

February, 2011

# Structures of Domains I and IV from YbbR are Representative of a Widely Distributed Protein Family

Adam Barb, *University of Georgia*

John R. Cort, *Pacific Northwest National Laboratory*

Jayaraman Seetharaman

Scott Lew

Hsiau-Wei Lee, *University of Georgia*, et al.

# Structures of domains I and IV from YbbR are representative of a widely distributed protein family

Adam W. Barb,<sup>1,2</sup> John R. Cort,<sup>2,3</sup> Jayaraman Seetharaman,<sup>2,4</sup> Scott Lew,<sup>2,4</sup> Hsiau-Wei Lee,<sup>1,2</sup> Thomas Acton,<sup>2,5,6</sup> Rong Xiao,<sup>2,5,6</sup> Michael A. Kennedy,<sup>2,7</sup> Liang Tong,<sup>2,4</sup> Gaetano T. Montelione,<sup>2,5,6</sup> and James H. Prestegard<sup>1,2\*</sup>

<sup>1</sup>Complex Carbohydrate Research Center, The University of Georgia, Athens, Georgia

<sup>2</sup>Northeast Structural Genomics Consortium, The State University of New Jersey, Rutgers, Piscataway, New Jersey 08854

<sup>3</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington

<sup>4</sup>Department of Biological Sciences, Columbia University, New York, New York

<sup>5</sup>Department of Molecular Biology and Biochemistry, The State University of New Jersey, Rutgers, Piscataway, New Jersey 08854

<sup>6</sup>Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

<sup>7</sup>Department of Biochemistry, Miami University, Miami, Ohio

Received 17 September 2010; Revised 19 November 2010; Accepted 19 November 2010

DOI: 10.1002/pro.571

Published online 8 December 2010 proteinscience.org

**Abstract:** YbbR domains are widespread throughout Eubacteria and are expressed as monomeric units, linked in tandem repeats or cotranslated with other domains. Although the precise role of these domains remains undefined, the location of the multiple YbbR domain-encoding *ybbR* gene in the *Bacillus subtilis* *glmM* operon and its previous identification as a substrate for a surfactin-type phosphopantetheinyl transferase suggests a role in cell growth, division, and virulence. To further characterize the YbbR domains, structures of two of the four domains (I and IV) from the YbbR-like protein of *Desulfitobacterium hafniense* Y51 were solved by solution nuclear magnetic resonance and X-ray crystallography. The structures show the domains to have nearly identical topologies despite a low amino acid identity (23%). The topology is dominated by  $\beta$ -strands, roughly following a “figure 8” pattern with some strands coiling around the domain perimeter and others crossing the center. A similar topology is found in the C-terminal domain of two stress-responsive bacterial ribosomal proteins, TL5 and L25. Based on these models, a structurally guided amino acid alignment identifies features of the YbbR domains that are not evident from naïve amino acid sequence alignments. A structurally conserved *cis*-proline (*cis*-Pro) residue was identified in both domains, though the local structure in the immediate vicinities surrounding this residue differed between the two models. The conservation and location of this *cis*-Pro, plus anchoring Val residues, suggest this motif may be significant to protein function.

**Keywords:** structural homolog; *cis*-proline; functional annotation; tandem domain

**Abbreviations:** *glmM*, phosphoglucosamine mutase; HSQC, heteronuclear single quantum coherence; NESG, Northeast Structural Genomics Consortium; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PDB, Protein Data Bank; PSVS, Protein Structure Validation Suite software; RDC, residual dipolar coupling; rmsd, root mean squared deviation; Sfp-PPTase, surfactin-type phosphopantetheinyl transferase.

Grant sponsor: National Institutes of Health Protein Structure Initiative; Grant number: U54-GM074958; Grant sponsor: A.W.B. was supported by a National Institutes of Health Ruth L. Kirschstein NRSA fellowship; Grant number: F32AR058084.

\*Correspondence to: James H. Prestegard, 315 Riverbend Road, Athens, GA 30602. E-mail: jpresteg@ccrc.uga.edu

## Introduction

Following closely on the heels of genome sequencing projects, the protein structure initiative has aimed to generate three-dimensional models of proteins with uncharacterized structures in an effort to provide representative structures in hundreds of protein families. It is anticipated that, in many cases, these can provide structural platforms to guide the annotation of genes with unknown function.<sup>1,2</sup> One of the 1200+ protein families selected for structural description, YbbR, shows little sequence similarity to other families and was predicted to represent a unique, undescribed portion of protein structural space.

YbbR domains were selected for study by the Northeast Structural Genomics (NESG) Consortium ([www.nesg.org](http://www.nesg.org)) based on their occurrence in a large, structurally uncharacterized domain family and the identification genes coding these domains in human gut metagenomics sequencing projects. YbbR domains were first identified as subunits of the larger YbbR protein of *Bacillus subtilis*. These domains are found in Eubacteria and are particularly common in genera of Gram-positive bacteria from the phylum Firmicutes but can be found in Gram-negative genera as well. These domains appear in soil-borne bacteria and extremophiles as well as human pathogens such as *Bacillus anthracis*, *Clostridium botulinum*, *Leptospira interrogans*, and *Staphylococcus aureus*. YbbR and YbbR-like proteins are classified as pfam<sup>3</sup> Protein Family PF07949, and occur in seven different arrangements from one lone domain to two, three, or four repeats or one, two, or three sequential domains following a DisA nucleotide-binding checkpoint domain (<http://pfam.sanger.ac.uk>). The *ybbR* locus of *B. subtilis* defines the common four domain, sequential repeat arrangement found in many *ybbR*-containing genomes and is located in the *glmM* operon immediately upstream of the glucosamine-1-phosphate mutase gene *glmM*.<sup>4</sup>

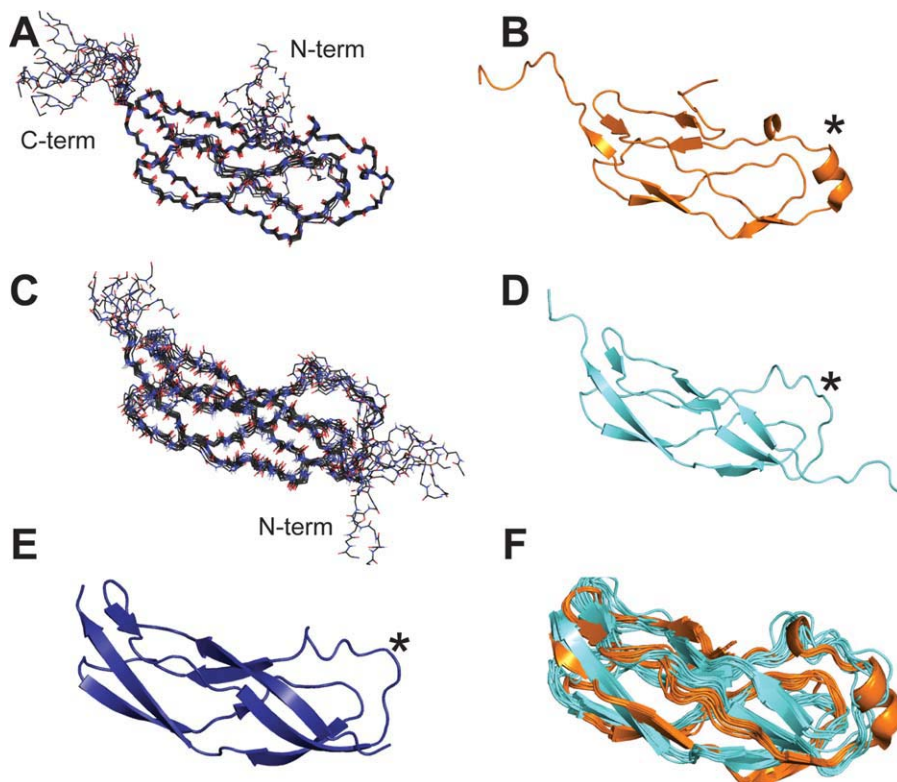
Despite the broad distribution of these domains and the essential role of the *glmM* operon in peptidoglycan biosynthesis, and thus cell growth and division,<sup>5</sup> the function(s) of the YbbR protein and the YbbR domains are unknown. YbbR domains are potentially important to the development of new antibacterial agents for two reasons: first, the proximity of *ybbR* to the *glmM* gene suggests that targeting YbbR domains may be part of a strategy to interrupt peptidoglycan synthesis, and second, the suggestion that YbbR domains are *in vivo* substrates for a surfactin-type phosphopantetheinyl transferase (Sfp-PPTase),<sup>6</sup> an important activator of nonribosomal peptidic virulence factor biosynthesis and a currently recognized antibiotic target,<sup>7,8</sup> suggests additional importance for targeting these domains in drug discovery efforts.

**Table I.** Summary of Conformationally Restricting Experimental Constraints for NMR Structures

	Domain I	Domain IV
NOE-based distance constraints		
Total	775	1127
Intraresidue [ $i = j$ ]	157	308
Sequential [ $ i - j  = 1$ ]	223	372
Medium range [ $1 <  i - j  < 5$ ]	73	148
Long-range [ $ i - j  \geq 5$ ]	322	299
NOE constraints per constrained residue	9.0	12.8
Dihedral-angle constraints	65	98
Residual dipolar coupling constraints		
Phage		
N: number (Q score)	n.d.	41 (0.1)
N-Co: number (Q score)	n.d.	49 (0.1)
Gel		
H-N: number (Q score)	64 (0.4)	43 (0.2)
PEG		
H-N: number (Q score)	61 (0.2)	n.d.
Total number of restricting constraints	965	1358
rmsd of backbone atoms		
Ordered regions <sup>a</sup>	0.6 Å	0.7 Å
Entire polypeptide	2.1 Å	2.1 Å
NOE violations per model >0.5Å	0.7	1.6
Dihedral angle violations per model >10°	0	0
rmsd		
Bond length	0.014 Å	0.021 Å
Bond angle	1.4°	1.5°
Ramachandran analysis		
Most favored region (%)	89.3%	88.5%
Allowed region (%)	10.7%	9.4%
Generously allowed (%)	0.0%	1.9%
PSVS <sup>11</sup> Z-scores		
Verify3D <sup>51</sup>	−2.41	−4.17
ProsaII <sup>52</sup>	−1.36	−2.61
Procheck (phi-psi) <sup>53</sup>	−1.97	−2.56
MolProbity <sup>54</sup>	−1.61	−0.47

<sup>a</sup> Ordered regions were Domain I (46): 9–17,24–26,31–37,50–55,66–72,75–78,82–91 and Domain IV (76): 9–24, 27–64, 67–68, 71–87, 89–91. n.d., not determined.

To structurally define the YbbR domains, structures of two of four predicted YbbR domains from the 437-residue YbbR-like protein Dhaf\_0833 from *Desulfitobacterium hafniense* were determined using solution nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography. *D. hafniense* is an anaerobic Gram-positive bacterium (phylum Firmicutes, class Clostridia) best known for its ability to degrade halogenated organic compounds.<sup>9,10</sup> Based on these structural studies, the global topology of the YbbR domain was identified, and unique structurally conserved characteristics common to the YbbR domain family were revealed. Structural similarity to proteins outside of the YbbR-like family was noted, possibly indicating more widespread occurrence of this domain in bacterial proteins.



**Figure 1.** Structures of Domain I and IV from the YbbR family protein of *Desulfotobacterium hafniense*. Panels A and C: The ensemble of 10 structures for Domain IV and I, respectively, as calculated from solution NMR spectroscopy-observed constraints. Nitrogen positions are depicted as blue sticks, and oxygen with red sticks. Panels B, D, and E: Ribbon diagrams of the lowest energy solution NMR structures of Domain IV (orange) and I (cyan) and the 1.9-Å X-ray crystal structure of Domain I (deep blue), respectively, show a unique topology that is identical for both of these domains. Panel F: An overlay of the solution structures of Domain I and IV show the similarity of these two molecules. The star marks the homology-modeled site of phosphopantetheine modification is identified by Walsh and coworkers (2005).

## Results

Structure-prediction tools utilizing YbbR domain primary sequences (mGenTHREADER and a BLAST search of the PDB) failed to identify significant homology to proteins deposited in the PDB. Plasmids containing different YbbR domains from different organisms were prepared, expressed in *E. coli*, and evaluated for structural characterization using solution NMR spectroscopy and X-ray crystallography. Of these, Domains I and IV of the YbbR-like protein of *D. hafniense* expressed with a high yield were soluble, stable, and either gave positive results in crystal trials or yielded  $^{15}\text{N}$ -heteronuclear single quantum coherence (HSQC) spectra consistent with proteins suitable for structural studies. Gel filtration chromatography indicated both Domains I and IV were monomeric (data not shown).  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  resonances of Domains I and IV were assigned and used to generate distance-dependent nuclear Overhauser effect (NOE) restraints. These distance constraints were combined with backbone dihedral angle estimates and residual dipolar coupling (RDC)-derived orientational constraints (Table I) for calculating structural models. In parallel, crystals of Domain I were obtained and found to diffract to 1.9 Å.

A complete dataset was collected, and the phases were determined using single wavelength anomalous diffraction of the selenium atoms incorporated as selenomethionine residues.

Models of the YbbR domains calculated from solution NMR measurements converged to an ensemble of structures with an root mean squared deviation (rmsd) for ordered backbone atoms of 0.6 Å for Domain I and 0.7 Å for Domain IV [Table I, Fig. 1(A,C)]. Protein Structure Validation Software Suite (PSVS) analysis<sup>11</sup> verified the high quality of these models with global structure quality Z-scores for the critical indices similar to those observed in high-resolution X-ray crystal structures (Table I). It is important to note that the ensemble rmsd determined by the PSVS package reflects atoms in regions of the protein that are constrained in the structure calculation (ordered residues). Residues that are not constrained due to a lack of measurable NOEs or RDCs appear disordered in the structural ensemble but should not be considered dynamic based on this classification. By this measure, the Domain IV models had a higher proportion of “ordered residues” with 76 residues used for the calculation compared with 46 residues with the models of Domain I (see the footnote in Table I). The

**Table II.** Summary of Crystallographic Information for Domain I

Space group	$P4_32_12$
Molecules per asymmetric unit	1
$V_M$	$2.11 \text{ \AA}^3 \text{ Da}^{-1}$
Unit cell parameters	$a = b = 44.8 \text{ \AA}$ $c = 86.0 \text{ \AA}$ $\alpha = \beta = \gamma = 90^\circ$
Resolution <sup>a</sup>	50–1.9 (1.93–1.90) $\text{\AA}$
Unique reflections	15491
Completeness	100 (100)
Redundancy	17.7 (14.3)
$R_{\text{merge}}$	0.058 (0.412)
$R_{\text{cryst}}$	0.224
$R_{\text{free}}$	0.258
No. of protein atoms	669
No. of water molecules	83
Average B factor	
Protein Main chain	$23.1 \text{ \AA}^2$
Protein Side chain	$32.5 \text{ \AA}^2$
Water molecules	$37.3 \text{ \AA}^2$
rmsd	
Bond lengths	$0.006 \text{ \AA}$
Bond angles	$1.60^\circ$
Ramachandran statistics	
Most favored region	91.8%
Allowed region	6.8%
Generously allowed	1.4%
PSVS <sup>11</sup> Z-scores	
Verify3D <sup>51</sup>	0.96
ProsaII <sup>52</sup>	–1.16
Procheck (phi-psi) <sup>53</sup>	–1.18
MolProbity <sup>54</sup>	–2.40

<sup>a</sup> values in parentheses are for the highest resolution shell.

X-ray diffraction derived model was likewise of high quality and has good global structure quality Z-scores [Fig. 1(E), Table II]. The NMR and X-ray structures of Domain I overlaid with a backbone rmsd over structured residues to 1.1  $\text{\AA}$ . Despite the moderate sequence identity between Domains I and IV of 23%, these domains were structurally very similar and the lowest energy models from NMR methods for the two domains overlaid with a global backbone rmsd of 2.2  $\text{\AA}$  [Fig. 1(F)].

The YbbR domains adopt the structure of elongated bent cylinders, roughly twice as long as their diameter, as shown in Figure 1(B,D,E). These domains are comprised almost entirely of  $\beta$ -pleated sheets and loops, though Domain IV had one short  $\alpha$ -helical segment [Fig. 1(B)]. The  $\beta$ -sheets were mostly antiparallel with a section of parallel  $\beta$ -sheet [Fig. 2(A)] as exemplified by residues 10–11 and 49–50 of Domain IV. The N-terminal segment of Domain I is well defined and nestled between two strands [Fig. 1(D)]; however, resonances belonging to the seven N-terminal residues of Domain IV were not observed in the spectra, and therefore, these residues did not converge in the ensemble of models [Fig. 1(A)]. The Domain IV N-terminal sequence was shortened relative to that of Domain I, as a result of construct optimization considerations used to

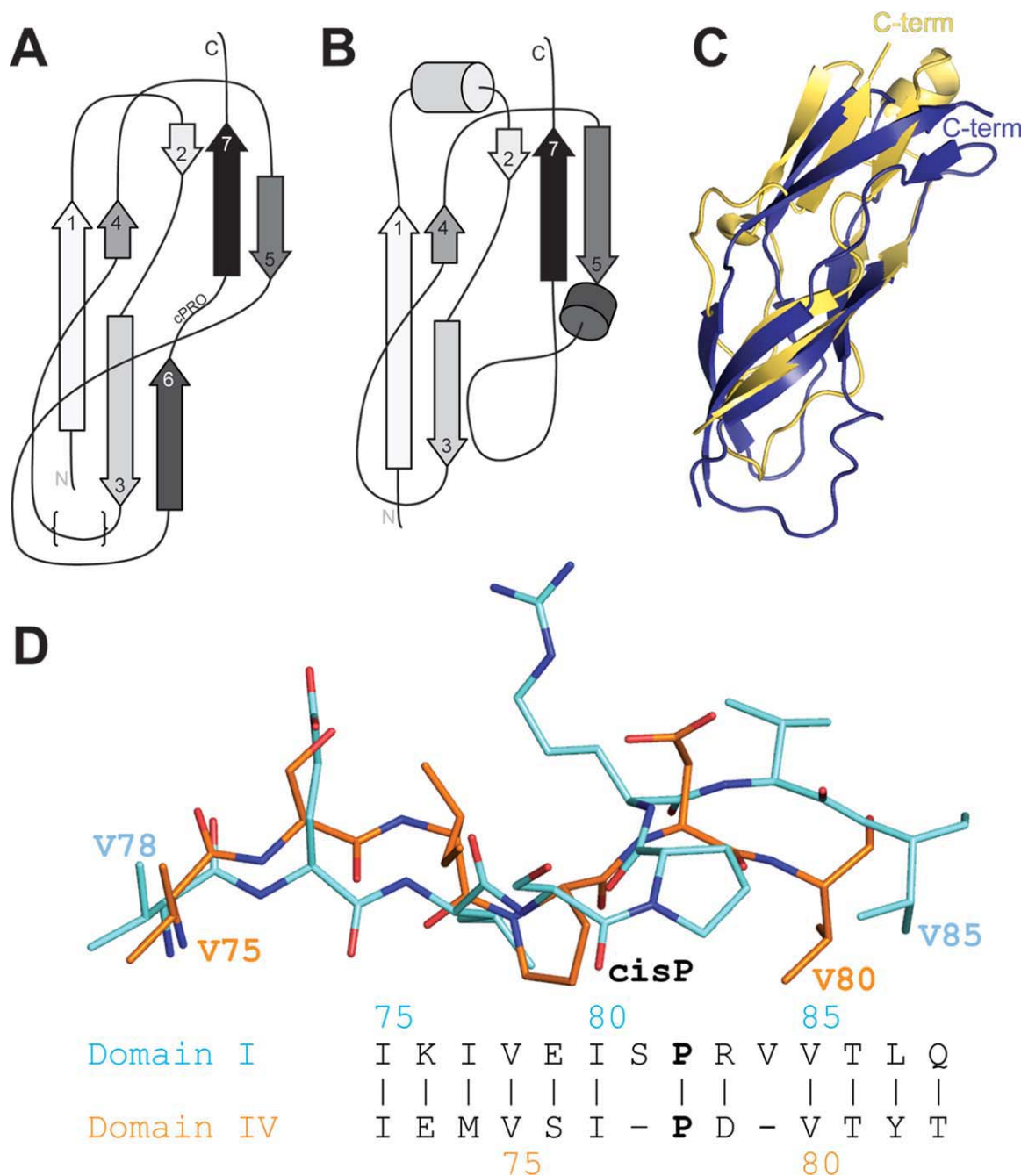
enhance success in protein sample production. In this case, the shortening may have precluded formation of stabilizing interactions and may account for the difference in rigidity.

The topology of the YbbR domain fold can be described as a “figure 8” with some strands coiling around the domain perimeter and others crossing the center as shown in Figure 2(A). A search for structurally similar proteins using the DALI and MarkUs bioinformatics tools<sup>12,13</sup> returned many matches to small segments of the YbbR domains. However, only two proteins from this list were topologically similar to the entire YbbR domain: the C-terminal domain of the TL5 and L25 ribosomal proteins from *Thermus thermophilus*<sup>14,15</sup> and *Deinococcus radiodurans*,<sup>16</sup> respectively. The amino acid sequence of L25 is 7.8 and 13.6% identical with Domain I and IV, respectively, and TL5 is 21.3 and 17% identical, respectively. TL5 and the Domain I X-ray structure can be superimposed with a backbone rmsd of 2.2  $\text{\AA}$  [Fig. 2(C)]. TL5 and Domain IV overlaid with an rmsd for backbone atoms of 3.8  $\text{\AA}$ . The general strand topology is conserved between the TL5 protein and both YbbR domains including the parallel  $\beta$ -sheets, though the loop regions are substantively different, and the TL5 protein lacks the sixth beta strand observed in the YbbR domains [Fig. 2(A,B)].

An interesting feature of the YbbR domains, not present in the TL5 and L25 proteins, is the presence of a *cis*-proline (*cis*-Pro) residue near the C-terminus (Domain I P82, Domain IV P78). Although the location of this *cis*-Pro conformation is conserved, the local structure in the immediate vicinity surrounding the *cis*-Pro differs in Domains I and IV, as shown in Figure 2(D). Domain IV has a shorter sequence surrounding the *cis*-Pro, and the extra residues in Domain I introduced bulges in the strand relative to Domain IV. Despite these differences, the locations of preceding and following Val residues (V78 and V85 for Domain I and V75 and V80 for Domain IV) were identical.

Three regions of Domain IV gave rise to weak or absent NMR signals, unlike Domain I, which had nearly uniform signal intensities. Weak NMR signals may reflect macromolecular dynamics and are often suggestive of functional involvement.<sup>17,18</sup> Measurements of residue-specific rotational correlation times from cross-correlation experiments for Domain IV were relatively uniform as shown in Figure 3(A), though the C-terminus had markedly reduced values, suggestive of motion on the ps-ns timescale. In addition, the average value of 4.7 ns for the ordered residues of Domain IV (7–87) was indicative of a monomeric form and consistent with the gel filtration chromatography result. Measurements of spin relaxation rates ( $R_{1s}$  and  $R_{2s}$ ) are likewise informative of macromolecular dynamics but, unlike the

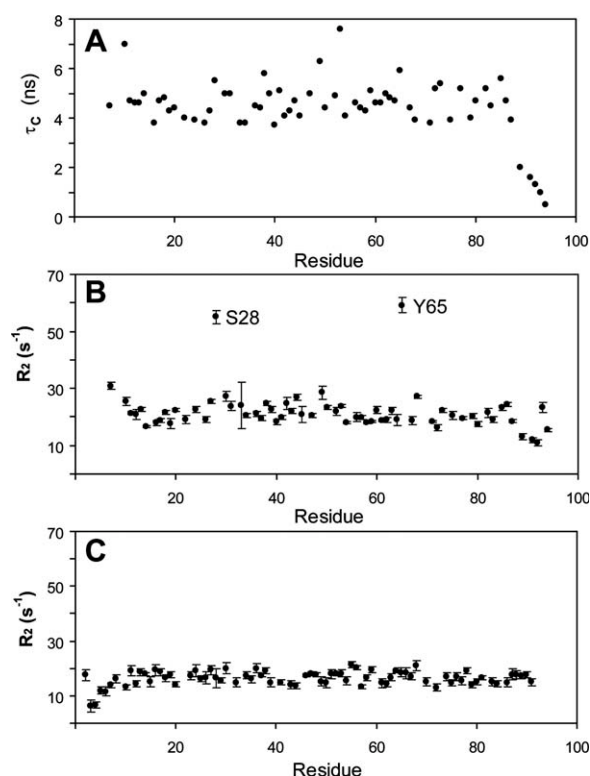




**Figure 2.** Unique structural features of Domains I and IV are conserved and are similar to the C-terminal domain of the TL5 ribosome-associated protein from *Thermus thermophilus*. Panel A: The unique topology of these domains forms a “figure 8”-like fold. The structurally variable region is depicted with {s}. Panel B: The topology of the TL5 protein is very similar, though one b-strand is missing and two additional helices are present. Panel C: An overlay of the Domain I X-ray structure (deep blue) and TL5 C-terminal domain (light orange) shows considerable differences. Panel D: The presence of a Pro residue preceded by a *cis* peptide bond is observed in structures of YbbR Domains I and IV though residues in the immediate vicinity are not conserved. However, the location of hydrophobic amino acid sidechains preceding the *cis*-Pro is conserved. Domain coloring is the same as in Figure 1.

correlation time measurements from cross-correlation experiments,  $R_2$ s can be sensitive to slow,  $\mu$ s-ms timescale motions. Decay rates for transverse magnetization ( $R_2$ ) of Domain IV amide nitrogen signals revealed greater values for residues in three regions of the protein (residues of the N-terminus, S28 and Y65) that also had reduced or absent signals in

some of the backbone assignment experiments [Fig. 3(B)]. This behavior is not observed for Domain I [Fig. 3(C)] and may reflect instability in Domain IV due to the shorter N-terminal relative to Domain I. These measurements are consistent with slow-timescale dynamics within Domain IV; however, the structural basis for these motions is not clear.



**Figure 3.** NMR spin relaxation measurements identify dynamic regions of Domain IV that are not present in Domain I. Panel A: A measurement of the rotational correlation time of each NH moiety of Domain IV showed a relatively homogenous distribution with an average for ordered residues of 4.7 ns, suggesting ps-ns timescale motions of the backbone are uniform excepting the highly mobile C-terminal residues. Panel B: Measurements of the  $R_2$  spin relaxation rate identified three regions of Domain IV with significantly larger values, suggesting the presence slow  $\mu$ s-ms timescale motions. These residues likewise have greatly reduced or absent intensity in three-dimensional heteronuclear backbone assignment experiments when compared to other residues in the protein. Panel C: Similar transverse relaxation measurements of Domain I (shown) and analysis of triple-resonance experiments identified no regions with enhanced relaxation rates.

## Discussion

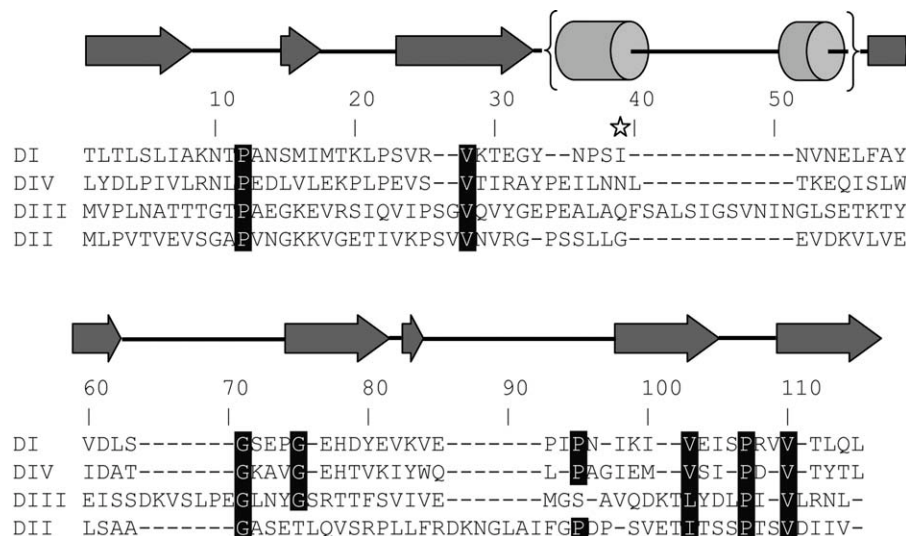
The models of Domains I and IV provide structural templates for homology modeling of other YbbR family proteins that may help to identify unusual conserved structural features, which reflect domain function. A refined amino acid alignment for the four YbbR domains of the *D. hafniense* protein based on the domain structures presented here shows conserved features of the domains as shown in Figure 4. Primarily because of the low sequence identity across all four of the domains (2.5 vs. 4.3% in the structure-based alignment) as judged using a non-structure based alignment, these features are not evident from a simple amino acid sequence alignment and highlight the importance of structural in-

formation in identifying features potentially relevant to function. The presence of a *cis*-Pro residue and the position of the Val residues identified above suggest that these features are conserved to preserve a local surface characteristic, because they alone are unlikely to dictate the shape of the entire domain.

The similarity of the YbbR domains to the TL5 and L25 ribosomal proteins may offer insight into the biochemical function of the YbbR domains. TL5 and L25 are believed to bind to the 5S ribosome through their N-terminal domains, but both have a functionally uncharacterized C-terminal domain with a topology similar to YbbR.<sup>14</sup> In a crystal structure of the *D. radiodurans* ribosome (PDB: 2zjr), for example, the L25 protein is engaged with the 5S ribosome, and the C-terminal domain does not appear to interact with any component of the ribosome; it instead lies on the periphery of the holomolecule.<sup>16</sup> *D. radiodurans* mutants lacking L25 are still viable, though growth is reduced.<sup>19</sup> L25 and TL5, also known as CTC proteins in different systems, are stress response factors,<sup>20,21</sup> though again, the role of the C-terminal domain in this response remains undefined.

The modular often repeated arrangement of YbbR-containing genes likewise is suggestive of function. Similar arrangements of repeated SH2/3 or PDZ protein-protein interaction domains are frequently observed in eukaryotic systems.<sup>22</sup> Multidomain scaffolds often serve to organize larger protein complexes and coordinate specific events, much like the Shank family of proteins in the nervous system of higher organisms.<sup>23</sup> Similar scaffold proteins have been described in bacterial proteomes including FtsZ, a scaffold for cytokinesis machinery,<sup>24</sup> and CheW, a scaffold for signal transduction in flagellar motor regulation.<sup>25</sup> Alternatively, YbbR repeats may not bind proteins but another biomolecule such as a carbohydrate or other cell wall/membrane component, working in parallel to obtain a high avidity interaction. Monomeric lectins, for example, often interact weakly with cognate ligands but commonly oligomerize to enhance affinity in manner that could likewise be achieved by sequential repetition.<sup>26</sup>

As mentioned in the introduction, the third domain of the YbbR protein from *B. subtilis* has been identified as a substrate for a Sfp-PPTase.<sup>6</sup> Walsh and coworkers identified the site of modification as the Ser residue in a D-S-E-L-F motif. The structure and location of this modification may now be studied with a homology model of the *B. subtilis* YbbR domain (not shown) built using the structures of Domains I and IV presented here. Based on this model, the Ser acceptor residue from the *B. subtilis* YbbR domain would reside at the solvent-exposed C-terminus of a helix found between  $\beta$ -strands 3 and 4 of the Domain IV structure presented here. The location of this residue, mapped onto the Domain IV



**Figure 4.** A structure-based sequence alignment of the four YbbR domains in the *D. hafniense* YbbR protein highlights regions of conservation not observed in a naïve amino acid alignment. Structurally conserved features are highlighted. Secondary structure features are shown above the alignment and are based on the domain structures presented herein. The homology modeled site of Sfp PPTase modification is shown with a star.

structure and aligned sequences, is shown in Figures 1 and 4. This location is similar in terms of local secondary structure and position on the protein/solvent boundary to the phosphopantetheine acceptor site on acyl carrier protein.<sup>27</sup> Interestingly, both Domain I and III of the *D. hafniense* protein contain a Ser residues in this vicinity that may be substrates for this type of modification (Fig. 4). Although *in vivo* modification of YbbR domain has not been demonstrated, these data are consistent with a conserved structural site for Sfp-PPTase modification. The potential of some, but not all YbbR domains, to accept this modification suggests these domains may assume a variety of roles *in situ*.

Other proteins displaying structures similar to the YbbR domains have appeared in the Protein Data Bank (PDB) since Domains I and IV were deposited and contain the unique elements observed in the YbbR structures. One protein from *Streptococcus thermophilus* (PDB: 2kxy; Liu and Prestegard, unpublished) is 18% identical at the amino acid level to Domain IV but shares the topology and *cis*-Pro architecture. Another protein from *Bacillus halodurans* is only 15% identical but shares a nearly identical topology (PDB: 2kq1; Wu, Szyperski *et al.*, unpublished). In addition, chemical shifts from the *B. halodurans* protein (Biological Magnetic Resonance Data Bank 16576) provide evidence for a *cis*-Pro residue similar to that described herein, based on the C $\beta$  and C $\gamma$  chemical shifts.<sup>28</sup>

Future genetic and biochemical studies will be required to unambiguously define the role of YbbR domains in a host organism. The structural and sequence conservation in a wide assortment of bacteria combined with the location in the genome, the relationship to a general stress response protein,

and role of some domains as an acceptor of phosphopantetheinylation combine to suggest an important role for these proteins in the function and survival of bacteria that express YbbR domains.

## Materials and Methods

NESG targets DhR29a (Domain IV) and DhR29b (Domain I) were cloned, expressed, and purified based on the standard procedures of NESG to produce a uniformly <sup>15</sup>N/<sup>13</sup>C-enriched protein samples for NMR spectroscopy and selenomethionine-labeled samples for X-ray crystallography.<sup>29,30</sup> DNA encoding residues 32–118 (Domain I; NESG ID DhR29B) or 321–409 (Domain IV; NESG ID DhR29A) from the YbbR family protein Dhaf\_0833 of *D. hafniense* Y51 was cloned into a pET21b vector with a non-cleavable C-terminal hexa-His tag. These pET expression vectors for Domain I (NESG DhR29B-31-118-21.1) and Domain IV (NESG DhR29A-321-409-21.2) are freely available from the PSI Materials Repository (<http://psimr.asu.edu/>). Transformed cells were cultured at 37°C in MJ9 minimal medium<sup>30</sup> containing (<sup>15</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and U-<sup>13</sup>C-glucose as the sole nitrogen and carbon sources. Proteins were purified using an AKTAexpress FPLC apparatus with a two-step protocol consisting of HisTrap HP affinity and HiLoad 26/60 Superdex 75 gel filtration chromatography. The purity of purified proteins (>98%) was verified with sodium dodecyl sulfate polyacrylamide gel electrophoresis and matrix-assisted laser-desorption ionization time of flight mass spectrometry. Extensive cloning, expression, purification, and biophysical characterization data are on-line in the NESG Spine database (<http://spine.nesg.org/target.cgi?id=DhR29>).<sup>31</sup>



### Solution NMR spectroscopy and refinement

The backbone assignment and NOE experiments for Domain IV and all RDC experiments were collected at the University of Georgia on a Varian Inova 600-MHz spectrometer equipped with a cryogenically cooled probe. The backbone assignment and NOE pulse sequences were supplied by Varian as part of the BioPack distribution. The backbone and NOE experiments for Domain I were collected at Pacific Northwest National Laboratory on a Varian Inova 600- and 750-MHz spectrometers with 5-mm room temperature probes. The NMR spectra of Domain IV were collected using protein concentrations of 0.9–1.1 mM in a buffer containing 20 mM MES, 200 mM sodium chloride, 5 mM calcium chloride, 0.02% sodium azide, 10 mM dithiothreitol, 1x protease inhibitors, 10% D<sub>2</sub>O, and 50  $\mu$ M DSS, pH 6.5 at 25 °C. NMR spectra for Domain I were collected using similar protein concentrations in a buffer containing 20 mM ammonium acetate, 200 mM sodium chloride, 10 mM calcium chloride, 0.02% sodium azide, and 5% D<sub>2</sub>O, pH 4.5 at 25°C. Sequence-specific backbone resonance assignments were determined using HNCO, HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH experiments. CCONH, HBHA(CO)NH, and HCCH-TOCSY experiments were used for sidechain assignment. The NOE distance constraints for structure calculations were derived from <sup>15</sup>N-edited NOESY-HSQC and <sup>13</sup>C-edited NOESY-HSQC (for the aliphatic region) taken with mixing times of 100 ms. Stereospecific valine and leucine methyl proton assignments were obtained using the method of Neri *et al.*<sup>32</sup> NMR data were processed using NMRPipe and analyzed using NMRViewJ (Domain IV) or SPARKY (Domain I).<sup>33,34</sup> Backbone resonance assignment completeness for Domains I and IV, respectively, was: 100 and 90% for NH, 91 and 84% for Co, 100 and 94% for Ca, and 100 and 95% for Cb. These values do not include the N-terminal Met or the His6 Tag residues.

For RDC measurements Domain I was partially aligned in a positively charged (50% 3-acrylamidopropyl-trimethylammonium chloride + 50% acrylamide) compressed gel medium.<sup>35</sup> The positively charged gel was cast in a 3.2-mm diameter plastic tube. The polymerized gel was first washed extensively in deionized water followed by washing with protein buffer to equilibrate pH. Finally, the gel was washed with deionized water to remove buffer. The swelled gel (~7-mm diameter) was trimmed to a length of 35 mm and dried in open air for 2 days. The gel pellet was swollen in a 5-mm Shigemi NMR tube using the protein solution. The plunger of the Shigemi tube was fixed at a height of 14 mm from the bottom of the tube.

Domain I was also partially aligned in 4.2% (v/v) C<sub>12</sub>E<sub>5</sub> polyethylene glycol (PEG) bicelle using previously published protocols.<sup>36</sup> Specifically, 16% (v/v)

C<sub>12</sub>E<sub>5</sub> PEG stock was prepared by mixing 50  $\mu$ L of PEG (Sigma Aldrich), 50  $\mu$ L of D<sub>2</sub>O, 200  $\mu$ L of protein buffer (0.02% NaN<sub>3</sub>, 10 mM DTT, 5 mM CaCl<sub>2</sub>, 200 mM NaCl, 1x Protease Inhibitors, 20 mM MES pH 6.5, 10% D<sub>2</sub>O, 50  $\mu$ M DSS), and 16  $\mu$ L of hexanol. The alignment medium was then prepared by mixing 55  $\mu$ L of 16% PEG stock with 145  $\mu$ L of protein and 20  $\mu$ L of D<sub>2</sub>O to reach 4.2%.

Domain IV was partially aligned in a negatively charged (50% 2-acrylamido-2-methyl-1-propanesulfonic acid + 50% acrylamide) compressed gel medium.<sup>35</sup> Preparation of the negatively charged gel was identical to the positively charged gel. Domain IV was also aligned in 12.5 mg/mL Pf1 phage (ASLA Biotech) using previously published protocols.<sup>37</sup> The alignment medium was prepared by mixing 55  $\mu$ L of Pf1 phage PEG stock with 145  $\mu$ L of protein and 20  $\mu$ L of D<sub>2</sub>O to reach 12.5 mg/mL concentration. In all cases, RDCs of <sup>1</sup>H-<sup>15</sup>N amide pairs were measured using a J-modulated experiment.<sup>38</sup> N–CO RDCs were measured according to Liu and Prestegard.<sup>39</sup>

Backbone dihedral angles for Domains I and IV were predicted using TALOS based on the assigned chemical shifts of HA, CA, CB, CO, and N.<sup>40</sup> Residue-specific rotational correlation times were estimated using the dipole-dipole/CSA cross-correlated relaxation of backbone amides by the method of Liu and Prestegard.<sup>41</sup> Antiphase HzNx *R*<sub>2</sub> rates were measured in an isotropic medium and fitted simultaneously with the isotropic <sup>1</sup>J values. The structure calculations were initially done using CYANA.<sup>42</sup> NOEs and RDCs from regions of Domain IV with motional influences, as characterized by the *R*<sub>2</sub> values, were not included in structure refinement. During the optimization stage for the NOE distance and dihedral angle constraints, 50 structures were calculated by CYANA and 20 structures with the lowest target energies were selected for analysis. Starting *D*<sub>a</sub> and *R* values for each partially aligning medium were calculated from principle order parameters determined in PALES.<sup>43</sup> The structural refinement was performed using NOE distance, dihedral angle, and orientation constraints using XPLOR-NIH; the top 10 structures with the lowest NOE violations of the 50 calculated structures were selected for final structure deposition.<sup>44</sup> The results of these refinements are summarized in Table I. The quality of the structural models was verified using the Protein Structure Validation Suite (PSVS) software (<http://psvs.nesg.org>).<sup>11</sup> Structural ensembles were deposited in the PDB as 2l5n (Domain I) and 2l3u (Domain IV). Chemical shift, NOESY peak list, and NOESY fid data have been deposited in the BioMagResDB (Domain I-16570 and Domain IV-16568). RDC data are available at <http://spine.nesg.org/rdc.cgi>. The rmsd determinations were made using CHIMERA<sup>45</sup> or PyMOL (Schrödinger, LLC). Alignments of the NMR and X-ray models of Domain I

included 74 C $\alpha$  atoms. Alignments of TL5 with Domains I and IV included the best 34 pairs of C $\alpha$  atoms.

### Crystallization, data collection, and structure determination

The selenomethionine derivative of Domain I (DhR29B) was crystallized by the hanging drop vapor diffusion method at room temperature. The protein sample ( $\sim 15$  mg/mL) was mixed with an equal volume ( $2 + 2$   $\mu$ L) of a precipitant solution containing 0.1M KBr, 0.1M Bis-tris (pH 7.0), and 30% (w/v) PEG 8000. Crystals grew to an approximate size of 0.4 mm  $\times$  0.8 mm  $\times$  0.8 mm in 2 weeks. A single-wavelength anomalous diffraction dataset was collected at beamline X4A of the Brookhaven National Laboratory from a flash-cooled crystal after soaking the crystal in a cryoprotectant solution containing 20% (v/v) glycerol with the mother liquor. Data were processed, scaled, and merged using the HKL2000 program package.<sup>46</sup> The structure was determined by SHELX.<sup>47</sup> ARP/wARP<sup>48</sup> and COOT<sup>49</sup> were used for model building. Several cycles of simulated annealing, minimization, and B-factor refinement were carried out using the CNS program package.<sup>50</sup> The R-factor, R-free, and geometry were checked during refinements. The R-free was calculated based on 10% of randomly selected data excluded from the refinement. The solvent molecules located in the difference electron-density maps ( $F_o - F_c$ ) were included in the final refinement. The crystallographic statistics for data collection and refinement are summarized in Table II. The resulting model was deposited in the PDB as 3lyw.

### Acknowledgments

The authors thank Yizhou Liu for assistance with the J-modulated RDC experiments and R. Belote, C. Ciccosanti, and K. Hamilton, for assistance in cloning and purifying the proteins. NMR data for Domain I were collected at 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

- Burley SK, Joachimiak A, Montelione GT, Wilson IA (2008) Contributions to the NIH-NIGMS protein structure initiative from the PSI production centers. *Structure* 16:5–11.
- Jaroszewski L, Li ZW, Krishna SS, Bakolitsa C, Wooley J, Deacon AM, Wilson IA, Godzik A (2009) Exploration of uncharted regions of the protein universe. *PLoS Biol* e1000205.
- 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. (2010) The Pfam protein families database. *Nucleic Acids Res* 38:D211–D222.
- Glanzmann P, Gustafson J, Komatsuzawa H, Ohta K, Berger-Bachi B (1999) glmM operon and methicillin-resistant glmM suppressor mutants in *Staphylococcus aureus*. *Antimicrob Agents Chemother* 43:240–245.
- Mengin-Lecreux D, vanHeijenoort J (1996) Characterization of the essential gene glmM encoding phosphoglucosamine mutase in *Escherichia coli*. *J Biol Chem* 271:32–39.
- Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, Kelleher NL, Kolter R, Walsh CT (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci USA* 102:15815–15820.
- Chalut C, Botella L, de Sousa-D'Auria C, Houssin C, Guilhot C (2006) The nonredundant roles of two 4'-phosphopantetheinyl transferases in vital processes of *Mycobacteria*. *Proc Natl Acad Sci USA* 103:8511–8516.
- Yasgar A, Foley TL, Jadhav A, Inglese J, Burkart MD, Simeonov A (2010) A strategy to discover inhibitors of *Bacillus subtilis* surfactin-type phosphopantetheinyl transferase. *Mol Biosyst* 6:365–375.
- Christiansen N, Ahring BK (1996) *Desulfitobacterium hafniense* sp nov, an anaerobic, reductively dechlorinating bacterium. *Intl J System Bacteriol* 46:442–448.
- Nonaka H, Keresztes G, Shinoda Y, Ikenaga Y, Abe M, Naito K, Inatomi K, Furukawa K, Inui M, Yukawa H (2006) Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes* 195. *J Bacteriol* 188:2262–2274.
- Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* 66:778–795.
- Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* 24:2780–2781.
- Petrey D, Fischer M, Honig B (2009) Structural relationships among proteins with different global topologies and their implications for function annotation strategies. *Proc Natl Acad Sci USA* 106:17377–17382.
- Gryaznova OI, Davydova NL, Gongadze GM, Jonsson BH, Garber MB, Liljas A (1996) A ribosomal proteins for *Thermus thermophilus* is homologous to a general shock protein. *Biochimie* 78:915–919.
- Fedorov R, Meshcheryakov V, Gongadze G, Fomenkova N, Nevskaya N, Selmer M, Laurberg M, Kristensen O, Al-Karadaghi S, Liljas A, Garber M, Nikonov S. (2001) Structure of ribosomal protein TL5 complexed with RNA provides new insights into the CTC family of stress proteins. *Acta Cryst D* 57:968–976.
- Harms JM, Wilson DN, Schlutzenzen F, Connell SR, Stachelhaus T, Zaborowska Z, Spahn CMT, Fucini P (2008) Translational regulation via L11: molecular switches on the ribosome turned on and off by thio-strepton and micrococin. *Mol Cell* 30:26–38.
- Mittermaier A, Kay LE (2006) New tools provide new insights in NMR studies of protein dynamics. *Science* 312:224–228.
- Palmer AG, III, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* 339:204–238.
- Korepanov AP, Gongadze GM, Garber MB, Court DL, Bubunenko MG (2007) Importance of the 5 S rRNA-binding ribosomal proteins for cell viability and translation in *Escherichia coli*. *J Mol Biol* 366:1199–1208.

20. Gardan R, Duche O, Leroy-Setrin S, Labadie J, Consortiu ELG (2003) Role of ctc from *Listeria monocytogenes* in osmotolerance. *Appl Environ Microbiol* 69:154–161.
21. Schmalisch M, Langbein I, Stulke J (2002) The general stress protein Ctc of *Bacillus subtilis* is a ribosomal protein. *J Mol Microbiol Biotech* 4:495–501.
22. Pawson T, Scott JD (1997) Signaling through scaffold, anchoring, and adaptor proteins. *Science* 278:2075–2080.
23. Sheng M, Kim E (2000) The Shank family of scaffold proteins. *J Cell Sci* 113:1851–1856.
24. Adams DW, Errington J (2009) Bacterial cell division: assembly, maintenance and disassembly of the Z ring. *Nat Rev Microbiol* 7:642–653.
25. McNally DF, Matsumura P (1991) Bacterial chemotaxis signaling complexes: formation of a CheA/CheW complex enhances autophosphorylation and affinity for CheY. *Proc Natl Acad Sci USA* 88:6269–6273.
26. Varki A (2009) Essentials of glycobiology, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. pp. xxix, 784 p.
27. Kim Y, Prestegard JH (1990) Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins* 8:377–385.
28. Schubert M, Labudde D, Oschkinat H, Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on C-13 chemical shift statistics. *J Biomol NMR* 24:149–154.
29. Acton TB, Gunsalus KC, Xiao R, Ma LC, Aramini J, Baran MC, Chiang YW, Climent T, Cooper B, Denisova 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.
30. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione BT, Nilsson B (1996) High-level production of uniformly <sup>15</sup>N- and <sup>13</sup>C-enriched fusion proteins in *Escherichia coli*. *J Biomol NMR* 7:131–141.
31. Goh CS, Lan N, Echols N, Douglas SM, Milburn D, Bertone P, Xiao R, Ma LC, Zheng D, Wunderlich Z, Acton T, Montelione GT, Gerstein M. (2003) SPINE 2: a system for collaborative structural proteomics within a federated database framework. *Nucleic Acids Res* 31: 2833–2838.
32. Neri D, Szyperski T, Otting G, Senn H, Wuthrich K (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional <sup>13</sup>C labeling. *Biochemistry* 28:7510–7516.
33. 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.
34. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278:313–352.
35. Cierpicki T, Bushweller JH (2004) Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral membrane proteins. *J Am Chem Soc* 126:16259–16266.
36. Ruckert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* 122:7793–7797.
37. Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct Biol* 5: 1065–1074.
38. Tjandra N, Grzesiek S, Bax A (1996) Magnetic field dependence of nitrogen-proton J splittings in N-15-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J Am Chem Soc* 118:6264–6272.
39. Liu YZ, Prestegard JH (2009) Measurement of one and two bond N-C couplings in large proteins by TROSY-based J-modulation experiments. *J Magn Res* 200:109–118.
40. Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289–302.
41. Liu YZ, Prestegard JH (2008) Direct measurement of dipole-dipole/CSA cross-correlated relaxation by a constant-time experiment. *J Magn Res* 193:23–31.
42. Guntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353–378.
43. Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J Am Chem Soc* 122:3791–3792.
44. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Res* 160:65–73.
45. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
46. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
47. Sheldrick GM (2008) A short history of SHELX. *Acta Cryst A* 64:112–122.
48. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–463.
49. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Cryst D* 60:2126–2132.
50. 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. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Cryst D* 54:905–921.
51. Eisenberg D, Luthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Macromol Cryst B* 277:396–404.
52. Sippl MJ (1993) Recognition of errors in 3-dimensional structures of proteins. *Proteins* 17:355–362.
53. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck—a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26: 283–291.
54. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, Richardson DC. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W383.