NJ 08854, USA

Jordi Benach,

Department of Biological Sciences and the Northeast Structural Genomics Consortium Columbia University, New York, NY, 10032, USA

Bomina Yu,

Department of Biological Sciences and the Northeast Structural Genomics Consortium Columbia University, New York, NY, 10032, USA

Mariam Abashidze,

Department of Biological Sciences and the Northeast Structural Genomics Consortium Columbia University, New York, NY, 10032, USA

Jayaraman Seetharaman,

Department of Biological Sciences and the Northeast Structural Genomics Consortium Columbia University, New York, NY, 10032, USA

Rong Xiao,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and the Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, and Robert Wood Johnson Medical School, UMDNJ, Piscataway, NJ 08854, USA

Thomas B. Acton,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and the Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, and Robert Wood Johnson Medical School, UMDNJ, Piscataway, NJ 08854, USA

Liang Tong, and

Department of Biological Sciences and the Northeast Structural Genomics Consortium Columbia University, New York, NY, 10032, USA

Gaetano T. Montelione

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and the Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, and Robert Wood Johnson Medical School, UMDNJ, Piscataway, NJ 08854, USA. Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, NJ 08854, USA

Gaetano T. Montelione: guy@cabm.rutgers.edu

Abstract

Protein domain family PF06855 (DUF1250) is a family of small domains of unknown function found only in bacteria, and mostly in the order *Bacillales* and *Lactobacillales*. Here we describe the solution NMR or X-ray crystal structures of three representatives of this domain family, MW0776 and MW1311 from *Staphylococcus aureus* and yozE from *Bacillus subtilis*. All three proteins adopt a four-helix motif similar to sterile alpha motif (SAM) domains. Phylogenetic analysis classifies MW1311 and yozE as functionally equivalent proteins of the UPF0346 family of unknown function, but excludes MW0776, which likely has a different biological function. Our structural characterization of the three domains supports this separation of function. The structures of MW0776, MW1311, and yozE constitute the first structural representatives from this protein domain family.

Keywords

yozE; MW0776; MW1311; UPF0345; DUF1250; PF06855; MW0776; SAM domain; X-ray crystal structure; solution NMR; structural genomics

Introduction

Orthologs are homologs in different organisms that have the same biological function. In many applications it is critical to distinguish orthologous relationships from simple homology. Several important tools have been developed to identify orthologous relationships. COGs are Clusters of Orthologous Groups of proteins, representing reciprocal “best hits” between protein sequences of multiple genomes [1]. Classes of Reciprocal Sequence Homologues (CRSH) (www.orthology.org) are groups of homologous proteins likely to have equivalent biological functions. These classes are identified based on three criteria: (i) reconstructed orthology, that is relationship via speciation rather than in-genome duplication, (ii) a relative similarity between different organisms that is consistent with other orthologous groups, and (iii) the absence of large insertions or deletions in the protein sequence.

Orthologous proteins should also share common structural features such as protein folds and surface electrostatics. The NIH Protein Structure Initiative has included a significant effort in structural coverage of large, broadly conserved protein domain families. As part of this effort, the Northeast Structural Genomics Consortium (NESG) targeted multiple proteins from Pfam domain families for which no structural representatives were known at the time of target selection. Structural descriptions of protein domain families could then complement and support the identification of orthologous proteins.

In this paper, we describe the three-dimensional structures of three proteins from protein domain family PF06855 (DUF1250) [2]: namely MW1311 (UniProtKB/TrEMBL ID: Q7A0W3 or Y1311_STAAW; NESG ID: ZR218), MW0776 (UniProtKB/TrEMBL ID: Q7A1E8_STAAW; NESG ID: ZR215) and yozE (UniProtKB/TrEMBL ID: O31864 or YOZE_BACSU; NESG ID: SR391). This domain family consists of hypothetical, acidic (pI ~4.5) proteins approximately 70 residues in length. Sequence alignment reveals that there is 24% sequence identity between yozE and MW1311, 20% between MW0776 and MW1311, and 19% between MW0776 and yozE (Fig. 1a). However, despite the sequence similarity between these proteins, the CRSH database of functionally equivalent bacterial proteins segregates yozE and MW0776 from MW1311, suggesting that they belong to two different families of orthologous proteins, both of which are comprised of domains of unknown function. The structures described here support the identification of MW1311 and yozE as orthologs, and the suggestion from CRSH analysis that MW0776, which has significantly different surface charge distribution, may have a different biological function.

Material and Methods

Cloning, expression and protein purification

Isotopically-enriched samples of MW0776 and yozE for NMR spectroscopy and selenomethionine-labeled MW1311 for X-ray crystallography were cloned, expressed, and purified following standard NESG protocols [3]. Specifically, the genes for MW0776 (UniProtKB/TrEMBL ID, Q7A1E8_STAAW) and MW1311 (UniProtKB/TrEMBL ID, Q7A0W3 or Y1311_STAAW) from *Staphylococcus aureus* and yozE (UniProtKB/TrEMBL ID, O31864 or YOZE_BACSU) from *Bacillus subtilis* were amplified from genomic DNA and cloned into the pET21_NESG vector (Novagen) in frame with a C-terminal affinity tag (LEHHHHHH). MW0776 had an additional 5 residues (MAGDP) added to the N-terminus. The resulting plasmids pZR215-21.1, pSR391-21.2, and pZR218-21.1, coding for the recombinant 80 residue MW0776, 82 residue yozE and 81 residue MW1311 proteins, respectively, were transformed into *E. coli* BL21(DE3) pMGK cells. Proteins were expressed overnight at 17 °C in MJ9 minimal media containing $U\text{-}(^{15}\text{NH}_4)_2\text{SO}_4$ and $U\text{-}^{13}\text{C}$ -glucose as the sole nitrogen and carbon sources for the isotopically-enriched NMR samples [4], and in MJ9 minimal media supplemented with selenomethionine, lysine, and phenylalanine, threonine, isoleucine, leucine, and valine for selenomethionine-labeled samples [5]. Proteins were purified using an ÄKTAexpress system (GE Healthcare) with a two-step protocol consisting of IMAC (HisTRAP HP) and gel filtration (HiLoad 26/60 Superdex 75) chromatography.

Final yields of 10 mg of [U - ^{13}C , ^{15}N]-MW0776 and 11 mg of [U -5%- ^{13}C , 100%- ^{15}N]-MW0776 were obtained per liter of cell culture and were concentrated to 0.7–0.9 mM in 10% $^2\text{H}_2\text{O}$, 50 μM DSS, 100 mM NaCl, 5 mM DTT, 0.02% (w/v) NaN_3 , 10 mM Tris-Cl pH 7.5. MW0776 is sensitive to pH changes and at a pH below 7.5, the sample heavily precipitated. Protein purity and molecular mass were confirmed by SDS-PAGE and MALDI-TOF mass spectrometry (9844.7920 Da observed vs. 9972.59 Da expected for [U - ^{13}C , ^{15}N]-MW0776, and 9459.6768 Da observed vs. 9578.54 Da expected for [U -5%- ^{13}C , 100%- ^{15}N]-MW0776).

Final yields of 67 mg of [U - ^{13}C , ^{15}N]-yozE and 50 mg of [U -5%- ^{13}C , 100%- ^{15}N]-yozE were obtained per liter of cell culture and were each concentrated to 1.35–1.5 mM in 10% $^2\text{H}_2\text{O}$, 50 μM DSS, 100 mM NaCl, 5 mM CaCl_2 , 10 mM DTT, 0.02% (w/v) NaN_3 , 20 mM NH_4OAc , pH 5.5. Protein purity and molecular mass were confirmed by SDS-PAGE and MALDI-TOF mass spectrometry (10545.16 Da observed vs. 10544.60 Da expected for [U - ^{13}C , ^{15}N]-yozE, and 10110.08 Da observed vs. 10120.39 Da expected for [U -5%- ^{13}C , 100%- ^{15}N]-yozE).

A final yield of 26 mg of selenomethionine-labeled MW1311 was obtained per liter of cell culture and concentrated to 12 mg/mL in 100 mM NaCl, 5 mM DTT, 0.02% (w/v) NaN_3 , 10 mM Tris-Cl pH 7.5. Protein purity and molecular mass were confirmed by SDS-PAGE and MALDI-TOF mass spectrometry (10076.89 Da observed vs. 10075.50 Da expected for the selenomethionine-labeled protein).

NMR spectroscopy and structure refinement of MW0776 and yozE

All NMR data for resonance assignment and structure determination were collected at 293 K on Bruker AVANCE 800 MHz or Varian INOVA 600 MHz spectrometers equipped with 5 mm TXI cryo probes, processed with *NMRPipe* [6] and visualized using SPARKY (Goddard and Kneller, University of California, San Francisco). All spectra were referenced to internal DSS. Complete ^1H , ^{13}C , and ^{15}N resonance assignments for MW0776 and yozE were determined using conventional triple resonance NMR methods. Backbone resonance assignments were made by *AutoAssign 2.1* (MW0776) or *AutoAssign 1.17* (yozE) [7] using peak lists for 2D ^1H - ^{15}N HSQC and 3D HNCOC, HNCA HN(CO)CACB and HNCACB spectra. Side chain assignment for MW0776 was completed manually using 3D HBHA(CO)NH, and (H)CCH-TOCSY experiments, and 3D HA(CACO)NH, HCCH-COSY, and C(CCO)NH-TOCSY experiments for yozE. Three-bond $^3J(\text{H}^{\text{N}}-\text{H}^{\alpha})$ scalar couplings were determined using the 3D HNHA experiment [8]. Stereospecific isopropyl methyl assignments for all Val and Leu residues were deduced from characteristic cross-peak fine structures in high resolution 2D ^1H - ^{13}C HSQC spectra of [U -5%- ^{13}C , 100%- ^{15}N]-MW0776 or [U -5%- ^{13}C , 100%- ^{15}N]-yozE [9]. Resonance assignments were validated using the *Assignment Validation Suite* (AVS) software package [10] and deposited in the BioMagResDB (BMRB accession number 16793 and 6976 for MW0776 and yozE, respectively).

The solution NMR structure of MW0776 was calculated using *CYANA 2.1* [11] supplied with peak intensities from a 3D simultaneous CN NOESY [12] ($\tau_{\text{m}} = 100$ ms) spectrum, together with broad dihedral angle constraints computed by *TALOS* [13] for ordered residues

with confidence scores of 10. The 20 structures with lowest target function out of 400 in the final cycle calculated were further refined by restrained molecular dynamics in explicit water using *CNS 1.2* [14, 15] and the PARAM19 force field, using the final NOE derived distance constraints, *TALOS* dihedral angle constraints, and hydrogen bond constraints derived from *AutoStructure 2.2.1* [16] and *CYANA*. The final refined ensemble of 20 structures was deposited into the Protein Data Bank (PDB ID 2KVS) and BioMagResDB (BMRB accession number, 16793).

The solution NMR structure of yozE was calculated using *AutoStructure 2.1.1* [16] supplied with peak intensities from a 3D ^{13}C edited NOESY and ^{15}N edited NOESY ($\tau_m = 80$ ms) spectrum, together with broad dihedral angle constraints computed by *HYPER* [17] and *TALOS* [13] ($\phi \pm 30^\circ$; $\psi \pm 30^\circ$) for ordered residues with confidence scores of 10. The 10 structures with lowest target function out of 56 in the final cycle calculated were further refined by restrained molecular dynamics in explicit water using *CNS 1.1* [14, 15] and the PARAM19 force field, using the final NOE derived distance constraints, *TALOS* dihedral angle constraints, and hydrogen bond constraints derived from *AutoStructure 2.1.1* [16]. The final refined ensemble of 10 structures was deposited into the Protein Data Bank (PDB ID 2FJ6) and BioMagResDB (BMRB accession number, 6976).

Structural statistics and global structure quality factors (Table 1), including *Verify3D* [18], *ProsaII* [19], *PROCHECK* [20], and *MolProbity* [21] raw and statistical Z-scores, were computed using the *PSVS 1.4* software package [22]. The global goodness-of-fit of the final structure ensemble with the NOESY peak list data and resonance assignments was determined using the *RPF* analysis program [23].

Crystallization, data collection and structure refinement of MW1311

Crystallization screening of MW1311 was carried out at the Hauptmann-Woodward Institute high-throughput screening facility [24] and then further optimized manually by hanging drop vapor diffusion reactions at 18 °C. The protein stock solution described above was mixed 1:1 with the precipitant containing 300 mM NH_4F , 18% PEG-3350, 50 mM MES, pH 6.5. Crystals of MW1311 appeared overnight and grew to full size in several days. The crystals were cryoprotected with paratone-N, prior to flash freezing in liquid propane for data collection at 100 K. A multi-wavelength anomalous diffraction (MAD) data set to 2.1 Å resolution was collected from a single crystal on beamline X4A at the National Synchrotron Light Source at Brookhaven National Laboratory. The diffraction images were processed with *HKL-2000* [25] (Table 2).

The crystal belongs to the orthorhombic space group $P2_12_12_1$ with cell dimensions of $a=29.22$ Å, $b=52.16$ Å, $c=115.27$ Å. Assuming two protomers per asymmetric unit, the V_m value is 2.2 Å³/Da, which is within the most probably range of values for proteins [26]. Of the 6 possible selenium sites in the ASU of the crystal, 4 were located with the program *BNP* [27]. *SOLVE/RESOLVE* [28] was used for phasing the reflections and automated model building, which correctly placed 75% of the residues with side chains in each protomer of ASU. The majority of the model, including the initiating selenomethionines, was manually built with the program *COOT* [29]. Further refinement involved iterations of manual model-building in *COOT* [29] followed by computational refinement in *CNS* [15] (Table 2) using

standard stereochemical restraints [30] in conjunction with a randomly selected R_{free} set comprising ~10% of the reflections. B-factors were subject to vicinal restraints (1.5–2.0 Å² and 2.0–2.5 Å² for main-chain and side-chain atoms, respectively). Strong noncrystallographic symmetry restraints (1050 kJ/Å², $\sigma_B = 1.5$ for main chain atoms and 418 kJ/Å², $\sigma_B = 2.5$ for side chain atoms) initially applied throughout the model were ultimately released completely to produce a considerable improvement in R_{free} . Water-molecules automatically identified using *CNS* were checked for consistency with $2F_o - F_c$ maps and hydrogen-bonding criteria.

The quality of the final structure was assessed with *PSVS* [22]. All residues were found in the most favored or additionally allowed regions of the Ramachandran Plot (Table 2). The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB ID 2O6K).

Results and Discussion

In the CRSH database of functionally equivalent bacterial proteins (www.orthology.org), MW1311 and yozE belong to the same family of orthologous domains containing the UPF0346 proteins of unknown function. This eubacterial family is found only in 2 classes (*Bacillales* and *Lactobacillales*) and 7 genera. Its representation in no archaeal organisms and only 7 eubacterial organisms, representing less than 2% of the >400 present in that database, suggests this protein family is unlikely to function in a conserved metabolic pathway and is more likely to function in secondary metabolism. In about half of the organisms containing an ortholog of UPF0346, the corresponding gene is located on the chromosome in close proximity to degV/degVA_1, ypmR/ycsK_1, and ypmS/yfaA_1. DegV is a hypothetical fatty acid binding protein [31], YpmR/ycsK_1 belongs to the ‘GDSL’ lipolytic enzyme family [32], and YpmS/YfaA_1 is a putative membrane-associated protein [33]. These protein families are also limited to *Bacillales* and *Lactobacillales* and the consistent co-occurrence, and often co-localization, of all four genes suggests that they may be involved in a similar pathway. However, the same phylogenetic analysis excludes MW0776 from the UPF0346 family and weakly suggests it belongs to a different family, also of unknown function, which includes yahF, a putative acyl-coA synthetase [34].

Three-dimensional structural analysis was used to complement the CRSH analysis for identification of orthologous proteins. The structures of MW1311 and MW0776 consist of a four α -helix bundle comprised of two orthogonal helix-turn-helix motifs, or helix hairpins (Fig. 1b,c), similar to the SAM domain-like fold in SCOP [35]. yozE has an additional small helix in the turn between the helices of the second helix hairpin (Fig. 1d). The two hairpins are tethered by a 8–10 residue loop. This loop is disordered in the NMR solution structure of MW0776 (Fig. 1c), but is well defined in the structures of MW1311 and yozE (Fig. 1b,d). In fact, the main helices in the MW1311 and yozE structures align very well; least-squares superposition gives a root-mean-square-deviation (rmsd) of 1.3 Å for alignment of 63 Ca atoms (Fig. 1e). MW0776 differs from the MW1311 and yozE structures in the orientation of the N- and C-terminal helices, with an overall rmsd of 3.2 and 2.9 Å with MW1311 (65 atoms) and yozE (70 atoms), respectively (Fig. 1e). The electrostatic surface potential of the

MW1311 and yozE protomers reveals a predominately negatively charged face on the side of the protein comprised of helices $\alpha 3$ and $\alpha 4$ (Fig. 1f,h). The opposite face of the protein, comprised of $\alpha 1$ and $\alpha 2$, is predominately positively charged (Fig. 1h). Surprisingly, this is the complete reverse of what is seen in the MW0776 structure (Fig. 1g).

Independent structural alignment analyses using the program DALI [36] show that MW0776, MW1311, and yozE are the most similar domains of known structure to each other when either one is used as the search model with Z-scores of 9.6 between MW1311 and yozE and ~5 between either of those and MW0776. Similar domains of known function were found to be the RNA-binding domains of the signal recognition particle (SRP, PDB id 2PXT) [37] and T7 RNA polymerase (PDB id 1H38) [38], with DALI Z-scores ranging from 3.2 to 4.8 for the different alignments. Structural alignment with SRP shows that the corresponding

RNA-binding region is negatively charged in MW0776, MW1311, and yozE (Fig. 1i), making it unlikely for these proteins to be able to bind RNA in the same way. Protein-protein and protein-RNA interactions *via* SAM domains have been reported for numerous proteins with diverse functions [39]. The charged patches on the surface of MW0776, MW1311, and yozE suggest that these proteins may also participate in protein-protein interactions. Analytical gel filtration chromatography with in-line static light-scattering and refractive index detectors show that the proteins are monomeric under the conditions used in the NMR studies or in the crystallization stock. However, the observed monomers of MW0776, MW1311, and yozE may homo- or hetero-oligomerize under different conditions and/or in the presence of binding partners. For example, MW1311 or yozE could have some role in fatty acid metabolism and interact with degV, ympR, or ypmS, given the co-occurrence and co-localization of these genes in ~80% of organisms containing an ortholog (www.orthology.org)

Summary

We have determined the X-ray crystal structure of MW1311 from *S. aureus* and the solution NMR structures of MW0776 from *S. aureus* and yozE from *B. subtilis*. These are the first structural representatives of the Pfam domain family PF06855. Phylogenetic and structural analyses indicate that this domain family can be further divided into at least two different subdomain families: the UPF0346 family comprised of MW1311 and yozE, and another family comprised of MW0776. However, the exact biological functions of these subdomains remain unknown.

Acknowledgments

The authors thank J.M. Aramini, M.C. Baran, C. Ciccocanti, K.E. Cunningham, J.K. Everett, Y. Fang, H. Janjua, L.-C. Ma, B. Rost, K. Shetty, and D. Wang for valuable scientific discussions and/or technical support, A. Laucella and G. DeTitta for performing initial crystallization screening, and R. Abramowitz and J. Schwarnof for access to beamline X4A at Brookhaven National Laboratory. The NESG expression vectors SR391-21.2, ZR218-21.1, and ZR215-21.1 have been deposited in the PSI Materials Repository (<http://psimr.asu.edu/>). This research was supported by the Protein Structure Initiative of the US National Institutes of Health, grants U54 GM074958 and U54 GM094597.

Abbreviations

NESG	Northeast Structural Genomics Consortium
PDB	Protein Data Bank
rmsd	root-mean-square-deviation

References

1. Tatusov RL, Koonin EV, Lipman DJ. Science. 1997; 278:631–637. [PubMed: 9381173]
2. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A. Nucleic Acids Res. 2010; 38:D211–222. [PubMed: 19920124]
3. Xiao R, Anderson S, Aramini J, Belote R, Buchwald WA, Ciccocanti C, Conover K, Everett JK, Hamilton K, Huang YJ, Janjua H, Jiang M, Kornhaber GJ, Lee DY, Locke JY, Ma LC, Maglaqui M, Mao L, Mitra S, Patel D, Rossi P, Sahdev S, Sharma S, Shastry R, Swapna GVT, Tong SN, Wang D, Wang H, Zhao L, Montelione GT, Acton TB. J Struct Biol. 2010; 172:21–33. [PubMed: 20688167]
4. Jansson M, Li YC, Jendeborg L, Anderson S, Montelione GT, Nilsson B. J Biomol NMR. 1996; 7:131–141. [PubMed: 8616269]
5. Doublet S, Kapp U, Aberg A, Brown K, Strub K, Cusack S. FEBS Lett. 1996; 384:219–221. [PubMed: 8617357]
6. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. J Biomolecular NMR. 1995; 6:277–293.
7. Moseley HN, Monleon D, Montelione GT. Methods Enzymol. 2001; 339:91–108. [PubMed: 11462827]
8. Vuister GW, Bax A. J Am Chem Soc. 1993; 115:7772–7777.
9. Neri D, Szyperski T, Otting G, Senn H, Wuthrich K. Biochemistry. 1989; 28:7510–7516. [PubMed: 2692701]
10. Moseley HN, Sahota G, Montelione GT. J Biomol NMR. 2004; 28:341–355. [PubMed: 14872126]
11. Herrmann T, Guntert P, Wuthrich K. J Mol Biol. 2002; 319:209–227. [PubMed: 12051947]
12. Pascal SM, Muhandiram DR, Yamazaki T, Forman-Kay JD, Kay LE. J Magn Reson Ser B. 1994; 103:197–201.
13. Cornilescu G, Delaglio F, Bax A. J Biomol NMR. 1999; 13:289–302. [PubMed: 10212987]
14. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Proteins. 2003; 50:496–506. [PubMed: 12557191]
15. 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. Acta Crystallogr D Biol Crystallogr. 1998; 54 (Pt 5):905–921. [PubMed: 9757107]
16. Huang YJ, Tejero R, Powers R, Montelione GT. Proteins. 2006; 62:587–603. [PubMed: 16374783]
17. Tejero R, Monleon D, Celda B, Powers R, Montelione GT. J Biomol NMR. 1999; 15:251–264. [PubMed: 10677828]
18. Luthy R, Bowie JU, Eisenberg D. Nature. 1992; 356:83–85. [PubMed: 1538787]
19. Sippl MJ. Proteins. 1993; 17:355–362. [PubMed: 8108378]
20. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. J Applied Crystallogr. 1993; 26:283–291.
21. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC. Nucleic Acids Res. 2007; 35:W375–383. [PubMed: 17452350]
22. Bhattacharya A, Tejero R, Montelione GT. Proteins. 2007; 66:778–795. [PubMed: 17186527]
23. Huang YJ, Powers R, Montelione GT. J Am Chem Soc. 2005; 127:1665–1674. [PubMed: 15701001]

24. Luft JR, Wolfley JR, Said MI, Nagel RM, Lauricella AM, Smith JL, Thayer MH, Veatch CK, Snell EH, Malkowski MG, Detitta GT. *Protein Sci.* 2007; 16:715–722. [PubMed: 17327388]
25. Otwinowski Z, Minor W. *Methods Enzymol.* 1997; 276:307–326.
26. Matthews BW. *J Mol Biol.* 1968; 33:491–497. [PubMed: 5700707]
27. Weeks CM, Miller R. *Acta crystallographica Section D, Biological crystallography.* 1999; 55:492–500.
28. Terwilliger TC. *Methods Enzymol.* 2003; 374:22–37. [PubMed: 14696367]
29. Cowtan EP. *Acta Crystallogr D Biol Crystallogr.* 2004; D60:2126–2132. [PubMed: 15572765]
30. Engh RA, Huber R. *Acta Crystallogr A.* 1991; 47:392–400.
31. Schulze-Gahmen U, Pelaschier J, Yokota H, Kim R, Kim SH. *Proteins.* 2003; 50:526–530. [PubMed: 12577257]
32. Masayama A, Kuwana R, Takamatsu H, Hemmi H, Yoshimura T, Watabe K, Moriyama R. *J Bacteriol.* 2007; 189:2369–2375. [PubMed: 17220230]
33. Beres SB, Richter EW, Nagiec MJ, Sumby P, Porcella SF, DeLeo FR, Musser JM. *Proc Natl Acad Sci U S A.* 2006; 103:7059–7064. [PubMed: 16636287]
34. Johnson TJ, Kariyawasam S, Wannemuehler Y, Mangiamale P, Johnson SJ, Doetkott C, Skyberg JA, Lynne AM, Johnson JR, Nolan LK. *J Bacteriol.* 2007; 189:3228–3236. [PubMed: 17293413]
35. Murzin AG, Brenner SE, Hubbard T, Chothia C. *J Mol Biol.* 1995; 247:536–540. [PubMed: 7723011]
36. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. *Bioinformatics.* 2008; 24:2780–2781. [PubMed: 18818215]
37. Keel AY, Rambo RP, Batey RT, Kieft JS. *Structure.* 2007; 15:761–772. [PubMed: 17637337]
38. Tahirov TH, Temiakov D, Anikin M, Patlan V, McAllister WT, Vassilyev DG, Yokoyama S. *Nature.* 2002; 420:43–50. [PubMed: 12422209]
39. Qiao F, Bowie JU. *Sci STKE.* 2005; 2005:re7. [PubMed: 15928333]
40. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. *Nucleic Acids Res.* 2003; 31:3497–3500. [PubMed: 12824352]
41. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. *Proc Natl Acad Sci U S A.* 2001; 98:10037–10041. [PubMed: 11517324]

Fig. 1. Structural representatives of the PF06855 protein domain family

(A) Sequence alignment of MW1311 and MW0776 from *Staphylococcus aureus* and yozE from *Bacillus subtilis* by ClustalW [40]. Boxed residues colored shaded red and white represent residues that have 100% or 50% sequence identity, respectively. The secondary structure elements found in the crystal structure of MW1311 are shown above the alignment. (B) The crystal structure of MW1311 from *Staphylococcus aureus*. The two protomers found in the asymmetric unit are shown. Each chain is colored from blue to red from its N-to C-terminus. The α -helices are labeled in one of the monomers. (C) Stereoview of the final ensemble of 20 conformers from the solution NMR structure of MW0776 from *Staphylococcus aureus*. The chain is colored from N-to C-terminus according to panel B. (D) Stereoview of the final ensemble of 10 conformers from the solution NMR structure of yozE from *Bacillus subtilis*. The chain is colored from N- to C-terminus according to panel B. (E) Least-squares superposition of MW1311 (colored from N-to C-terminus according to panel B), MW0776 (pink) and yozE (grey). The aligned proteins are rotated $\sim 90^\circ$ to the left as that shown in panels C and D to better view the structural differences between the proteins. Electrostatic surface potential [41] of (F) MW1311 (shown as in panel B), (G) MW0776,

and **(H)** yozE. The left of panel G and H show the protein orientated as in panel C and D, respectively. Fully saturated red and blue colors represent, respectively, negative and positive potentials of ± 5 kT at an ionic strength of 0.2 M. **(I)** Least-squares superposition of yozE (electrostatic potential colored as in panel H) and the signal recognition particle (SRP, PDB id 2PXT) [37]. Only the RNA of the SRP structure is shown.

Table 1Summary of NMR and structural statistics^a

	MW0076		yozE	
Completeness of resonance assignments ^b :				
Backbone (%)	94		95.3	
Side chain (%)	92		88.2	
Aromatic (%)	82		100	
Stereospecific methyl (%)	100		100	
Conformationally-restricting constraints:				
Distance constraints				
Total	971		1083	
intra-residue ($i = j$)	93		165	
sequential ($ i - j = 1$)	301		305	
medium range ($1 < i - j < 5$)	378		335	
long range ($ i - j \geq 5$)	199		278	
Dihedral angle constraints	107		208	
Hydrogen bond constraints	52		42	
No. of constraints per residue	13.1		16.7	
No. of long range constraints per residue	2.7		3.5	
Residual constraint violations ^c :				
Average No. of distance violations per structure:				
0.1 – 0.2 Å	2.1		9.1	
0.2 – 0.5 Å	3.2		2.5	
> 0.5 Å	4.5		0	
Average No. of dihedral angle violations per structure:				
1 – 10°	0		3.6	
> 10°	0		0	
Model Quality ^{c,d} :				
RMSD backbone atoms (Å)	1.0		0.5	
RMSD heavy atoms (Å)	1.5		0.9	
RMSD bond lengths (Å)	0.009		0.011	
RMSD bond angles (°)	0.6		1.5	
MolProbity Ramachandran statistics:				
most favored regions (%)	94.1		92.0	
allowed regions (%)	5.5		6.9	
disallowed regions (%)	0.4		1.1	
Global quality scores (Raw/Z-score):				
Verify3D	0.23	−3.69	0.34	−1.93
ProsaII	0.28	−1.53	0.46	−0.79
Procheck (phi-psi) ^d	0.11	0.75	0.10	0.71
Procheck (all) ^d	0.04	0.24	−0.20	−1.18

	MW0076		yozE	
Molprobability clash	13.45	-0.78	16.16	-1.25
RPF Scores ^e :				
Recall/Precision	0.952	0.928	0.960	0.797
F-measure/DP-score	0.94	0.794	0.871	0.758
Model Contents:				
Ordered residue ranges ^d	8–12, 18–30, 35–37, 43–77		4–11, 18–29, 32–49, 51–73	
BMRB accession code:	16793		6976	
PDB accession code:	2KVS		2FJ6	

^aStructural statistics were computed for the ensemble of 20 (MW0776) or 10 (yozE) deposited structures.

^bComputed using AVS software [10] 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 6-His tag.

^cCalculated using PSVS 1.4 program [22]. Average distance violations were calculated using the sum over r^{-6} (MW0776) or center averaging (yozE). The abbreviation RMSD represents root mean square deviation.

^dBased on ordered residue ranges [S(phi) + S(psi) > 1.8].

^eRPF scores [23] reflecting the goodness-of-fit of the final ensemble of structures (all residues with chemical shift assignments are included in the analysis).

Table 2**X-ray data and refinement statistics^a**

	MW1311
Crystal Parameters	
Space group	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions:	
a, b, c (Å)	29.22, 52.16, 115.27
α β, γ, (°)	90, 90, 90
Matthews coefficient (Å ³ /Da)	2.18
Solvent Content (%)	43.53
Data Collection^b	
Wavelength (Å)	0.97916, 0.97949, 0.96863
Resolution (Å)	20.0 -2.1 (2.12 – 2.10)
R _{sym} (%) ^b	10.4 (41.5)
No. of unique reflections	18638
No. of reflections in R _{free} set	1771
Mean Redundancy	6.0 (5.7)
Overall completeness (%)	99.6 (99.9)
Mean I/σ _I	17.4 (8.6)
Refinement Residuals^c (f 2σ_f):	
R _{free} (%)	21.7
R _{work} (%)	20.8
Completeness, I 2σ _I (%)	93.9 (85.7)
Model Quality^d	
RMSD bond lengths (Å)	0.006
RMSD bond angles (°)	1.1
MolProbity Ramachandran statistics:	
most favored (%)	97.9
allowed (%)	2.1
disallowed (%)	0.0
Mean main chain B-factor (Å ²)	21.1
Mean overall B-factor (Å ²)	22.7
Mean solvent B-factor (Å ²)	30.0
Model Contents	
Protomers in ASU	2
Protein residues	A & B: 4–73
No. of protein atoms	1258
No. of water molecules	82
PDB accession code:	2O6K

^aEntries in parentheses report data from the limiting resolution shell. Data collection and refinement statistics come from *SCALEPACK* [25] and *CNS* [15] respectively. The abbreviations RMSD and ASU stand for root-mean-square deviation and asymmetric unit, respectively.

^b All observations with $I \geq 3\sigma_I$ were merged and included in calculating data quality statistics.

^c Reflections with $\sigma_f \leq 1.34$ were included in calculating R-factors.

^d Calculated using PSVS 1.4 program [22].