

Garry W. Buchko,<sup>a,b\*</sup> Stephen N. Hewitt,<sup>a,c</sup> Alberto J. Napuli,<sup>a,c</sup> Wesley C. Van Voorhis<sup>a,c</sup> and Peter J. Myler<sup>a,d,e</sup>

<sup>a</sup>Seattle Structural Genomics Center for Infectious Disease, <http://www.ssgcid.org>, USA, <sup>b</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, USA, <sup>c</sup>Department of Medicine, University of Washington, Seattle, Washington, USA, <sup>d</sup>Seattle Biomedical Research Institute, Seattle, Washington, USA, and <sup>e</sup>Department of Medical Education and Biomedical Informatics and Department of Global Health, University of Washington, Seattle, Washington, USA

Correspondence e-mail:  
 garry.buchko@pnpl.gov

Received 25 January 2011  
 Accepted 19 February 2011

**PDB Reference:** *Bm*-YffB, 2kok.

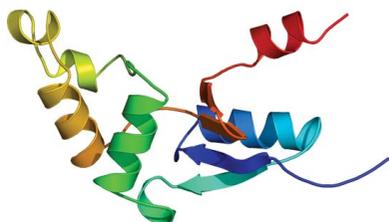
## Solution structure of an arsenate reductase-related protein, YffB, from *Brucella melitensis*, the etiological agent responsible for brucellosis

*Brucella melitensis* is the etiological agent responsible for brucellosis. Present in the *B. melitensis* genome is a 116-residue protein related to arsenate reductases (*Bm*-YffB; BR0369). Arsenate reductases (ArsC) convert arsenate ion ( $\text{H}_2\text{AsO}_4^-$ ), a compound that is toxic to bacteria, to arsenite ion ( $\text{AsO}_2^-$ ), a product that may be efficiently exported out of the cell. Consequently, *Bm*-YffB is a potential drug target because if arsenate reduction is the protein's major biological function then disabling the cell's ability to reduce arsenate would make these cells more sensitive to the deleterious effects of arsenate. Size-exclusion chromatography and NMR spectroscopy indicate that *Bm*-YffB is a monomer in solution. The solution structure of *Bm*-YffB (PDB entry 2kok) shows that the protein consists of two domains: a four-stranded mixed  $\beta$ -sheet flanked by two  $\alpha$ -helices on one side and an  $\alpha$ -helical bundle. The  $\alpha/\beta$  domain is characteristic of the fold of thioredoxin-like proteins and the overall structure is generally similar to those of known arsenate reductases despite the marginal sequence similarity. Chemical shift perturbation studies with  $^{15}\text{N}$ -labeled *Bm*-YffB show that the protein binds reduced glutathione at a site adjacent to a region similar to the  $\text{HX}_3\text{CX}_3\text{R}$  catalytic sequence motif that is important for arsenic detoxification activity in the classical arsenate-reductase family of proteins. The latter observation supports the hypothesis that the ArsC-YffB family of proteins may function as glutathione-dependent thiol reductases. However, comparison of the structure of *Bm*-YffB with the structures of proteins from the classical ArsC family suggest that the mechanism and possibly the function of *Bm*-YffB and other related proteins (ArsC-YffB) may differ from those of the ArsC family of proteins.

### 1. Introduction

*Brucella melitensis* is a facultative intracellular bacterial pathogen that exhibits a host preference for goats and sheep. It is one of the *Brucella* species identified as being responsible for brucellosis, a zoonotic disease that causes abortions and stillbirths in animals and Malta fever in humans. The disease is transmitted to humans primarily *via* contact with infected animals (alive or dead) or the consumption of unpasteurized dairy products (Young, 1995). In undeveloped regions of the Mediterranean, Asia, Africa and Latin America brucellosis infections are a major problem in livestock and human populations, causing severe economic hardship (Corbel, 1997). In humans, the disease is multi-systemic and the symptoms are nonspecific, including fever, chills, malaise, dementia, fatigue, headaches, nausea, vomiting and constipation. Infections can be successfully treated with antimicrobial agents that are able to penetrate well into the host cells, often a combination of doxycycline and streptomycin or doxycycline and rifampin for prolonged periods of time (Young, 1995; Rubinstein *et al.*, 1991). There is much interest in the pathobiology of strains of *Brucella* because it is a potential agent of biological warfare and its pathogenicity is unique in that the organism does not display any obvious 'classical' virulence factors (Moreno & Moriyon, 2002).

Arsenic, a naturally occurring metalloid element that is frequently abundant in the environment (Messens & Silver, 2006), is a human carcinogen (Shi *et al.*, 2004) that is also toxic to most forms of life. The toxic effects arise from the reactivity of arsenic ions with protein



thiols. To counter the toxic effects, a family of arsenic detoxification enzymes has evolved that convert arsenate ion ( $\text{H}_2\text{AsO}_4^-$ ), the highly reactive form of arsenic, to arsenite ion ( $\text{AsO}_2^-$ ), a compound that may be effectively transported outside the cell (Mukhopadhyay *et al.*, 2002). In *B. melitensis*, YffB, a protein with marginal sequence similarity to the classical family of arsenate reductases (ArsC), is found that may play a role in reducing arsenate (DeVecchio *et al.*, 2002). This 13.5 kDa protein (*Bm*-YffB) is a potential drug target because if arsenate reduction is this protein's major biological function, contributing to the organism's virulence, then disabling this protein and the cell's ability to reduce arsenate would make *B. melitensis* more sensitive to the deleterious effects of endogenous arsenate. Towards understanding the biological function of *Bm*-YffB and providing a blueprint for structure-based drug design (Myler *et al.*, 2009) based on this protein, the solution structure of *Bm*-YffB was determined. Its thermostability was measured by CD spectroscopy and its structure was compared with those of a similar protein, *Pseudomonas aeruginosa* YffB (*Pa*-YffB; PDB entry 1rwl; Telyakov *et al.*, 2004), and a protein known to reduce arsenate, *Escherichia coli* ArsC (*Ec*-ArsC; PDB entry 1id9; Martin *et al.*, 2001).

## 2. Materials and methods

### 2.1. Cloning, expression and purification

The *Bm*-YffB gene (BR0369; YP\_4138591.1) was amplified using the genomic DNA of *B. melitensis* biovar Abortus 2308 and the oligonucleotide primers 5'-GGGTCCTGGTTCGATGAGTGTGACCATTACGGCATC-3' (forward) and 5'-CTTGTCGTGCTGTTTATTATAGCTTAAATAAGCTTCATACTGCG-3' (reverse) (Invitrogen, Carlsbad, California, USA). The amplified *Bm*-YffB gene was then gel-purified, treated with T4 DNA polymerase and annealed into the *Nru*I/*Pme*I-digested expression vector AVA0421 at a site that provided a 21-residue tag containing six consecutive histidine residues (MAHHHHHHMGTLEAQTQGPGS-) at the N-terminus of the expressed protein (Choi *et al.*, 2011). The recombinant plasmid was transformed into *E. coli* BL21(DE3)R3-pRARE2 cells (a gift from SGC Toronto, Toronto, Ontario, Canada) using a heat-shock method. Uniformly  $^{15}\text{N}$ - and  $^{15}\text{N}$ -,  $^{13}\text{C}$ -labeled *Bm*-YffB was obtained by growing the transformed cells (310 K) in minimal medium (Miller) containing  $^{15}\text{NH}_4\text{Cl}$  (1 mg ml $^{-1}$ ) and D-[ $^{13}\text{C}_6$ ]-glucose (2.0 mg ml $^{-1}$ ) supplemented with  $\text{FeCl}_3$  (50  $\mu\text{g ml}^{-1}$ ) and the antibiotics chloramphenicol (35  $\mu\text{g ml}^{-1}$ ) and ampicillin (100  $\mu\text{g ml}^{-1}$ ). Once the cells reached an  $\text{OD}_{600}$  of  $\sim 0.8$ , the cells were cooled to 298 K and protein expression was induced with isopropyl  $\beta$ -D-1-thiogalactopyranoside (0.026  $\mu\text{g ml}^{-1}$ ). After approximately 5 h, the cells were harvested by mild centrifugation and frozen at 193 K. The frozen pellet was later thawed and resuspended in  $\sim 35$  ml lysis buffer (0.3 M NaCl, 50 mM sodium phosphate, 10 mM imidazole, pH 8.0) brought to 0.2 mM phenylmethylsulfonyl fluoride (PMSF) prior to three passes through a French press (SLM Instruments, Rochester, New York, USA). Following 60 s sonication (SLM Instruments, Rochester, New York, USA) the cell debris was pelleted by centrifugation at 25 000g for 1 h in a JA-20 rotor (Beckman Instruments, Fullerton, California, USA). The supernatant was then passed through a 0.45  $\mu\text{m}$  syringe filter and applied onto an Ni-NTA affinity column (Qiagen, Valencia, California, USA) containing  $\sim 25$  ml resin. The column was washed stepwise by gravity with 40 ml buffer (0.3 M NaCl, 50 mM sodium phosphate pH 8.0) containing increasing concentrations of imidazole (5, 10, 20, 50 and 250 mM). *Bm*-YffB eluted primarily in the 250 mM imidazole wash. Following exchange into 3C cleavage buffer by overnight dialysis in 4 l cleavage buffer (150 mM

NaCl, 20 mM Tris-HCl, pH 8.4) the protein was concentrated to  $\sim 2$  ml (Amicon Centriprep-10) and the N-terminal polyhistidine tag was removed by overnight incubation with 3C protease (1  $\mu\text{g}$  per 50  $\mu\text{g}$  target protein) at 279 K (Bryan *et al.*, 2011). Using a flow rate of 1.0 ml min $^{-1}$ , the reaction solution was then loaded onto a Superdex 75 HiLoad 16/60 column (GE Healthcare, Piscataway, New Jersey, USA) to simultaneously purify the protein and exchange it into NMR buffer (100 mM NaCl, 20 mM Tris-HCl, 1.0 mM dithiothreitol, pH 7.1). The band containing *Bm*-YffB (retention time 78 min) was collected and the volume was reduced (Amicon Centriprep-10) to generate NMR samples in the 1–2 mM range (Lowry analysis). SDS-PAGE analysis of the final NMR samples showed the protein to be greater than  $\sim 95\%$  pure.

### 2.2. Circular dichroism spectroscopy

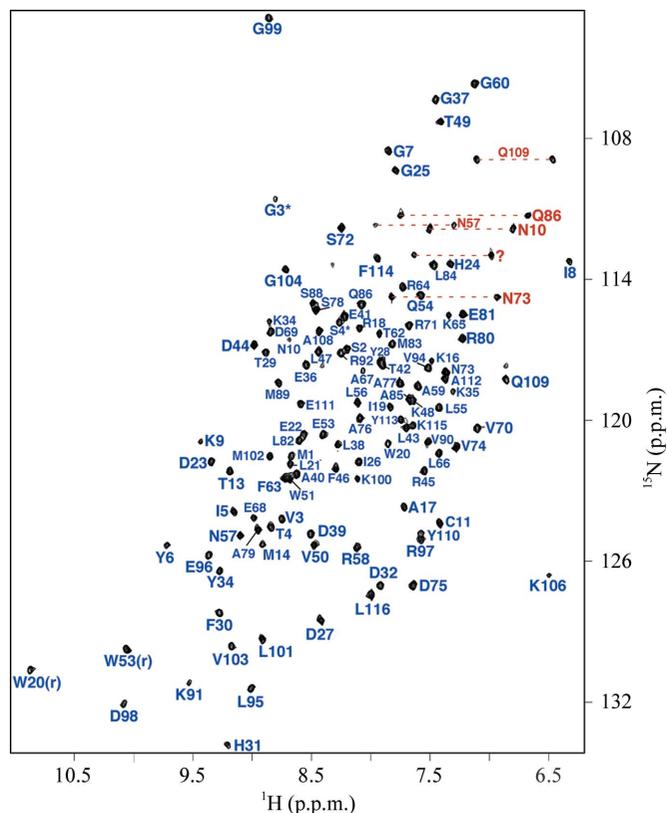
An Aviv Model 410 spectropolarimeter (Lakewood, New Jersey, USA) calibrated with an aqueous solution of ammonium D-(+)-camphorsulfonate was used to collect circular dichroism data from a 0.05 mM *Bm*-YffB sample in NMR buffer in a quartz cell of 0.1 cm path length. A thermal denaturation curve was obtained by recording the ellipticity at 216 nm in 2.0 K intervals from 283 to 353 K. A quantitative estimation of the melting temperature,  $T_m$ , was obtained by taking a first derivative of the thermal denaturation curve using the Aviv software (Greenfield, 2006). Steady-state wavelength spectra for *Bm*-YffB were recorded in 0.5 nm increments between 200 and 260 nm at 298 and 353 K. Each reported steady-state wavelength spectrum was the result of averaging two consecutive scans with a bandwidth of 1.0 nm and a time constant of 1.0 s. These spectra were processed by subtracting a blank spectrum from the protein spectrum and then automatically line-smoothing the data using the Aviv software.

### 2.3. Nuclear magnetic resonance spectroscopy

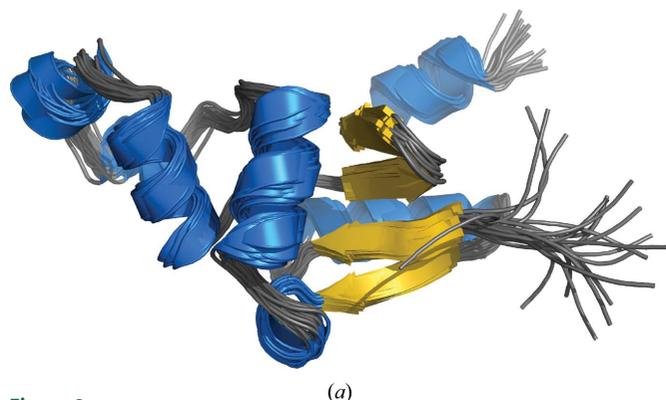
Varian 800-, 750- and 600-Inova spectrometers equipped with  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  triple-resonance probes and pulse-field gradients were used to collect the NMR data required for resonance assignments and structure determination. The NMR data, which were collected from 1–2 mM samples at 293 K, were processed with *Felix*2007 (Felix NMR Inc., San Diego, California, USA) and analyzed with *Sparky* (v.3.115; Goddard & Kneller, 2008). Assignments of the  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts for the backbone and side-chain resonances were made from standard 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC,  $^1\text{H}$ - $^{13}\text{C}$  HSQC, HBCBC-GCDHD and HBCBCGDCHE experiments and 3D HNCACB, CBCA(CO)NH, HNCO, HCC-TOCSY-NNH and CC-TOCSY-NNH experiments using Varian *Protein Pack* pulse programs. Chemical shifts were referenced to DSS (DSS = 0 p.p.m.) using indirect methods (Wishart *et al.*, 1995). Distance restraints were obtained from a suite of 3D  $^{13}\text{C}$ - and  $^{15}\text{N}$ -edited NOESY-HSQC experiments using a mixing time of 80 ms. Deuterium-exchange studies were performed by lyophilizing a  $^{15}\text{N}$ -labeled NMR sample, re-dissolving it in 99.8%  $\text{D}_2\text{O}$  and immediately collecting  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra 10, 20 and 60 min after exchange. An overall rotational correlation time,  $t_c$ , was estimated from backbone-amide  $^{15}\text{N}$   $T_1/T_{1\rho}$  ratios (Farrow *et al.*, 1994; Buchko *et al.*, 2008). A chemical shift perturbation experiment was performed by adding aliquots of reduced glutathione in NMR buffer (50 mM) to a 0.5 mM sample of  $^{15}\text{N}$ -labeled *Bm*-YffB. Following gentle agitation,  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were collected at glutathione:*Bm*-YffB molar ratios of 0.3:1, 0.6:1, 1:1 and 2:1.

## 2.4. Structure calculations

The chemical shifts for *Bm*-YffB were assigned using conventional methods (Cavanagh *et al.*, 1996) and were deposited in the Biological Magnetic Resonance Data Bank (BMRB) under accession No. 16517. Using these  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift assignments and the peak-picked data from  $^{13}\text{C}$ - and  $^{15}\text{N}$ -edited NOESY-HSQC experiments as initial inputs, structure calculations were performed iteratively using *CYANA* (v.2.1; Güntert, 2004). 184 dihedral angle restraints for both  $\varphi$  and  $\psi$  were introduced on the basis of the elements of secondary structure identified in the early structural ensembles and *TALOS*



**Figure 1**  
Assigned  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of double-labeled *Bm*-YffB collected at 293 K in NMR buffer (100 mM NaCl, 20 mM Tris-HCl, 1.0 mM DTT, pH 7.1) at a  $^1\text{H}$  resonance frequency of 750 MHz. Side-chain  $-\text{NH}_2$  resonances are indicated by dashed horizontal lines (red) and the exchangeable ring resonances for Trp20 and Trp53 are identified with an 'r'.



**Figure 2**  
(a) Superposition of the cartoon representations of the ensemble of structures calculated for *Bm*-YffB (PDB entry 2kok), with  $\alpha$ -helices colored blue and  $\beta$ -strands colored gold. (b) Cartoon representation of the structure most similar to the average structure of the ensemble, with the four  $\beta$ -strands and seven  $\alpha$ -helices labeled.

**Table 1**

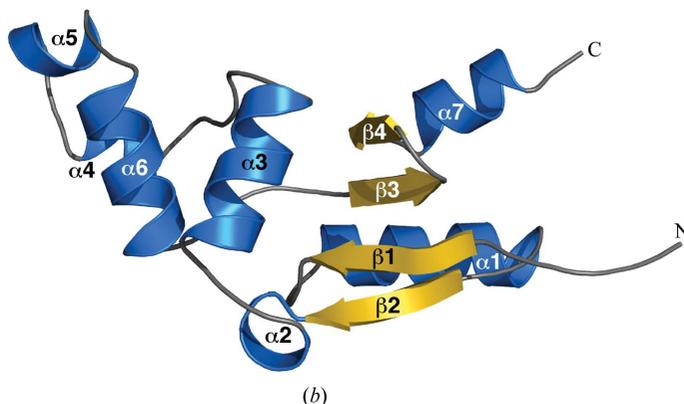
Summary of the structural statistics for *Bm*-YffB.

All statistics are for the 20-structure ensemble deposited in the PDB (2kok).

Restraints for structure calculations	
Total NOEs	1490
Intraresidue NOEs	436
Sequential ( $i, i + 1$ ) NOEs	449
Medium-range ( $i, i + j; 1 < j \leq 4$ ) NOEs	282
Long-range ( $i, i + j; j > 4$ ) NOEs	323
$\varphi$ -angle restraints	94
$\psi$ -angle restraints	94
Hydrogen-bond restraints	84
Structure calculations	
No. of structures calculated	100
No. of structures used in ensemble	20
Structures with restraint violations	
Distance restraint violations $> 0.05 \text{ \AA}$	0
Dihedral restraint violations $> 2^\circ$	0
R.m.s.d. from mean $^\dagger$ ( $\text{\AA}$ )	
Backbone N-C $^\alpha$ -C=O atoms	$0.69 \pm 0.16$
Heavy atoms	$1.19 \pm 0.11$
Ramachandran plot summary for selected residues (from PROCHECK) $^\ddagger$	
Most favored regions (%)	89.8
Additionally allowed regions (%)	10.2
Generously favored regions (%)	0.0
Disallowed regions (%)	0.0
Global quality scores for ordered residues $^\S$	
PROCHECK (all)	-2.84 (-0.48)
PROCHECK ( $\varphi, \psi$ )	-0.28 (-0.15)
MolProbity clash score	-0.86 (13.91)

$^\dagger$  Calculated for the ordered residues Val3-Ile8, Cys11-Ala59, Thr61-Asp98 and Lys100-Lys115.  $^\ddagger$  Calculated for the central ordered core, Ser2-Ile8 and Cys11-Lys115, using *PSVS*.  $^\S$  Z scores; values in parentheses are raw values.

calculations (Cornilescu *et al.*, 1999). Near the end of the iterative process, 84 hydrogen-bond restraints (1.8–2.0 and 2.7–3.0  $\text{\AA}$  for the NH–O and N–O distances, respectively) were introduced into the structure calculations on the basis of proximity in early structure calculations and the observation of slowly exchanging amides in the deuterium-exchange experiment. On the basis of the chemical shift difference between the proline  $^{13}\text{C}^\beta$  and  $^{13}\text{C}^\gamma$  atoms (Schubert *et al.*, 2002), Pro93 was placed in the *cis* conformation. From the final set of 100 calculated structures, the 20 with the lowest target function were selected and this ensemble was deposited in the Protein Data Bank (PDB) under PDB code 2kok. Structural quality was assessed using the *Protein Structure Validation Suite* (*PSVS*; v.1.3; Bhattacharya *et al.*, 2007). Note that the deposited structures were not refined with explicit water (Linge *et al.*, 2003) because these calculations continuously introduced unfavorable steric clashes into the structures despite numerous attempts to adjust the parameters. A summary of the structure statistics is provided in Table 1.



**Table 2**

Chemical shift difference between the proline  $^{13}\text{C}^\beta$  and  $^{13}\text{C}^\gamma$  atoms in *Bm*-YffB.

Residue	$\Delta\beta\gamma$ (p.p.m.)
Pro2*	5.0
Pro51	5.0
Pro67	4.6
Pro87	4.2
Pro93	8.7
Pro107	4.2
Average for <i>trans</i> -Pro	$4.51 \pm 1.17^\dagger$
Average for <i>cis</i> -Pro	$9.64 \pm 1.27^\dagger$

$^\dagger$  Average values obtained from Schubert *et al.* (2002).

The amino-acid sequence of *Bm*-YffB deposited in the PDB and BMRB is numbered sequentially, Gly1–Leu120, starting with the four non-native residues (GPGS-) that remained after 3C protease treatment. However, here the first four non-native residues are numbered sequentially with asterisks (Gly1\*–Ser4\*) and the first native residue, Met5 in the PDB and BMRB depositions, is labeled Met1.

### 3. Results and discussion

#### 3.1. Solution structure of *Bm*-ArsC

The elution time of *Bm*-YffB on a Superdex 75 HiLoad 16/60 column was within a range consistent with a monomeric ~14 kDa protein (data not shown). Such a conclusion was corroborated by the experimentally estimated rotational correlation time determined for *Bm*-YffB,  $9.1 \pm 0.2$  ns (293 K), a value that is more consistent with a monomeric ~14 kDa protein than an ~28 kDa dimer (Bhattacharjya *et al.*, 2004). The line widths and chemical shift dispersion of the  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectrum for *Bm*-YffB, shown in Fig. 1, were also characteristic of a folded monomeric protein with a molecular weight in the 15 kDa range. As illustrated in Fig. 1, 121 of the expected 123 amide resonances for *Bm*-YffB [120 – (6 prolines + Gly1\*)] were assigned in the  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectrum. Amide cross-peaks for Asp16 and Phe109 were not unambiguously identified. On the basis of these amide assignments and extensive assignment of the  $^{13}\text{C}^\alpha$  and side-chain proton and carbon chemical shifts (BMRB ID 16517), an ensemble of structures was calculated (Fig. 2*a*) that satisfied all of the available experimental NMR data (NOEs, chemical shifts, deuterium-exchange experiments and *TALOS* calculations).

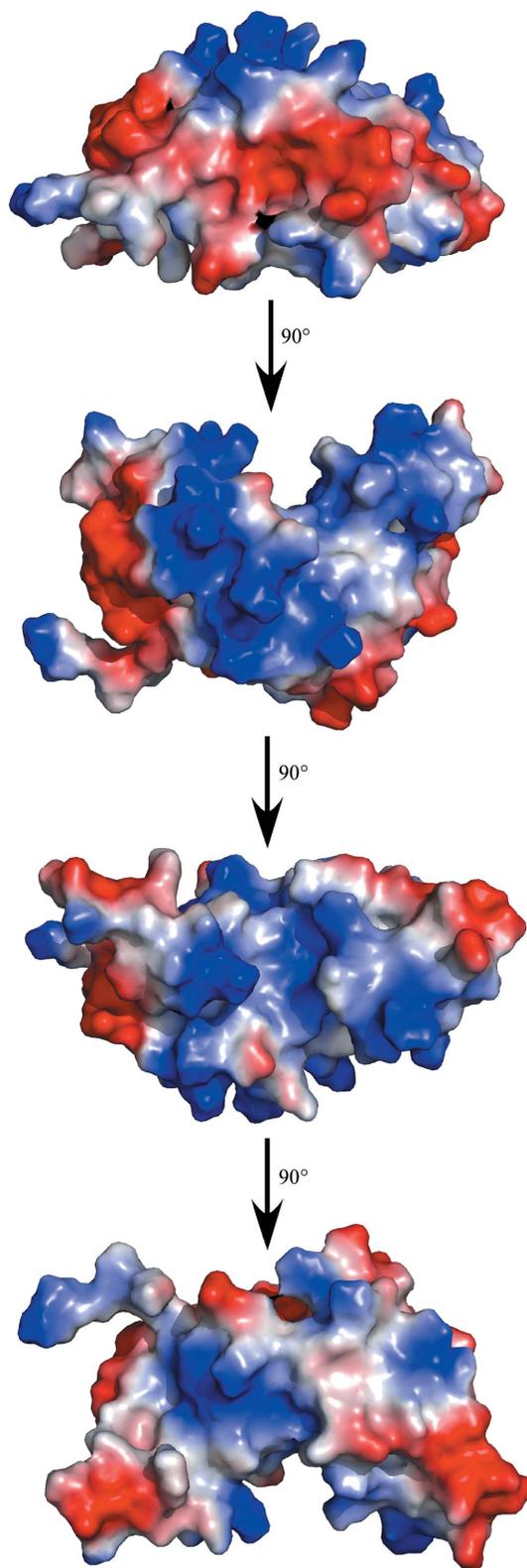
As summarized in Table 1, a total of 1490 interproton distance restraints, 84 hydrogen-bond restraints and 188 dihedral angle restraints were used in the final structure calculations. Each member of the final ensemble of 20 calculated structures agreed well with the experimental data, with no upper limit violation of greater than 0.05 Å and no torsion-angle violation of greater than 2°. Analysis of the ensemble with the *PSVS* validation-software package (Bhattacharjya *et al.*, 2007) also showed that the quality of the structures in the final ensemble was good. The Ramachandran statistics for all residues in the ensemble were overwhelmingly in acceptable space



**Figure 3**

Stereoview showing a cartoon representation of the structure most similar to the average structure in the ensemble, with the protein rainbow-colored (ROYGBIV) from the N-terminus to the C-terminus.

[90% of the ( $\phi$ ,  $\psi$ ) pairs for *Bm*-YffB were found in the most favored regions and 10% were within additionally allowed regions] and all the structure-quality Z scores were acceptable ( $> -5$ ).



**Figure 4**

Maps generated using *PyMOL* (DeLano, 2002) of the electrostatic potential at the solvent-accessible surface of *Bm*-YffB. The long axis of the protein is illustrated and is sequentially rotated 90° about the horizontal axis four times.

The final set of 20 calculated structures in the ensemble converge well, as shown mathematically by the statistics in Table 1 and visually by the superposition in Fig. 2(a). The r.m.s.d.s of the structured core regions (Val3–Ile8, Cys11–Ala59, Thr61–Asp98 and Lys100–Lys115) in the ensemble from the mean structure are 0.7 Å for the backbone atoms (N–C $\alpha$ –C=O) and 1.2 Å for all heavy atoms. Fig. 2(b) illustrates the single structure in the ensemble that is nearest to the mean structure and Fig. 3 shows a stereoview of this single structure. The protein consists of two domains: a four-stranded mixed  $\beta$ -sheet flanked by two  $\alpha$ -helices on one face and an  $\alpha$ -helical bundle. The  $\alpha/\beta$  domain is composed of a  $\beta_4$ – $\beta_3$ – $\beta_1$ – $\beta_2$   $\beta$ -sheet, with  $\beta_4$ – $\beta_3$ – $\beta_1$  aligned antiparallel and  $\beta_1$ – $\beta_2$  aligned parallel.  $\alpha$ -Helix 1 (Asp12–His24) is tucked in behind  $\beta_1$ – $\beta_2$ , and  $\alpha_7$  (Pro107–Phe114) is tucked in behind  $\beta_4$ – $\beta_3$ . The  $\alpha/\beta$  domain is characteristic of the fold of thioredoxin-like proteins and usually contains a *cis*-proline on the N-terminal side of  $\beta_3$  that plays a role in the integrity of the active site (Martin, 1995). *Bm*-YffB also contains a proline, Pro93, at this position and analysis of the  $^{13}\text{C}^\beta$  and  $^{13}\text{C}^\gamma$  chemical shifts for this residue indicates that it is also in the *cis* conformation (Table 2; Schubert *et al.*, 2002). The other domain in *Bm*-YffB is an  $\alpha$ -helical bundle dominated by  $\alpha_3$  (Ala40–Thr49) and  $\alpha_6$  (Ala76–Ala85), with three short helices [ $\alpha_2$  (Tyr33–Glu36),  $\alpha_4$  (Thr61–Lys65) and  $\alpha_5$  (Glu68–Ser72)] around them. Hydrophobic interactions between the side chains of residues on the interface of these two domains, including Phe46 ( $\alpha_3$ ), Thr49 ( $\alpha_3$ ), Tyr6 ( $\beta_1$ ) and Leu101 ( $\beta_4$ ), assist in holding the domains together. As shown in Fig. 4, the protein contains a polarized distribution of charges on its surface, with a dominance of positive surfaces. Such a distribution of charges has been observed in the ArsC protein family, as well as for *Pa*-YffB, and would favor the binding of anions such as arsenate (Teplakov *et al.*, 2004).

### 3.2. Circular dichroism profile and thermal stability of *Bm*-ArsC

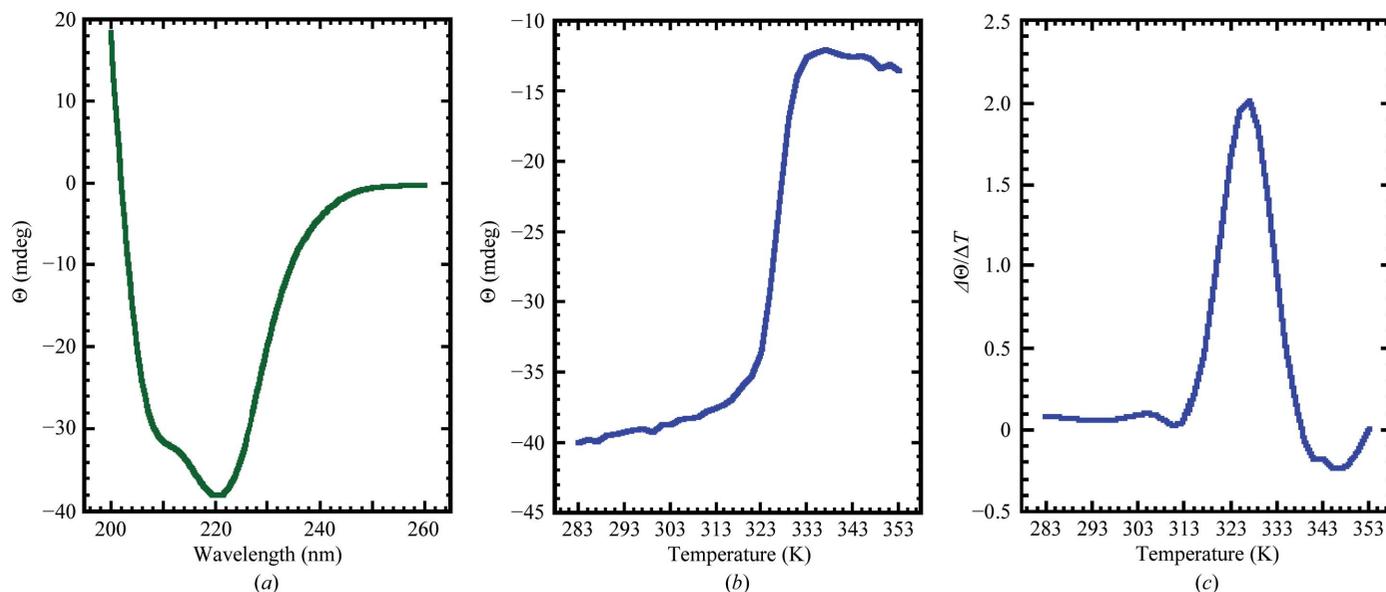
Circular dichroism (CD) spectroscopy is very sensitive to changes in a protein's backbone and, consequently, is a powerful tool to rapidly probe the conformation of proteins in solution and to assess the effect of variables such as pH, salt content and temperature on

the structure of a protein (Woody, 1974). Fig. 5(a) shows the steady-state CD spectrum of *Bm*-YffB collected at 298 K. The dominant feature of the spectrum is characteristic of  $\alpha$ -helical secondary structure: a double minimum at approximately 220 and 208 nm and an extrapolated maximum around 195 nm (Holzwarth & Doty, 1965; Greenfield, 2006). Such an observation is expected given the amount of helical structure (46%) observed in the solution structure of the protein (Figs. 2 and 3). Note that the double minimum is skewed and is more intense around 220 nm, which is likely to be due to the contributions of other elements of secondary structure to the CD steady-state spectrum.

By monitoring the increase in the ellipticity at a specific wavelength with increasing temperature, the thermal stability of a protein may be measured and a melting temperature ( $T_m$ ) estimated for the transition between a structured and an unstructured state (Karantzeni *et al.*, 2003). As shown in Fig. 5(b), a gradual increase in ellipticity at 220 nm is observed for *Bm*-YffB up to  $\sim$ 323 K, followed by a more rapid increase in ellipticity that tails off and plateaus at  $\sim$ 333 K. Visual inspection of the sample after heating to 353 K showed evidence of precipitation, indicating that the unfolding was irreversible and that the CD data may not be analyzed thermodynamically (Karantzeni *et al.*, 2003). However, a quantitative estimation of the  $T_m$  for this transition may still be obtained by assuming a two-state model and taking a first derivative of the curve shown in Fig. 5(b) (Greenfield, 2006). The maximum of this first derivative, shown in Fig. 5(c), is 326.6 K.

### 3.3. Comparison with related structures

The PDB was searched for structures similar to *Bm*-YffB using the DALI search engine (Holm & Rosenström, 2010). The search indicated that the structure of *Bm*-YffB was most similar ( $Z$  score = 13.4) to that of the conserved hypothetical protein YffB from *P. aeruginosa* (PDB entry 1rw1; Teplakov *et al.*, 2004). This similarity is evident in Fig. 6, which shows a superposition of the two structures with the program *SuperPose* (Maiti *et al.*, 2004). It is evident that both proteins have the same number of  $\beta$ -strands and  $\alpha$ -helices organized into a similar two-domain structure. Indeed, using *Bm*-YffB residues Ser2–



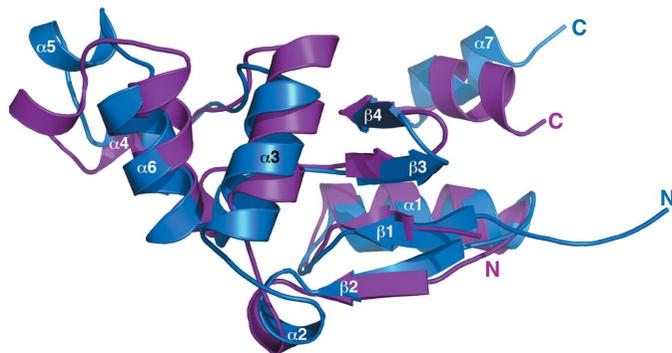
**Figure 5** (a) Circular dichroism steady-state wavelength spectrum for *Bm*-YffB (0.05 mM) in NMR buffer collected at 298 K. (b) The CD thermal melt for *Bm*-YffB obtained by measuring the ellipticity at 226 nm in 2.0 K intervals between 283 and 353 K. (c) The first derivative of the thermal melt curve shows that the protein has a melting temperature of 326.6 K.

Lys115 the backbone r.m.s.d. is 3.78 Å between *Bm*-YffB and the equivalent region in *Pa*-YffB. Such similarity between the structures of the two proteins is reasonable given that the amino-acid sequences of the two proteins are 56% identical and 70% similar.

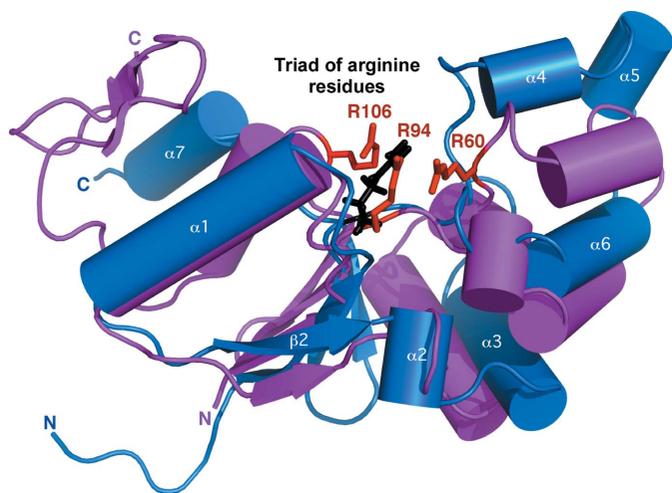
After *Pa*-YffB, the DALI search identified a number of proteins with *Z* scores between 11.3 and 10.0 that were annotated as arsenate reductases (ArsC) or Spx regulatory proteins. While the structure of *Bm*-YffB produced high DALI *Z* scores with these two types of proteins, the sequence identity between *Bm*-YffB and these proteins was only 14–27%. The Spx protein is a global transcription regulatory protein that does not bind DNA as part of its regulatory role but instead binds to the  $\alpha$ -subunit of RNA polymerase to control gene expression in response to disulfide-stress conditions (Nakano *et al.*, 2005). In Spx it is hypothesized that disulfide-bond formation in a CXXC motif triggers events that result in an increase in the cellular levels of thioredoxin and thioredoxin reductase (Nakano *et al.*, 2003). Because *Bm*-YffB is a monomer and contains only a single cysteine residue, it is unlikely that it functions as a regulatory protein using a mechanism similar to that which may be employed by Spx. On the other hand, *Bm*-YffB contains a sequence that is somewhat similar to the HX<sub>3</sub>CX<sub>3</sub>R catalytic sequence motif that is important for arsenic

detoxification activity in the arsenate-reductase family of proteins (ArsC; Martin *et al.*, 2001); hence, it is more likely that *Bm*-YffB has a function related to arsenate reduction.

Fig. 7 shows a superposition of a representative structure of *Bm*-YffB (PDB entry 2kok) with the crystal structure of *E. coli* ArsC (*Ec*-ArsC) in the native form (PDB entry 1id9; Martin *et al.*, 2001) using the program *SuperPose* (Maiti *et al.*, 2004). Apart from an extra two-stranded  $\beta$ -sheet at the C-terminus of *Ec*-ArsC, the two structures are generally similar to each other: a four-stranded mixed  $\beta$ -sheet is flanked by a number of  $\alpha$ -helices that generate a similar overall fold. One major difference between the two structures, highlighted in Fig. 7, is the presence of a triad of arginine residues, Arg60, Arg94 and Arg106 (red), in *Ec*-ArsC that is absent in *Bm*-YffB. These three arginines have been reported to be essential for enzymatic function, forming a thiasahydroxyl adduct when bound to arsenate (Martin *et al.*, 2001). In the *Bm*-YffB structure only one arginine, Arg96 (black), is found in this area. Such a major structural difference suggests a difference in substrate specificity between *Bm*-YffB and *Ec*-ArsC and it has been postulated that the ArsC-YffB family of proteins might be able to bind glutathione in the absence of arsenate (Teplakov *et al.*, 2004).



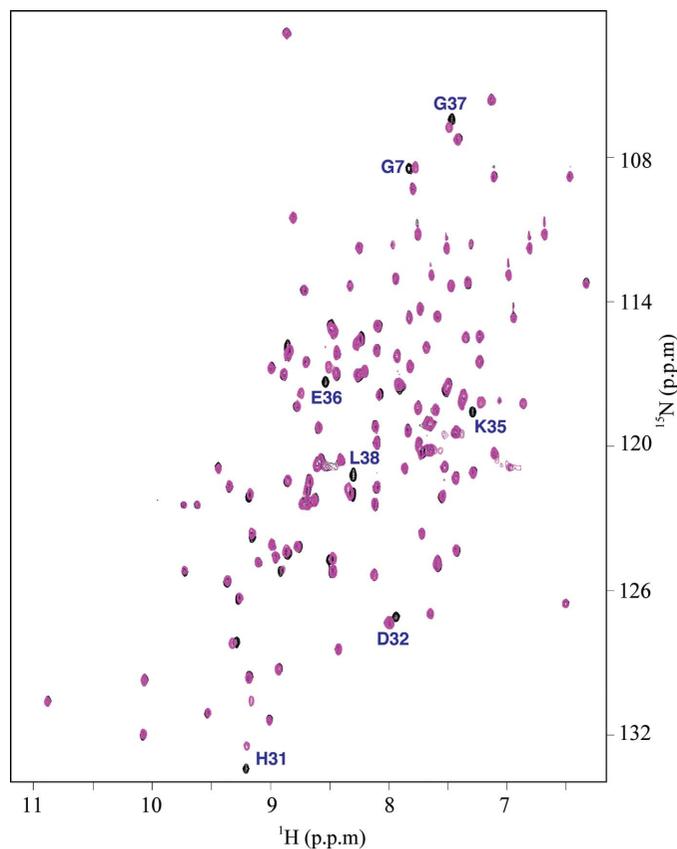
**Figure 6**  
Superposition of the structure closest to the mean for *Bm*-YffB (PDB entry 2kok, blue) on the crystal structure determined for *Pa*-YffB (PDB entry 1rw1, magenta; Teplakov *et al.*, 2004) using the program *SuperPose* (Maiti *et al.*, 2004).



**Figure 7**  
Superposition of the structure closest to the mean for *Bm*-YffB (PDB entry 2kok, blue) on the crystal structure determined for *E. coli* ArsC (PDB entry 1id9, magenta; Martin *et al.*, 2001) using the program *SuperPose* (Maiti *et al.*, 2004). The side chains of the three arginine residues in *Ec*-ArsC (Arg60, Arg94 and Arg107) that form a thiasahydroxyl adduct that is essential for enzymatic function are labeled and colored red. The only residue in *Bm*-YffB that is equivalent to those in the arginine triad of *Ec*-ArsC is Arg96; this side chain is colored black.

### 3.4. Chemical shift perturbation studies with reduced glutathione

In support of the hypothesis that the ArsC-YffB family of proteins may bind glutathione in the absence of arsenate ions, it has been



**Figure 8**  
Overlay of the <sup>1</sup>H–<sup>15</sup>N HSQC spectrum of <sup>15</sup>N-labeled *Bm*-YffB (black) on the spectrum collected in the presence of a 2:1 molar ratio of glutathione:*Bm*-YffB (magenta). Residues that were significantly perturbed are labeled. The spectrum at an ~1:1 molar ratio of glutathione:*Bm*-YffB was similar to that shown here at a 2:1 molar ratio. Data were collected at a proton resonance frequency of 600 MHz at 293 K.

shown by mass spectrometry that *P. aeruginosa* YffB can bind glutathione (Teplyakov *et al.*, 2004). One of the advantages of structure determination by NMR-based methods is that following the chemical shift assignment of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum, not only is it possible to quickly detect ligand binding to the protein by chemical shift perturbation experiments, but, it is also possible to map the location of the binding interaction on the three-dimensional structure of the protein (Buchko *et al.*, 1999; Zuiderweg, 2002). This is possible because protein–ligand interactions often manifest as changes in the chemical environment of the nuclei at the interface of ligand binding which are accompanied by perturbations in the measurable chemical shifts of the backbone  $^1\text{H}^{\text{N}}$  and  $^{15}\text{N}$  resonances. Upon identifying the resonances that experience a binding-dependent chemical shift or intensity perturbation, it is possible to map the location of the binding site onto the structure of the protein. Fig. 8 is the result of such an experiment following the titration of reduced glutathione into a  $^{15}\text{N}$ -labeled sample of *Bm*-YffB. At a 2:1 molar ratio of reduced glutathione:*Bm*-YffB, a subset of seven resonances are observed to shift. In Fig. 9 these perturbed resonances are mapped onto the structure of *Bm*-YffB and it is observed that they all cluster into one area (red) adjacent to the potential  $\text{GX}_3\text{CX}_3\text{K}$  catalytic sequence motif (yellow). The latter observation adds support to the hypothesis that *Bm*-YffB and other proteins in the YffB family may function as glutathione-dependent thiol reductases.

#### 4. Conclusions

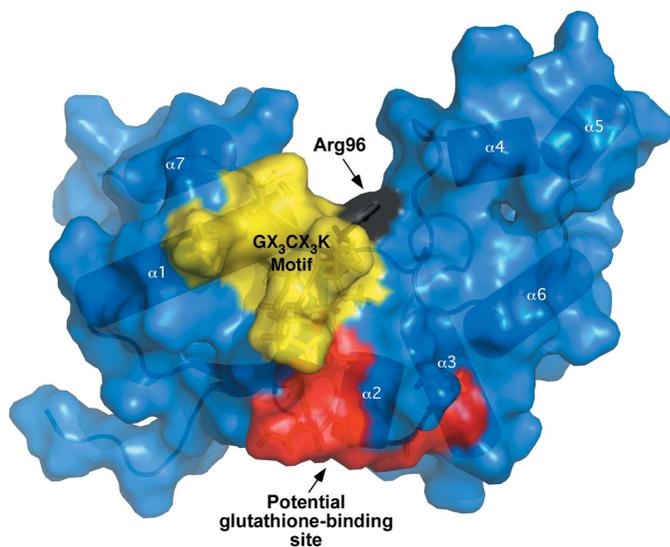
The solution structure determined for *Bm*-YffB is similar to the crystal structure reported for the YffB protein from *P. aeruginosa*: a four-stranded mixed  $\beta$ -sheet flanked by two  $\alpha$ -helices on one side and an  $\alpha$ -helical bundle. While the protein sequences of *Bm*-YffB and *Pa*-YffB are marginally similar to the amino-acid sequences of classical arsenate reductases (ArsC), comparison of the structures of YffB proteins and ArsC proteins suggest that the substrate specificities and mechanisms of these two families of proteins may differ. Chemical shift perturbation studies with reduced glutathione cor-

roborate the hypothesis that YffB proteins may function as glutathione-dependent thiol reductases. Further biochemical and biophysical studies will be necessary in order to identify the substrate and work out the details of the mechanism before *Bm*-YffB may be exploited for structure-based drug design. The solution structure and biophysical data presented here for *Bm*-YffB will facilitate these efforts.

This research was funded by the National Institute of Allergy and Infectious Diseases, National Institute of Health, Department of Health and Human Services under Federal Contract No. HHSN272200700057C. The SSGCID internal ID for *Bm*-YffB is BrabA.00007.a. The majority of the research presented here was conducted at the W. R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the US Department of Energy's Office of Biological and Environmental Research (BER) program located at Pacific Northwest National Laboratory (PNNL). Battelle operates PNNL for the US Department of Energy.

#### References

- Bhattacharjya, S., Xu, P., Gingras, R., Shaykhtudinov, R., Wu, C., Whiteway, M. & Ni, F. (2004). *J. Mol. Biol.* **344**, 1071–1087.
- Bhattacharya, A., Tejero, R. & Montelione, G. T. (2007). *Proteins*, **66**, 778–795.
- Bryan, C. M., Bhandari, J., Napuli, A. J., Leibly, D. J., Choi, R., Kelley, A., Van Voorhis, W. C., Edwards, T. E. & Stewart, L. J. (2011). *Acta Cryst.* **F67**, 1010–1014.
- Buchko, G. W., Daughdrill, G. W., de Lorimier, R., Rao, B. K., Isern, N. G., Lingbeck, J. M., Taylor, J. S., Wold, M. S., Gochin, M., Spicer, L. D., Lowry, D. F. & Kennedy, M. A. (1999). *Biochemistry*, **38**, 15116–15128.
- Buchko, G. W., Tarasevich, B. J., Bekhazi, J., Snead, M. L. & Shaw, W. J. (2008). *Biochemistry*, **47**, 13215–13222.
- Cavanagh, J., Palmer, A. G., Fairbrother, W. & Skelton, N. (1996). *NMR Spectroscopy: Principles and Practice*. San Diego: Academic Press.
- Choi, R., Kelley, A., Leibly, D., Nakazawa Hewitt, S., Napuli, A. & Van Voorhis, W. (2011). *Acta Cryst.* **F67**, 998–1005.
- Corbel, M. J. (1997). *Emerg. Infect. Dis.* **3**, 213–221.
- Cornilescu, G., Delaglio, F. & Bax, A. (1999). *J. Biomol. NMR*, **13**, 289–302.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- DelVecchio, V. G. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 443–448.
- Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994). *Biochemistry*, **33**, 5984–6003.
- Goddard, T. D. & Kneller, D. G. (2008). *Sparky 3*. University of California, San Francisco, USA. <http://www.cgl.ucsf.edu/home/sparky/>.
- Greenfield, N. J. (2006). *Nature Protoc.* **6**, 2527–2535.
- Güntert, P. (2004). *Methods Mol. Biol.* **278**, 353–378.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Holzwarth, G. M. & Doty, P. (1965). *J. Am. Chem. Soc.* **87**, 218–228.
- Karantzeni, I., Ruiz, C., Liu, C.-C. & Licata, V. J. (2003). *Biochem. J.* **374**, 785–792.
- Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. & Nilges, M. (2003). *Proteins*, **50**, 496–506.
- Maiti, R., van Domselaar, G. H., Zhang, H. & Wishart, D. S. (2004). *Nucleic Acids Res.* **23**, W590–W594.
- Martin, J. L. (1995). *Structure*, **3**, 245–250.
- Martin, P., DeMel, S., Shi, J., Gladysheva, T., Gatti, D. L., Rosen, B. P. & Edwards, B. F. (2001). *Structure*, **9**, 1071–1081.
- Messens, J. & Silver, S. (2006). *J. Mol. Biol.* **362**, 1–17.
- Moreno, E. & Moriyon, I. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 1–3.
- Mukhopadhyay, R., Rosen, B. P., Phung, L. T. & Silver, S. (2002). *FEMS Microbiol. Rev.* **26**, 311–325.
- Myler, P. J., Stacy, R., Stewart, L. J., Staker, B. L., Van Voorhis, W. C. & Buchko, G. W. (2009). *Infect. Disord. Drug Targets*, **9**, 493–506.
- Nakano, S., Erwin, K. N., Ralle, M. & Zuber, P. (2005). *Mol. Microbiol.* **55**, 498–510.
- Nakano, S., Küster-Schöck, E., Grossman, A. D. & Zuber, P. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 13603–13608.
- Rubinstein, E., Lang, R., Shasha, B., Hagar, B., Diamanstein, L., Joseph, G., Anderson, M. & Harrison, K. (1991). *Antimicrob. Agents Chemother.* **35**, 1925–1927.



**Figure 9**

Surface representation of the structure of *Bm*-YffB shown with a labeled transparent cartoon representation. Regions that may potentially be important for the function of the protein are labeled. The lone equivalent residue to those present in the arginine triad of *Ec*-ArsC, Arg96, is colored black, residues in the potential  $\text{GX}_3\text{CX}_3\text{K}$  catalytic sequence motif are colored yellow and the residues perturbed by the addition of reduced glutathione (labeled in Fig. 8) are colored red.

- Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. (2002). *J. Biomol. NMR*, **24**, 149–154.
- Shi, H., Shi, X. & Liu, K. J. (2004). *Mol. Cell. Biochem.* **255**, 67–78.
- Teplyakov, A., Pullalarevu, S., Obmolova, G., Doseeva, V., Galkin, A., Herzberg, O., Dauter, M., Dauter, Z. & Gilliland, G. L. (2004). *BMC Struct. Biol.* **4**, 5.
- Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. & Sykes, B. D. (1995). *J. Biomol. NMR*, **6**, 135–140.
- Woody, R. W. (1974). In *Peptides, Polypeptides and Proteins*, edited by E. R. Blout, F. A. Bovey, M. Goodman & N. Lotan. New York: John Wiley & Sons.
- Young, E. J. (1995). *Clin. Infect. Dis.* **21**, 283–289.
- Zuiderweg, E. R. (2002). *Biochemistry*, **41**, 1–7.