

Herbert L. Axelrod,<sup>a,b</sup> Debanu Das,<sup>a,b</sup>  
 Polat Abdubek,<sup>b,c</sup> Tamara  
 Astakhova,<sup>b,d</sup> Constantina Bakolitsa,<sup>b,e</sup>  
 Dennis Carlton,<sup>b,f</sup> Connie Chen,<sup>b,c</sup>  
 Hsiu-Ju Chiu,<sup>a,b</sup> Thomas Clayton,<sup>b,f</sup>  
 Marc C. Deller,<sup>b,f</sup> Lian Duan,<sup>b,d</sup> Kyle  
 Ellrott,<sup>b,d</sup> Carol L. Farr,<sup>b,f</sup> Julie  
 Feuerhelm,<sup>b,c</sup> Joanna C. Grant,<sup>b,c</sup> Anna  
 Grzechnik,<sup>b,f</sup> Gye Won Han,<sup>b,f</sup> Lukasz  
 Jaroszewski,<sup>b,d,e</sup> Kevin K. Jin,<sup>a,b</sup>  
 Heath E. Klock,<sup>b,c</sup> Mark W. Knuth,<sup>b,c</sup>  
 Piotr Kozbial,<sup>b,e</sup> S. Sri Krishna,<sup>b,d,e</sup>  
 Abhinav Kumar,<sup>a,b</sup> Winnie W. Lam,<sup>a,b</sup>  
 David Marciano,<sup>b,f</sup> Daniel  
 McMullan,<sup>b,g</sup> Mitchell D. Miller,<sup>a,b</sup>  
 Andrew T. Morse,<sup>b,d</sup> Edward  
 Nigoghossian,<sup>b,c</sup> Amanda Nopakun,<sup>b,f</sup>  
 Linda Okach,<sup>b,c</sup> Christina Puckett,<sup>b,c</sup>  
 Ron Reyes,<sup>a,b</sup> Natasha Sefcovic,<sup>b,e</sup>  
 Henry J. Tien,<sup>b,f</sup> Christine B. Trame,<sup>a,b</sup>  
 Henry van den Bedem,<sup>a,b</sup> Dana  
 Weekes,<sup>b,e</sup> Tiffany Wooten,<sup>b,c</sup>  
 Qingping Xu,<sup>a,b</sup> Keith O. Hodgson,<sup>b,h</sup>  
 John Wooley,<sup>b,d</sup> Marc-André Elsliger,<sup>b,f</sup>  
 Ashley M. Deacon,<sup>a,b</sup> Adam  
 Godzik,<sup>b,d,e</sup> Scott A. Lesley,<sup>b,c,f</sup> and  
 Ian A. Wilson<sup>b,f,\*</sup>

<sup>a</sup>Stanford Synchrotron Radiation Lightsources, SLAC  
 National Accelerator Laboratory, Menlo Park, CA,  
 USA, <sup>b</sup>Joint Center for Structural Genomics,  
 http://www.jcsg.org, USA, <sup>c</sup>Protein Sciences  
 Department, Genomics Institute of the Novartis  
 Research Foundation, San Diego, CA, USA, <sup>d</sup>Center  
 for Research in Biological Systems, University of  
 California, San Diego, La Jolla, CA, USA, <sup>e</sup>Program  
 on Bioinformatics and Systems Biology, Sanford–  
 Burnham Medical Research Institute, La Jolla, CA,  
 USA, <sup>f</sup>Department of Molecular Biology, The  
 Scripps Research Institute, La Jolla, CA, USA,  
<sup>g</sup>Protein Therapeutics Department, Genomics  
 Institute of the Novartis Research Foundation, San  
 Diego, CA, USA, and <sup>h</sup>Photon Science, SLAC  
 National Accelerator Laboratory, Menlo Park, CA,  
 USA

Correspondence e-mail: wilson@scripps.edu

Received 3 March 2010

Accepted 27 May 2010

**PDB References:** B8FYU2\_DESHY, 2glz;  
 Q9HJ63\_THEAC, 2gvi; Q2LQ23\_SYNAS,  
 3d00.

## Structures of three members of Pfam PF02663 (FmdE) implicated in microbial methanogenesis reveal a conserved $\alpha+\beta$ core domain and an auxiliary C-terminal treble-clef zinc finger

Examination of the genomic context for members of the FmdE Pfam family (PF02663), such as the protein encoded by the *fmdE* gene from the methanogenic archaeon *Methanobacterium thermoautotrophicum*, indicates that 13 of them are co-transcribed with genes encoding subunits of molybdenum formylmethanofuran dehydrogenase (EC 1.2.99.5), an enzyme that is involved in microbial methane production. Here, the first crystal structures from PF02663 are described, representing two bacterial and one archaeal species: B8FYU2\_DESHY from the anaerobic dehalogenating bacterium *Desulfitobacterium hafniense* DCB-2, Q2LQ23\_SYNAS from the syntrophic bacterium *Syntrophus aciditrophicus* SB and Q9HJ63\_THEAC from the thermoacidophilic archaeon *Thermoplasma acidophilum*. Two of these proteins, Q9HJ63\_THEAC and Q2LQ23\_SYNAS, contain two domains: an N-terminal thioredoxin-like  $\alpha+\beta$  core domain (NTD) consisting of a five-stranded, mixed  $\beta$ -sheet flanked by several  $\alpha$ -helices and a C-terminal zinc-finger domain (CTD). B8FYU2\_DESHY, on the other hand, is composed solely of the NTD. The CTD of Q9HJ63\_THEAC and Q2LQ23\_SYNAS is best characterized as a treble-clef zinc finger. Two significant structural differences between Q9HJ63\_THEAC and Q2LQ23\_SYNAS involve their metal binding. First, zinc is bound to the putative active site on the NTD of Q9HJ63\_THEAC, but is absent from the NTD of Q2LQ23\_SYNAS. Second, whereas the structure of the CTD of Q2LQ23\_SYNAS shows four Cys side chains within coordination distance of the Zn atom, the structure of Q9HJ63\_THEAC is atypical for a treble-cleft zinc finger in that three Cys side chains and an Asp side chain are within coordination distance of the zinc.

### 1. Introduction

The Pfam family PF02663 (FmdE; Finn *et al.*, 2008) currently contains 204 proteins from 74 bacterial and 39 archaeal species (Pfam v.24; http://pfam.sanger.ac.uk/). In thermophilic methanogenic archaea, co-transcription of the *fmdE* gene with downstream genes encoding catalytic subunits of formylmethanofuran dehydrogenase (EC 1.2.99.5) has been reported (Hochheimer *et al.*, 1996, 1998; Vorholt *et al.*, 1996). Formylmethanofuran dehydrogenase is a multi-subunit enzyme that contains tungsten (Bertram *et al.*, 1994) or molybdenum as well as iron–sulfur clusters (Hochheimer *et al.*, 1996), and catalyzes the first step in the formation of methane from carbon dioxide in methanogenic and sulfate-reducing microorganisms (Thauer *et al.*, 2008; Hallam *et al.*, 2004; Liu & Whitman, 2008). The proximity of *fmdE* to genes encoding the catalytic subunits suggests a role in methanogenesis for proteins in PF02663. These observations are consistent with environmental genomic studies, in which the *fmdE* gene was identified in microorganisms from anaerobic marine sediments which are believed to have a significant impact on the global environment by consuming methane (reverse methanogenesis), affecting the levels of atmospheric methane as a greenhouse gas (Hallam *et al.*, 2004).

The genomes of many nonmethanogenic microorganisms also encode proteins in PF02663. Genes from three microbes, *DSY1837* from *Desulfitobacterium hafniense* DCB-2 (UniProt B8FYU2\_DESHY), an anaerobic dehalogenating bacterium; *Ta1109* from *Thermoplasma acidophilum* (UniProt Q9HJ63\_THEAC), a thermo-

acidophilic archaeon; and *SYN\_00638* from *Syntrophus aciditrophicus* SB (UniProt Q2LQ23\_SYNAS), a syntrophic bacterium, encode proteins with molecular weights of 17.4, 23.1 and 21.5 kDa with calculated isoelectric points of 5.95, 6.13 and 6.21, respectively. Their structures, which are the first reported for the PF02663 Pfam family, were determined using the semi-automated high-throughput pipeline of the Joint Center for Structural Genomics (JCSG; Lesley *et al.*, 2002) as part of the NIH National Institute of General Medical Sciences' Protein Structure Initiative (PSI).

## 2. Materials and methods

### 2.1. Protein production and crystallization

Clones for *DSY1837*, *Ta1109* and *SYN\_00638* were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method (Klock *et al.*, 2008). The gene encoding *DSY1837* (GenBank YP\_002459451.1; UniProt B8FYU2\_DESHY) was amplified by polymerase chain reaction (PCR) from *D. hafniense* DCB-2 genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and I-PIPE primers (forward, 5'-ctgtacttccagggcATGTGCGTAGAAAAACCCTTGGGAAC-3'; reverse, 5'-aattaagtcgcttaAACTATTTTACTCAGTTGTCCCGGA-3'; target sequence in upper case) that included sequences for the predicted 5' and 3' ends. The expression vector pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSDK-IHHHHHHENLYFQ/G), was PCR-amplified with V-PIPE (Vector) primers. V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *Escherichia coli* GeneHogs (Invitrogen) competent cells were transformed with the V-PIPE/I-PIPE mixture and dispensed onto selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Expression was performed in a selenomethionine-containing medium at 310 K. Cells were induced after 1.5 h using 0.11% (*w/v*) arabinose and were allowed to grow for an additional 3 h before harvesting. Selenomethionine was incorporated *via* inhibition of methionine biosynthesis (Van Duyne *et al.*, 1993), which does not require a methionine-auxotrophic strain.

At the end of fermentation, lysozyme was added to the culture to a final concentration of 250  $\mu\text{g ml}^{-1}$  and the cells were harvested and frozen. After one freeze-thaw cycle, the cells were sonicated in lysis buffer [50 mM HEPES pH 8.0, 50 mM NaCl, 10 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine-HCl (TCEP)] and the lysate was clarified by centrifugation at 32 500g for 30 min. The soluble fraction was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer, the resin was washed with wash buffer [50 mM HEPES pH 8.0, 300 mM NaCl, 40 mM imidazole, 10% (*v/v*) glycerol, 1 mM TCEP] and the protein was eluted with elution buffer [20 mM HEPES pH 8.0, 300 mM imidazole, 10% (*v/v*) glycerol, 1 mM TCEP]. The eluate was buffer-exchanged with TEV buffer (20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP) using a PD-10 column (GE Healthcare) and incubated with 1 mg TEV protease per 15 mg of eluted protein. The protease-treated eluate was run over nickel-chelating resin (GE Healthcare) pre-equilibrated with HEPES crystallization buffer (20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP) and the resin was washed with the same buffer. The flowthrough and wash fractions were combined and concentrated to 15 mg  $\text{ml}^{-1}$  by centrifugal ultrafiltration (Millipore) for crystallization trials. B8FYU2\_DESHY was crystallized at 277 K using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard JCSG crystallization protocols (Lesley *et al.*, 2002). The crystallization reagent used was composed of 0.2 M MgCl<sub>2</sub>

and 20.0% PEG 3350. Ethylene glycol was added to the crystal as a cryoprotectant to a final concentration of 10% (*v/v*). Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM; Cohen *et al.*, 2002) at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, California, USA). The crystal was indexed in the primitive orthorhombic space group *P*<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub>. The oligomeric state of B8FYU2\_DESHY in solution was determined using a 1 × 30 cm Superdex 200 size-exclusion column (GE Healthcare; Klock *et al.*, 2008) coupled with miniDAWN (Wyatt Technology) static light-scattering (SEC/SLS) and Optilab differential refractive-index detectors (Wyatt Technology). The mobile phase consisted of 20 mM Tris pH 8.0, 150 mM NaCl and 0.02% (*w/v*) sodium azide.

The *Ta1109* gene (GenBank CAC12236.1; UniProt ID Q9HJ63\_THEAC) was amplified from *T. acidophilum* DSM1728 genomic DNA. Cloning (forward primer, 5'-ctgtacttccagggcATGGAGAAA-CTGAATTTTCGGAATTCCAG-3'; reverse primer, 5'-aattaagtcgcttaTTTCTTGCCGTAGTAATCAGGCTTGCAC-3'; target sequence in upper case), expression and purification were performed as described for B8FYU2\_DESHY. Purified Q9HJ63\_THEAC was concentrated to 14 mg  $\text{ml}^{-1}$  for crystallization trials and was crystallized at 277 K using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard JCSG crystallization protocols (Lesley *et al.*, 2002). The crystallization reagent used was composed of 0.2 M magnesium nitrate and 20.0% PEG 3350. The crystal was indexed in the monoclinic space group *C*<sub>2</sub>. A second crystal was obtained using a solution consisting of 10.0% PEG 8000, 0.2 M zinc acetate and 0.1 M MES pH 6.0. These crystals were indexed in the *I*-centered orthorhombic space group *I*222. A third crystal was grown in a solution consisting of 0.2 M magnesium nitrate and 20.0% PEG 3350 and was indexed in the tetragonal space group *P*<sub>4</sub><sub>2</sub><sub>2</sub><sub>1</sub>. Ethylene glycol was added to the crystals as cryoprotectant to a final concentration of 15% (*v/v*). Initial screening for diffraction and oligomeric state determination were performed as described for B8FYU2\_DESHY.

The *SYN\_00638* gene (GenBank CP000252; UniProt Q2LQ23\_SYNAS) was amplified from *S. aciditrophicus* SB genomic DNA. Cloning (forward primer, 5'-ctgtacttccagggcATGACAGCACGTAA-TATTTTGTCTTAC-3'; reverse primer, 5'-aattaagtcgcttaAAGAT-AAGGCGACCTCCCTGGCAGCTC-3'; target sequence in upper case), expression and purification were performed as described for B8FYU2\_DESHY. Purified Q2LQ23\_SYNAS was concentrated to 20 mg  $\text{ml}^{-1}$  for crystallization trials and was crystallized at 277 K using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard JCSG crystallization protocols (Lesley *et al.*, 2002). The crystallization reagent was composed of 0.01 M nickel chloride, 20.0% PEG MME 2000 and 0.1 M Tris pH 8.5. Glycerol was added to the crystal as a cryoprotectant to a final concentration of 10% (*v/v*). Initial screening for diffraction and oligomeric state determination were carried out as described for B8FYU2\_DESHY. The crystal was indexed in the tetragonal space group *P*<sub>4</sub><sub>1</sub><sub>2</sub><sub>1</sub>.

### 2.2. Data collection, structure solution and refinement

X-ray diffraction data were collected on beamline 9-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) at wavelengths corresponding to the high-energy remote ( $\lambda_1$ ), inflection ( $\lambda_2$ ) and peak ( $\lambda_3$ ) wavelengths of a three-wavelength selenium multi-wavelength anomalous diffraction (Se-MAD) experiment for the *P*<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> crystal form of B8FYU2\_DESHY and the *C*<sub>2</sub> crystal form of Q9HJ63\_THEAC. Three-wavelength Se-MAD data were collected on beamline 11-1 at SSRL for Q2LQ23\_SYNAS. Additional

**Table 1**

Summary of crystal parameters, data-collection and refinement statistics for B8FYU2\_DESHY (PDB entry 2glz).

Values in parentheses are for the highest resolution shell.

	$\lambda_1$ MADSe	$\lambda_2$ MADSe	$\lambda_3$ MADSe
Space group	$P2_12_12_1$		
Unit-cell parameters (Å)	$a = 46.42, b = 84.79, c = 100.71$		
Data collection			
Wavelength (Å)	0.91837	0.97927	0.97905
Resolution range (Å)	28.26–1.45 (1.49–1.45)	28.25–1.49 (1.53–1.49)	28.27–1.49 (1.53–1.49)
No. of observations	524343	482691	501396
No. of unique reflections	71199	65618	65729
Completeness (%)	99.9 (99.9)	99.9 (99.5)	99.9 (99.7)
Mean $I/\sigma(I)$	17.9 (1.6)	18.0 (2.1)	18.2 (2.1)
$R_{\text{merge}}$ on $I^\dagger$ (%)	7.1 (73.0)	7.6 (58.1)	7.4 (79.8)
$R_{\text{meas}}$ on $I^\ddagger$ (%)	7.6 (82.3)	8.1 (65.6)	7.9 (85.6)
Model and refinement statistics			
Resolution range (Å)	27.2–1.45		
No. of reflections (total)	71126		
No. of reflections (test)	3593		
Completeness (%)	99.8		
Data set used in refinement	$\lambda_1$ MADSe		
Cutoff criterion	$ F  > 0$		
$R_{\text{cryst}}^\S$	0.171		
$R_{\text{free}}^\P$	0.198		
Stereochemical parameters			
Restraints (r.m.s.d. observed)			
Bond angles (°)	1.86		
Bond lengths (Å)	0.017		
Average isotropic $B$ value (Å <sup>2</sup> )	27.1		
ESU $^\ddagger$ based on $R_{\text{free}}$ (Å)	0.061		
Protein residues/atoms	297/2441		
Waters/solvent molecules/ions	431/18/4		

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the scaled intensity of the  $i$ th measurement and  $\langle I(hkl) \rangle$  is the mean intensity for that reflection.  $^\ddagger R_{\text{meas}}$  is the redundancy-independent  $R_{\text{merge}}$  (Diederichs & Karplus, 1997; Weiss, 2001).  $^\S R_{\text{cryst}} = \sum_{hkl} |F_{\text{obs}}| - |F_{\text{calc}}| / \sum_{hkl} |F_{\text{obs}}|$ , where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the calculated and observed structure-factor amplitudes, respectively.  $^\P R_{\text{free}}$  is the same as  $R_{\text{cryst}}$  but for 5.1% of the total reflections chosen at random and omitted from refinement.  $^\ddagger$  Estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994; Cruickshank, 1999).

diffraction data for Q9HJ63\_THEAC were collected from the two other crystal forms ( $I222$  and  $P4_22_12_1$ ) on beamlines 11-1 and 9-2 at SSRL at wavelengths of 1.00 and 0.9790 Å, respectively. MAD phasing for Q9HJ63\_THEAC was carried out using the  $C2$  crystal data and further refinement was performed using the  $I222$  data at a higher resolution of 1.87 Å after molecular replacement with *Phaser* (McCoy, 2007) using the model obtained from the  $C2$  data. All data sets were collected at 100 K using either an ADSC Quantum 315 detector (beamline 11-1) or a MAR Mosaic 325 CCD detector (beamline 9-2). The data were integrated and scaled using either *MOSFLM* (Leslie, 1992) and *SCALA* from the *CCP4* program suite (Collaborative Computational Project, Number 4, 1994) or the *XDS* and *XSCALE* programs (Kabsch, 1993, 2010a,b). Data statistics are summarized in Table 1 for B8FYU2\_DESHY, in Tables 2 and 3 for Q9HJ63\_THEAC and in Table 4 for Q2LQ23\_SYNAS. The selenium substructures for the three proteins were solved with *SHELXD* (Sheldrick, 2008) and the MAD phases were refined with *autoSHARP* for Q9HJ63\_THEAC and Q2LQ23\_SYNAS (Vonnrhein *et al.*, 2007) and *SOLVE* (Terwilliger & Berendzen, 1999) for B8FYU2\_DESHY. The mean figures of merit were 0.45, 0.37 and 0.35, respectively. Automatic model building was performed with either *ARP/wARP* (Cohen *et al.*, 2004) or *RESOLVE* (Terwilliger, 2002). Model completion was performed using *Coot* (Emsley & Cowtan, 2004) and refinement was accomplished using *REFMAC5* (Winn *et al.*, 2003). Refinement statistics are summarized in Tables 1, 3 and 4 for B8FYU2\_DESHY, Q9HJ63\_THEAC and Q2LQ23\_SYNAS, respectively.

**Table 2**

Summary of crystal parameters and data-collection statistics for Q9HJ63\_THEAC in the  $C2$  crystal form.

Values in parentheses are for the highest resolution shell.

	$\lambda_1$ MADSe	$\lambda_2$ MADSe	$\lambda_3$ MADSe
Space group	$C2$		
Unit-cell parameters (Å, °)	$a = 108.68, b = 52.63, c = 88.83, \beta = 121.3$		
Data collection			
Wavelength (Å)	0.91837	0.97180	0.97903
Resolution range (Å)	29.67–2.00 (2.05–2.00)	29.66–2.00 (2.05–2.00)	29.66–2.00 (2.05–2.00)
No. of observations	78977	78479	78656
No. of unique reflections	28904	28864	28891
Completeness (%)	99.1 (98.9)	99.0 (96.7)	99.0 (97.3)
Mean $I/\sigma(I)$	9.4 (2.3)	8.5 (2.1)	8.6 (2.0)
$R_{\text{merge}}$ on $I^\dagger$ (%)	8.1 (50.0)	9.2 (53.6)	9.6 (58.1)
$R_{\text{meas}}$ on $I^\ddagger$ (%)	10.1 (62.3)	11.4 (66.9)	12.0 (72.5)

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the scaled intensity of the  $i$ th measurement and  $\langle I(hkl) \rangle$  is the mean intensity for that reflection.  $^\ddagger R_{\text{meas}}$  is the redundancy-independent  $R_{\text{merge}}$  (Diederichs & Karplus, 1997; Weiss, 2001).

### 2.3. Validation and deposition

The quality of the crystal structure was analyzed using the *JCSG Quality Control* server (see <http://smb.slac.stanford.edu/jcsg/QC/>). This server verifies the stereochemical quality of the model using *AutoDepInputTool* (Yang *et al.*, 2004), *MolProbity* (Chen *et al.*, 2010) and *WHAT IF* v.5.0 (Vriend, 1990), the agreement between the atomic model and the data using *SFCHECK* v.4.0 (Vaguine *et al.*, 1999) and *RESOLVE* (Terwilliger, 2002), the protein sequence using *ClustalW* (Chenna *et al.*, 2003), the atom occupancies using *MOLEMAN2* (Kleywegt *et al.*, 2001) and the consistency of NCS pairs. It also evaluates differences in  $R_{\text{cryst}}/R_{\text{free}}$ , expected  $R_{\text{free}}/R_{\text{cryst}}$  and maximum/minimum  $B$  values by parsing the refinement log file and PDB header. The *EBI PISA* server (Krissinel & Henrick, 2007) was used to analyze the protein quaternary structure. Figs. 1(a), 1(b) and 1(c) were adapted from *PDBsum* (Laskowski, 2009) and the other figures were prepared using *PyMOL* (DeLano Scientific). Atomic coordinates and experimental structure factors for B8FYU2\_DESHY at 1.45 Å resolution, Q9HJ63\_THEAC at 1.87 Å resolution and Q2LQ23\_SYNAS at 1.90 Å resolution have been deposited in the PDB and are accessible under codes 2glz, 2gvi and 3d00, respectively.

## 3. Results and discussion

### 3.1. Overall structures

The crystal structure of B8FYU2\_DESHY (Fig. 1a) was determined by MAD at 1.45 Å resolution. Data-collection, model and refinement statistics are summarized in Table 1. The final model includes two protein molecules (residues 3–151 for chain A; residues 4–151 for chain B), 18 ethylene glycol molecules, one Zn atom, one Ni atom and 427 water molecules in the asymmetric unit. No electron density was observed for a few residues at the N- and C-termini of both chains (GlyA0, MseA1, CysA2, ValA152, GlyB0, MseB1, CysB2, ValB3 and ValB152) or for side-chain atoms of ValA3, GluA4, AspA43, ArgA117, GluA118, ArgA119, IleA151, GluB4, AspB43, HisB111, AspB113, ArgB117 and IleB151. The Matthews coefficient ( $V_M$ ; Matthews, 1968) was 2.82 Å<sup>3</sup> Da<sup>-1</sup> and the estimated solvent content was 56.4%. The Ramachandran plot produced by *MolProbity* (Davis *et al.*, 2004) showed that 99% of the residues are in favored regions, with no outliers. B8FYU2\_DESHY is composed of five  $\beta$ -strands ( $\beta 1$ – $\beta 5$ ) and six  $\alpha$ -helices ( $\alpha 1$ – $\alpha 6$ ) (Fig. 1a). The total

**Table 3**

Summary of crystal parameters, data-collection and refinement statistics for Q9HJ63\_THEAC (PDB entry 2gvi).

Values in parentheses are for the highest resolution shell.

	$\lambda_1$
Space group	<i>I</i> 222
Unit-cell parameters (Å)	<i>a</i> = 78.60, <i>b</i> = 97.65, <i>c</i> = 75.27
Data collection	
Wavelength (Å)	1.000
Resolution range (Å)	30.08–1.87 (1.94–1.87)
No. of observations	92701
No. of unique reflections	24247
Completeness (%)	99.8 (99.9)
Mean <i>I</i> σ( <i>I</i> )	10.5 (1.7)
<i>R</i> <sub>merge</sub> on <i>I</i> † (%)	10.3 (83.1)
<i>R</i> <sub>meas</sub> on <i>I</i> ‡ (%)	12.0 (96.8)
Model and refinement statistics	
Resolution range (Å)	30.1–1.87
No. of reflections (total)	24246
No. of reflections (test)	1229
Completeness (%)	99.7
Cutoff criterion	<i>F</i>   > 0
<i>R</i> <sub>cryst</sub> §	0.190
<i>R</i> <sub>free</sub> ¶	0.217
Stereochemical parameters	
Restraints (r.m.s.d. observed)	
Bond angles (°)	1.64
Bond lengths (Å)	0.014
Average isotropic <i>B</i> value (Å <sup>2</sup> )	31.1
ESU†† based on <i>R</i> <sub>free</sub> (Å)	0.12
Protein residues/atoms	201/1599
Waters/solvent molecules/ions	129/15/5

†  $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the scaled intensity of the *i*th measurement and  $\langle I(hkl) \rangle$  is the mean intensity for that reflection. ‡  $R_{\text{meas}}$  is the redundancy-independent  $R_{\text{merge}}$  (Diederichs & Karplus, 1997; Weiss, 2001). §  $R_{\text{cryst}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$ , where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the calculated and observed structure-factor amplitudes, respectively. ¶  $R_{\text{free}}$  is the same as  $R_{\text{cryst}}$  but for 5.1% of the total reflections chosen at random and omitted from refinement. †† Estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994; Cruickshank, 1999).

$\beta$ -sheet and  $\alpha$ -helical contents are 24% and 58%, respectively. The monomer consists of a central five-stranded, mixed  $\beta$ -sheet (21345 topology) with one solvent-exposed face, while the other is covered by three  $\alpha$ -helices. A distinctive feature of the structure is the protrusion of two helices ( $\alpha 4$  and  $\alpha 5$ ) and a connecting loop (residues 99–138) from the core of each molecule.

The crystal structure of Q9HJ63\_THEAC (Fig. 1*b*) was initially determined by MAD from the *C*2 crystal form at 2.0 Å resolution. Molecular replacement was then used to determine the structure of the *I*222 crystal form at 1.87 Å resolution. Data-collection, model and refinement statistics are summarized in Tables 2 and 3. The final model includes one protein molecule (residues 1–201), one unknown ligand (UNL), five Zn atoms, six ethylene glycol molecules, eight acetate ions and 129 water molecules in the asymmetric unit. No electron density was observed for a few residues at the N- and C-termini (Gly0, Lys202 and Lys203) or for side-chain atoms of Mse1, Glu2, Lys3, Arg117, Glu10, Lys35, Arg155, Glu163 and Lys192. The Matthews coefficient ( $V_M$ ; Matthews, 1968) for the *I*222 form was 2.87 Å<sup>3</sup> Da<sup>-1</sup> and the estimated solvent content was 56.8%. The Ramachandran plot produced by *MolProbity* (Davis *et al.*, 2004) showed that 99% of the residues are in favored regions, with no outliers. Q9HJ63\_THEAC is composed of 11  $\beta$ -strands ( $\beta 1$ – $\beta 11$ ) and ten  $\alpha$ -helices ( $\alpha 1$ – $\alpha 10$ ) (Fig. 1*b*). The total  $\beta$ -sheet,  $\alpha$ -helical and  $3_{10}$ -helical contents are 24, 58 and 2.5%, respectively. In addition to the N-terminal  $\alpha + \beta$  core domain (NTD; residues 1–157), which is similar to that of B8FYU2\_DESHY, Q9HJ63\_THEAC also has a C-terminal domain (CTD) with a treble-clef, zinc finger-like motif

**Table 4**

Summary of crystal parameters, data-collection and refinement statistics for Q2LQ23\_SYNAS (PDB code 3d00).

Values in parentheses are for the highest resolution shell.

	$\lambda_1$ MADSe	$\lambda_2$ MADSe
Space group	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2	
Unit-cell parameters (Å)	<i>a</i> = 54.36, <i>b</i> = 54.36, <i>c</i> = 136.72	
Data collection		
Wavelength (Å)	0.9184	0.9782
Resolution range (Å)	29.4–1.90 (1.95–1.90)	29.4–1.90 (1.95–1.90)
No. of observations	118636	118329
No. of unique reflections	16954	16972
Completeness (%)	99.9 (99.9)	99.9 (99.9)
Mean <i>I</i> σ( <i>I</i> )	15.6 (2.0)	16.2 (2.0)
<i>R</i> <sub>merge</sub> on <i>I</i> † (%)	7.8 (113.4)	7.7 (106.1)
<i>R</i> <sub>meas</sub> on <i>I</i> ‡ (%)	8.4 (122.3)	8.4 (114.5)
Model and refinement statistics		
Resolution range (Å)	29.4–1.90	
No. of reflections (total)	16902	
No. of reflections (test)	855	
Completeness (%)	99.9	
Data set used in refinement	$\lambda_1$ MADSe	
Cutoff criterion	<i>F</i>   > 0	
<i>R</i> <sub>cryst</sub> §	0.233	
<i>R</i> <sub>free</sub> ¶	0.268	
Stereochemical parameters		
Restraints (r.m.s.d. observed)		
Bond angles (°)	1.60	
Bond lengths (Å)	0.019	
Average isotropic <i>B</i> value (Å <sup>2</sup> )	35.2	
ESU†† based on <i>R</i> <sub>free</sub> (Å)	0.16	
Protein residues/atoms	184/1408	
Waters/ions	42/2	

†  $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the scaled intensity of the *i*th measurement and  $\langle I(hkl) \rangle$  is the mean intensity for that reflection. ‡  $R_{\text{meas}}$  is the redundancy-independent  $R_{\text{merge}}$  (Diederichs & Karplus, 1997; Weiss, 2001). §  $R_{\text{cryst}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$ , where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the calculated and observed structure-factor amplitudes, respectively. ¶  $R_{\text{free}}$  is the same as  $R_{\text{cryst}}$  but for 5.1% of the total reflections chosen at random and omitted from refinement. †† Estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994; Cruickshank, 1999).

(Grishin, 2001; residues 169–201); it is connected to the N-terminal domain *via* an 11-residue linker.

The crystal structure of Q2LQ23\_SYNAS (Fig. 1*c*) was determined by MAD at 1.90 Å resolution. Data-collection, model and refinement statistics are summarized in Table 4. The final model includes one protein molecule (residues 1–190), one chloride anion, one Zn atom and 42 water molecules in the asymmetric unit. The smaller than expected number of ordered water molecules for a 1.9 Å resolution structure coincides with elevated  $R_{\text{cryst}}$  and  $R_{\text{free}}$  values of 23.3% and 26.8%, respectively. One possible explanation for the larger than expected *R* values is the anisotropy of the diffraction intensities, with a spread in the values of the three principal components of 21.4 Å<sup>2</sup> and with diffraction intensity falling off more significantly in the *a*\* and *b*\* directions compared with the *c*\* direction. No electron density was observed for residues A121–A126 or for side-chain atoms of GluA16, LysA17, AspA48, ArgA56, GluA95, LysA105, GlnA110, LysA118, LysA120, GluA128, ArgA129, LysA132, GluA136, LysA148, LysA150, GluA155, LysA156, LysA157, HisA158, LysA159 and LysA161. The Matthews coefficient ( $V_M$ ; Matthews, 1968) for Q2LQ23\_SYNAS was 2.33 Å<sup>3</sup> Da<sup>-1</sup> and the estimated solvent content was 47.1%. The Ramachandran plot produced by *MolProbity* showed that 96.1% of the residues are in favored regions, with no outliers. Q2LQ23\_SYNAS (Fig. 1*c*) is composed of seven  $\beta$ -strands ( $\beta 1$ – $\beta 7$ ) and nine  $\alpha$ -helices ( $\alpha 1$ – $\alpha 9$ ). The total  $\beta$ -sheet,  $\alpha$ -helical and  $3_{10}$ -helical contents are 18, 56 and 4.9%, respectively. Q2LQ23\_SYNAS displays a similar architecture to Q2HJ63\_THEAC, with a larger NTD (residues 1–154) and a smaller, treble-clef zinc-finger

domain CTD coupled together through a nine-residue linker (residues 155–163). The linkers in Q9HJ63\_THEAC and Q2LQ23\_SYNAS separate the NTD and CTD domains so that the closest edges of the two domains are  $\sim 20$  Å apart.

### 3.2. Oligomerization

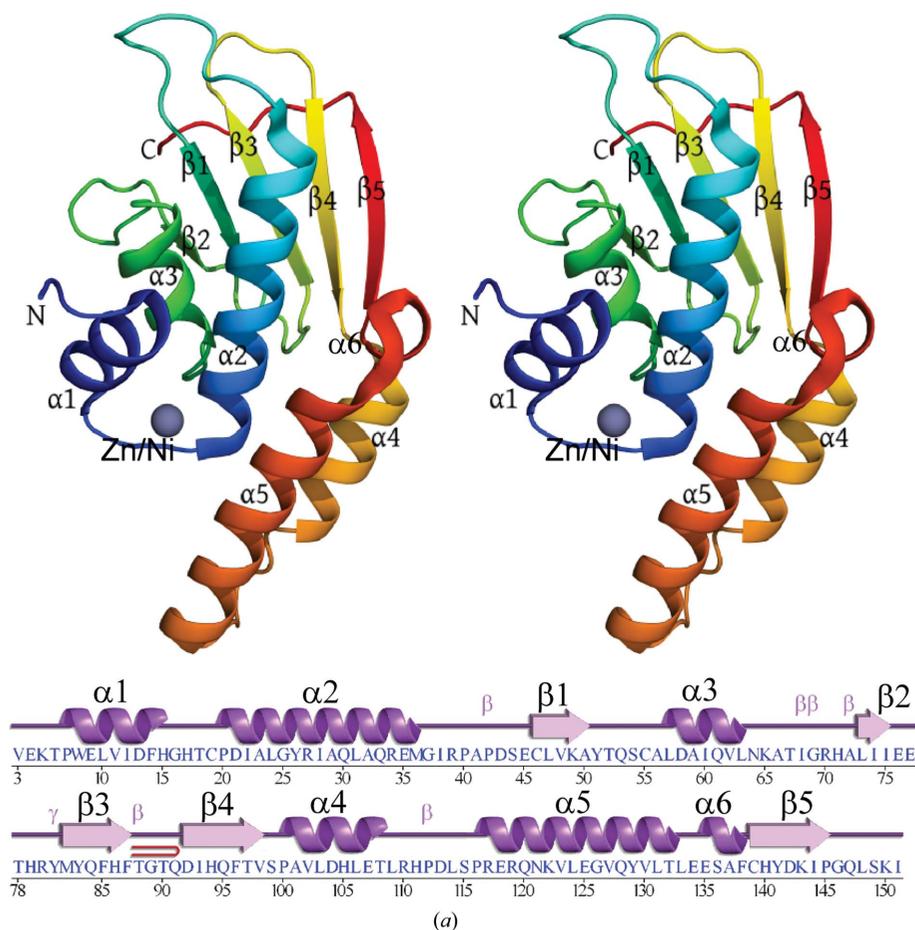
B8FYU2\_DESHY, Q9HJ63\_THEAC and Q2LQ23\_SYNAS contain stable dimeric interfaces of 2030, 5860 and 4350 Å<sup>2</sup>, respectively, as predicted by *PISA* (Krissinel & Henrick, 2007). Analytical size-exclusion chromatography coupled with static light scattering also supports these assignments in solution, suggesting that a dimer is the functionally relevant oligomer for each. The asymmetric unit dimer for B8FYU2\_DESHY is approximately S-shaped, with several close-range monomer–monomer interactions between residues on helix  $\alpha 3$  (Fig. 2*a*). The dimer has two prominent C-shaped grooves that extend along its surface parallel to the twofold axis; they are  $\sim 15$  Å wide and are exposed to solvent at either end (Fig. 2*a*). All crystal forms of Q9HJ63\_THEAC (Fig. 2*b*) and Q2LQ23\_SYNAS (Fig. 2*c*) show similar twofold-symmetric, domain-swapped dimers in which the NTD and the CTD of one polypeptide chain are separated by an 11-residue linker and the CTD is anchored to the NTD of the symmetry-related monomer. Analysis of the structures of the Q9HJ63\_THEAC and Q2LQ23\_SYNAS dimers using *CASTp*

(Dundas *et al.*, 2006) shows an  $\sim 20$  Å wide surface depression (Figs. 2*b* and 2*c*) that is large enough to accommodate a fairly large ligand.

### 3.3. Metal-ion binding in the NTD

A metal ion-binding site was identified at the bottom of the C-shaped groove in B8FYU2\_DESHY (Fig. 2*a*). The metal ion is solvent-accessible and within coordination distance of His15, His17, Cys19 and Cys55 (Fig. 2*a*, Table 5). X-ray anomalous scattering measurements indicated that the site had a mixed occupancy of zinc and nickel. The total occupancy of the zinc and nickel cations was reduced to 0.75 to match the observed scattering at this site, with a zinc:nickel ion stoichiometric ratio of 2.6:1 estimated from the ratio of their anomalous difference map peak heights. The guanidinium side chain of Arg70 from the other subunit in the dimer is within hydrogen-bonding distance of the carbonyl O atom of His15 and stacks parallel to the side chain of His17, which coordinates the metal (Fig. 2*a*).

X-ray fluorescence emission spectroscopy from the C2 crystals of Q9HJ63\_THEAC indicated the presence of zinc. To corroborate that zinc was bound at specific sites in the structure and not just in the bulk solvent, anomalous difference maps were calculated from data collected at wavelengths above and below the zinc X-ray absorption edge. One of the binding sites was located on the NTD (Fig. 2*b*, Table



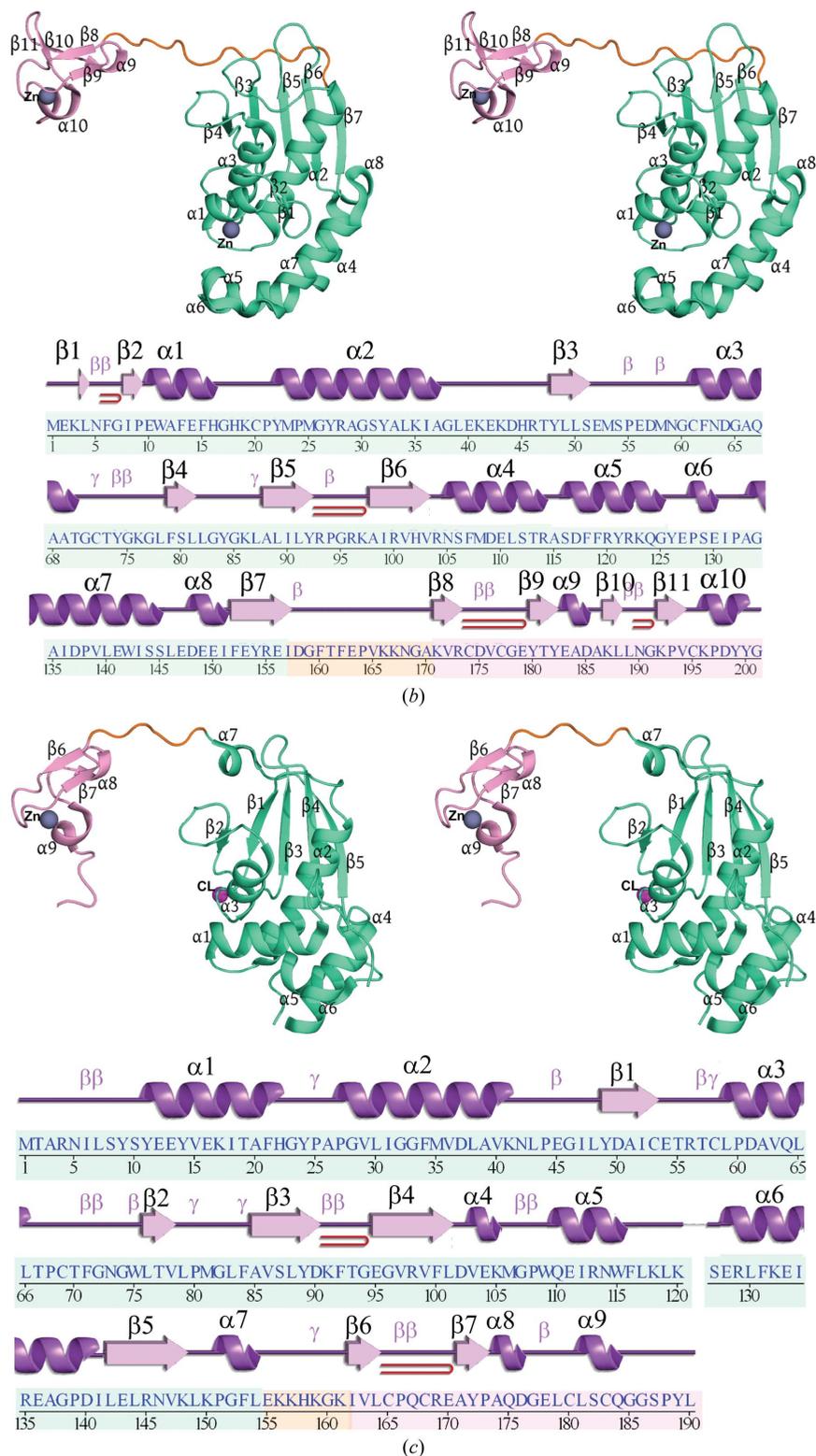
**Figure 1**

Crystal structures of (a) B8FYU2\_DESHY, (b) Q9HJ63\_THEAC and (c) Q2LQ23\_SYNAS. The polypeptide backbones are shown as stereo ribbon diagrams. Below the ribbon representations are the secondary-structure elements superimposed on the primary sequence. The  $\alpha$ -helices,  $3_{10}$ -helices,  $\beta$ -strands,  $\beta$ -turns and  $\gamma$ -turns are indicated.  $\beta$ -Hairpins are depicted as red loops. (a) For B8FYU2\_DESHY, the protein ribbon is color-coded from the N-terminus (blue) to the C-terminus (red). Helices  $\alpha 1$ – $\alpha 4$  and  $\beta$ -strands ( $\beta 1$ – $\beta 6$ ) are indicated. A dual-occupancy zinc/nickel-binding site in the vicinity of the putative active site on the  $\alpha + \beta$  core and the zinc-finger domain is shown as a gray sphere.

## structural communications

5) and a second on the CTD (Fig. 2*b*, Table 5). All three crystal forms show zinc binding at the same two sites, suggesting that these sites are

functionally relevant (note that two of the three crystal forms, C2 and P4<sub>2</sub>2<sub>1</sub>2, are devoid of exogenous zinc in the crystallization condi-



**Figure 1 (continued)**

(*b*) For Q9HJ63\_THEAC, helices  $\alpha 1$ – $\alpha 10$  and  $\beta$ -strands ( $\beta 1$ – $\beta 11$ ) are indicated. The subregions of the structure, the core domain (NTD), linker and C-terminal zinc-finger domain (CTD), and the background of the corresponding sequence are colored turquoise, orange and pink, respectively. Zn atoms are shown as gray spheres. (*c*) For Q2LQ23\_SYNAS, helices H1–H10 and  $\beta$ -strands ( $\beta 1$ – $\beta 7$ ) are indicated with subregions of the structure colored as in (*b*). A chloride ion in the vicinity of the putative active site is shown as a magenta sphere and the Zn atom bound to the zinc-finger domain is shown as a gray sphere.

**Table 5**

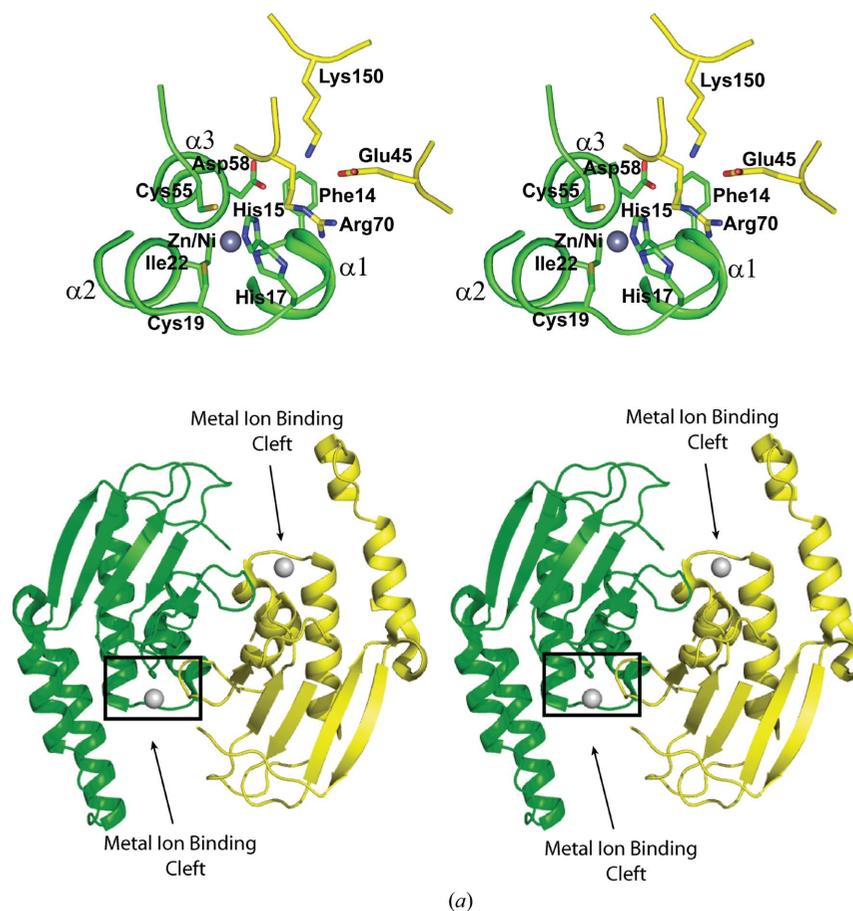
Metal-ion ligands and coordination geometry in the B8FYU2\_DESHY, Q9HJ63\_THEAC and Q2LQ23\_SYNAS structures.

Protein (UniProt designation)	Metal ion	Ligands	Interatomic distance (Å)	Coordination geometry
B8FYU2_DESHY	Ni	His15 NE2	2.2	Tetrahedral
		His17 NE2	1.9	
		Cys19 SG	2.6	
	Zn	Cys55 SG	2.2	
		His15 NE2	2.1	
		His17 NE2	1.9	
Q9HJ63_THEAC	Zn, N-terminal domain	Cys19 SG	2.4	Tetrahedral
		Cys55 SG	2.4	
		His16 NE2	2.0	
		His18 NE2	2.0	
	Zn, C-terminal domain	Cys20 SG	2.4	
		Cys61 SG	2.8	
		Cys174	2.3	
		Cys177	2.3	
		Cys195	2.3	
		Asp198	2.0	
Q2LQ23_SYNAS	Zn, C-terminal domain	Cys165	2.4	Tetrahedral
		Cys168	2.5	
		Cys180	2.4	
		Cys183	2.5	

tions). The *I*222 crystal form also showed four additional zinc-binding sites, which are likely to be attributable to the presence of zinc acetate in the crystallization experiments.

In Q9HJ63\_THEAC the zinc-binding site on the NTD is situated on a loop connecting the N-terminal  $\alpha$ -helices ( $\alpha 1$  and  $\alpha 2$ ). The zinc is within coordination distance of His16, His18, Cys20 and Cys61 (Fig. 2*b*). These side chains are conserved in B8FYU2\_DESHY, in which the NTD metal ion-binding site occupies a similar position. In the *I*222 crystal form of Q9HJ63\_THEAC, unexplained electron density near the zinc and Cys61 was modeled as an unknown ligand (UNL; Fig. 2*b*). The UNL is only 1.8 Å from the S atom of the conserved Cys61, which is consistent with a thioester bond between the protein and the UNL.

This binding site and the UNL are located within an elongated cleft on the surface of the dimer that is approximately 30 Å long and 10 Å wide (Fig. 2*b*). Each dimer contains two symmetry-related clefts positioned ~25 Å apart that are assembled from both subunits, including portions of the zinc-finger domain and its  $\beta$ -strand bridging the N- and C-terminal domains. In Q2LQ23\_SYNAS no zinc is bound to the NTD. It is worth noting that two of the zinc-binding residues in B8FYU2\_DESHY and Q9HJ63\_THEAC are not conserved in Q2LQ23\_SYNAS: His15 and Cys19 (B8FYU2\_DESHY numbering) are replaced by Tyr and Ala, respectively (Fig. 2*c*). Instead, an occupied anion-binding site was identified in Q2LQ23\_SYNAS (Fig. 2*c*) and was modeled as a chloride based on the electron density being within 3.5 Å of the polypeptide backbone N atoms of Arg56 and Gly82 and the presence of chloride in the crystallization reagent. The chloride is bound near the end of the central  $\beta$ -sheet facing

**Figure 2**

Stereo ribbon representations and close-up views of the structure surrounding the metal ion-binding sites in (a) B8FYU2\_DESHY, (b) Q9HJ63\_THEAC and (c) Q2LQ23\_SYNAS. (a) Stereo diagram of the structure surrounding one of the zinc/nickel-binding sites (top) of the B8FYU2\_DESHY dimer (bottom) and indicated by a rectangle. The metal ion-binding clefts on the dimer are indicated.

towards the extended stretch of polypeptide connecting the NTD and the CTD on the symmetry-related subunit.

3.4. Metal-ion binding in the CTD

The bound zinc on the zinc-finger domain of Q9HJ63\_THEAC shows a somewhat atypical coordination mode, with the side chains of Cys174, Cys177, Cys195 and Asp198 within ligation distance (Fig. 2b, Table 5). Typically, zinc ions in treble-clef zinc fingers are within coordination distance of Cys or His residues. Atypical coordination modes in which Asp or Glu act as ligands for the zinc have been observed previously in the zinc-finger domains of the mouse LIM-ldb1 LID complex (Deane *et al.*, 2004; PDB code 1rut), the human integrin-linked kinase ankyrin-repeat domain in complex with the PINCH1 LIM1 domain (Chiswell *et al.*, 2008; PDB code 3f6q), LIM domains 1 and 2 in complex with the LIM-interacting domain of LDB1 from mouse (Jeffries *et al.*, 2006; PDB code 2dfy) and the

heterodimeric core primase from *Sulfolobus solfataricus* (Lao-Sirieix *et al.*, 2005; PDB code 1zt2). Recently, the structure of a prokaryotic homolog of the transcriptional regulator of Ros from *Agrobacterium tumefaciens* was reported in which an Asp also replaces a Cys as a zinc ligand in the Cys<sub>2</sub>His<sub>2</sub> domain (Baglivo *et al.*, 2009). Q2LQ23\_SYNAS also has a single zinc-binding site on the zinc-finger domain, although here the zinc-chelating residues (Cys165, Cys168, Cys180 and Cys183; Fig. 2c, Table 5) are more typical.

3.5. Structural comparisons of the PF02663 proteins

Whereas 48 PF02663 proteins, including B8FYU2\_DESHY, are comprised of only a single NTD-like sequence motif, 98 others, including Q9HJ63\_THEAC and Q2LQ23\_SYNAS, also contain a C-terminal extension of ~40 amino acids with conserved cysteine and aspartic acid residues. The structures of Q9HJ63\_THEAC and Q2LQ23\_SYNAS show that these conserved residues form a zinc-

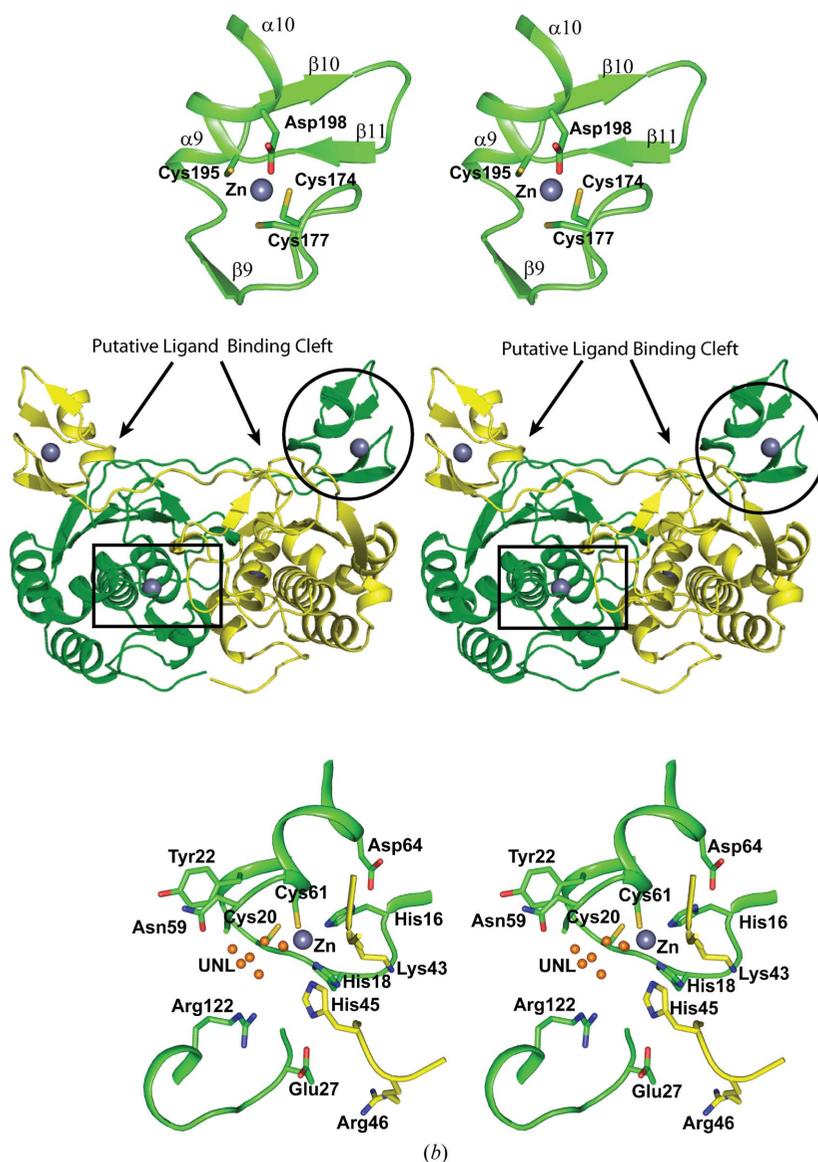


Figure 2 (continued) (b) Stereo diagram of one of the zinc-binding sites on the  $\alpha+\beta$  core domains (bottom), on the Q9HJ63\_THEAC dimer (middle) and on one of the zinc-finger domains (top). The sites on the NTD and CTD are indicated by a rectangle and a circle, respectively. An unidentified ligand (UNL) modeled at the putative active site on the  $\alpha+\beta$  core domain in the I222 crystal form is shown as orange spheres. A large putative binding cleft on the surface of the dimer is indicated.

binding site on a zinc-finger domain. Two other proposed domain architectures in the PF02263 family, for which structures have not yet been determined, include an NTD fused to a molybdopterine-binding domain (PF00994) and an NTD fused to a domain from the uncharacterized protein family UPF0066 (PF01980).

Pairwise structural comparisons of B8FYU2\_DESHY, Q9HJ63\_THEAC and Q2LQ23\_SYNAS (Fig. 3) revealed that the NTDs of B8FYU2\_DESHY and Q9HJ63\_THEAC are the most similar. The NTDs of B8FYU2\_DESHY and Q9HJ63\_THEAC (Fig. 3a) contain two conserved sequence motifs. The first motif, with a consensus sequence FHGHxC (Phe14–Cys19; B8FYU2\_DESHY numbering), contains three residues that coordinate the bound metal and is located on a loop connecting  $\alpha 1$  and  $\alpha 2$  (Figs. 1a and 1b). The second motif contains Asp58, Gln61 and Thr67 (B8FYU2\_DESHY numbering) and is located along the twofold-symmetry axis at the dimer interface.

The overall fold of the zinc-finger domains of Q9HJ63\_THEAC (residues 171–201) and Q2LQ23\_SYNAS (residues 162–190) are similar, with an r.m.s.d. of 1.1 Å for 24 superposed  $C^\alpha$  atoms. Two conserved Cys residues on the first  $\beta$ -loop of the CTD coordinate zinc (*i.e.* the zinc knuckle). These loops are located between  $\beta 8$  and  $\beta 9$  (Fig. 2b) and between  $\beta 6$  and  $\beta 7$  (Fig. 2c) in Q9HJ63\_THEAC and Q2LQ23\_SYNAS, respectively. The remaining zinc ligands (*i.e.* the two other Cys residues in Q2LQ23\_SYNAS and a Cys and an Asp in Q9HJ63\_THEAC) are located near the C-terminal  $\alpha$ -helix H10 (Figs. 2b and 2c).

### 3.6. Comparison with other structures

A *DALI* (Holm & Sander, 1995) search revealed that the NTD domain of Q9HJ63\_THEAC shows structural similarity to the intervening domain of 3-phosphoglycerate dehydrogenase from *Mycobacterium tuberculosis* (PDB code 3dc2; *DALI* Z score = 6.5, 7% sequence identity, 3.1 Å r.m.s.d. overlap of 96  $C^\alpha$  atoms; Dey *et al.*, 2008) and to a fragment from an iron–sulfur-dependent L-serine dehydratase from *Legionella pneumophila* (PDB code 2iqq; *DALI* Z score = 4.3, 7% sequence identity, 2.7 Å r.m.s.d. overlap of 78  $C^\alpha$  atoms). The low sequence identity between the NTD and the *DALI* hits suggests alternate functions for PF02263. In addition, four of the five strands in the  $\beta$ -sheet ( $\beta 1$ ,  $\beta 2$ ,  $\beta 3$  and  $\beta 4$  in Fig. 1a) and one of the  $\alpha$ -helices ( $\alpha 3$  in Fig. 1a) on the NTD are topologically equivalent to corresponding secondary-structure elements in the thioredoxin-like fold (Qi & Grishin, 2005; Martin, 1995). Therefore, the NTD can be classified as a type I circular permutation of the thioredoxin-like fold (Qi & Grishin, 2005), although thioredoxins are not reported to contain an equivalent metal ion-binding site, in contrast to the circularly permuted PF02263 NTD.

A *FATCAT* search of the PDB shows that the structure of the zinc-finger CTD on Q9HJ63\_THEAC is similar to the individual treble-clef zinc-finger subdomains of several eukaryotic LIM-like proteins (Gamsjaeger *et al.*, 2007; Krishna *et al.*, 2003). A similar search shows that the zinc-finger domain of Q2LQ23\_SYNAS is structurally similar to the phosphatidylinositol-3-phosphate-specific membrane-targeting

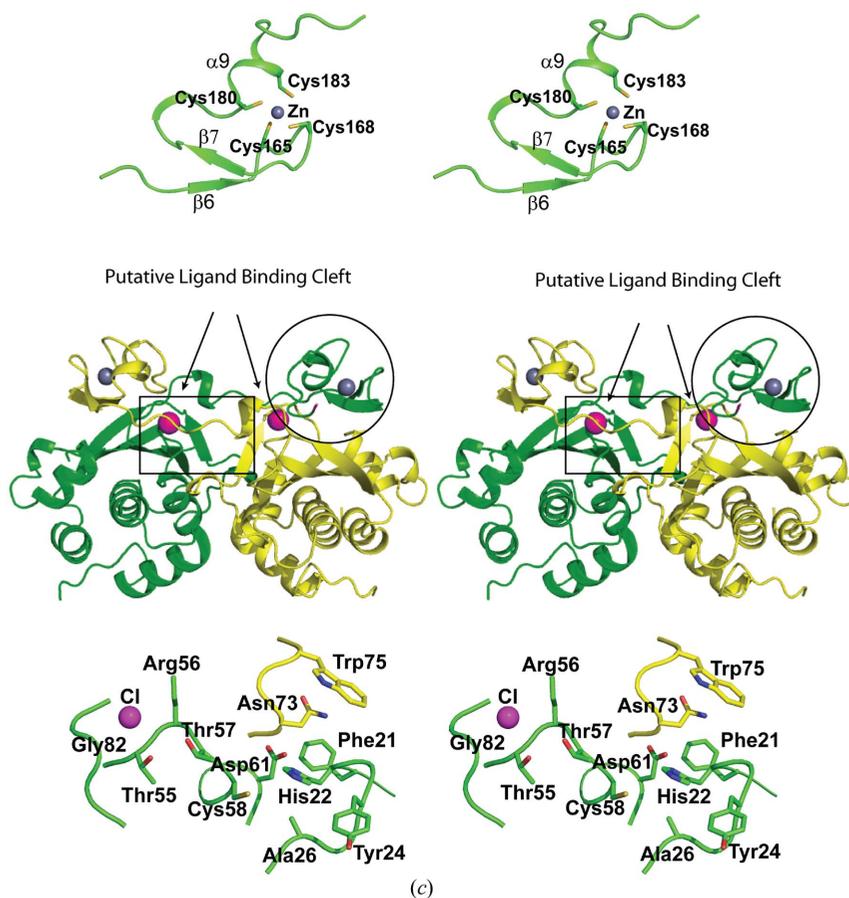


Figure 2 (continued)

(c) Stereo diagram of one of the putative active-site clefts (bottom; indicated by a rectangle), the Q2LQ23\_SYNAS dimer (middle) and one of the zinc-finger domains (top; indicated by a circle). The O, N, and S atoms on the side chains are shown in red, blue and yellow, respectively. Bound metal atoms and chloride anions are shown as gray and magenta spheres, respectively. A large putative binding cleft on the surface of the dimer is indicated.

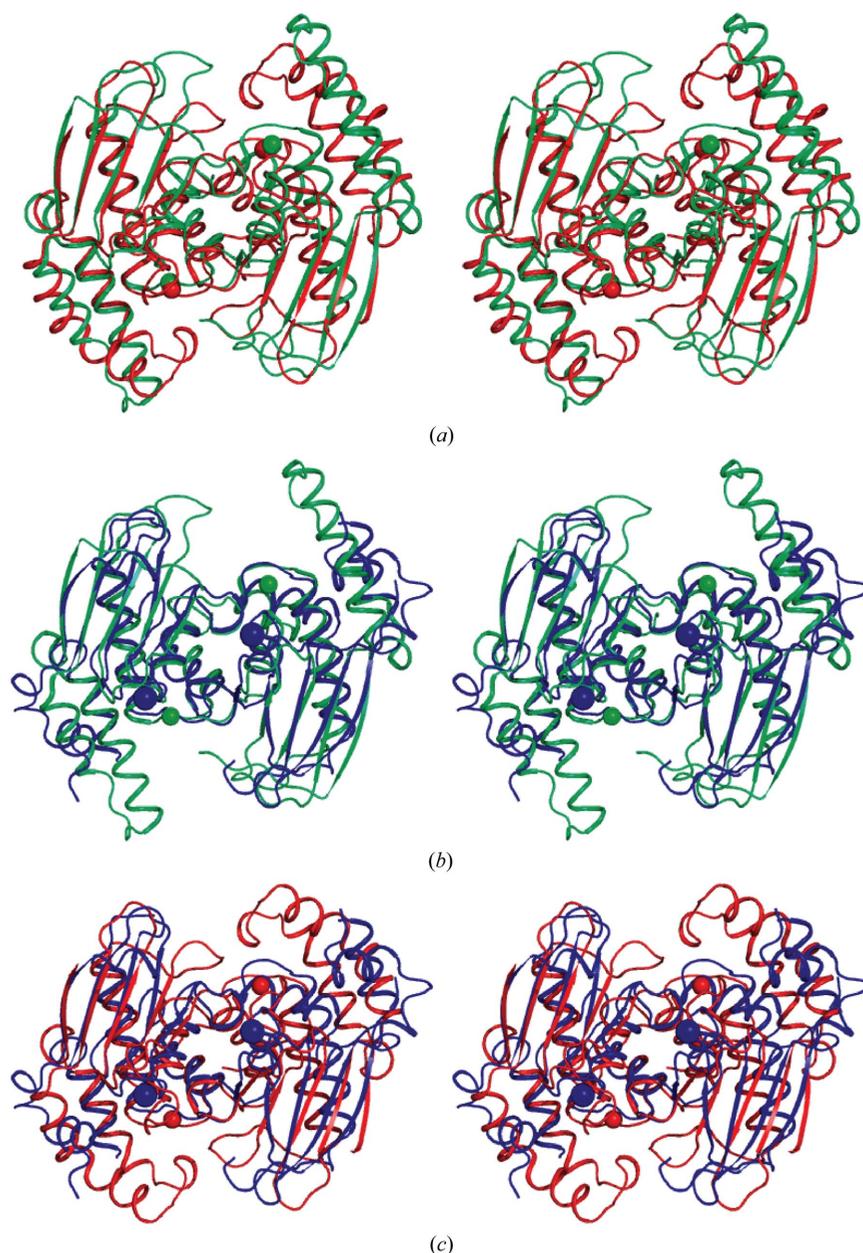
binding FYVE domain of vps27p from *Saccharomyces cerevisiae* (Misra & Hurley, 1999; PDB code 1vfy).

### 3.7. Functional implications

The identification of a treble-clef, zinc-finger domain on Q9HJ63\_THEAC and Q2LQ23\_SYNAS indicates that some PF02663 family members may be involved in transcriptional regulation or protein-protein interactions. However, since the range of functions performed by zinc fingers is diverse, a more detailed functional annotation remains a challenge at present. It has been suggested that a PF02663 homolog in *Methanosarcina barkeri* could be a chaperone

(Vorholt *et al.*, 1996). Chaperone activity has also been proposed based on the structure of thioredoxin-2 from the photosynthetic bacterium *Rhodobacter capsulatus* (Ye *et al.*, 2007; PDB code 2ppt). However, in contrast to the structures of the three PF02663 proteins described here, the zinc-finger domain is at the N-terminal end of the protein and the motif for the zinc finger in thioredoxin-2 is a zinc ribbon distinct from the treble-clef motif in the PF02663 structures.

Previous investigations have established that in some organisms *fmdE* is co-transcribed with genes encoding the catalytic subunits of a key methanogenic enzyme. Genome-context analysis indicates that only a handful (13 of 208) of genes corresponding to PF02663 members are adjacent to and likely to be co-transcribed with genes



**Figure 3**

Pairwise comparison of the  $\alpha+\beta$  core-domain structures of three PF02663 homologs. (a) Stereo diagram showing the superposition of the ribbon traces for (a) B8FYU2\_DESHY (PDB code 2glz; green) and Q9HJ63\_THEAC (PDB code 2gvi; red). The Zn/Ni atoms in B8FYU2\_DESHY are shown as green spheres and the Zn atoms from Q9HJ63\_THEAC are shown as red spheres. (b) Stereo diagram showing the superposition of the ribbon traces for B8FYU2\_DESHY (PDB code 2glz; green) and Q2LQ23\_SYNAS (PDB code 3d00; blue). The Zn/Ni atoms in B8FYU2\_DESHY are shown as green spheres and the chloride ions from Q2LQ23\_SYNAS are shown as blue spheres. (c) Stereo diagram showing the superposition of the ribbon traces for Q9HJ63\_THEAC (PDB code 2gvi; red) and Q2LQ23\_SYNAS (PDB code 3d00; blue). The Zn/Ni atoms in Q9HJ63\_THEAC are shown as red spheres and the chloride ions from Q2LQ23\_SYNAS are shown as blue spheres.

encoding the catalytic subunits of molybdenum formylmethanofuran dehydrogenase. Sequence analyses, combined with the structure determinations described here, indicate that 12 of these genes are likely to be part of an *fmd* operon with a two-domain NTD + zinc-finger architecture, whereas an *fmdE* homolog from *M. barkeri* has a one-domain NTD-like architecture. However, most of the genes encoding PF02663 homologs, irrespective of domain architecture, are adjacent to genes encoding metal-ion transporters. These results indicate the absence of a strict correlation between domain architecture and gene context; nevertheless, the results do suggest a possible involvement in metal-ion transport.

#### 4. Conclusions

The structures of three members of PF02663 enhance our understanding of the role of these proteins in microbes. Individual proteins within this family display differences in domain architectures, metal-ion binding propensities and dimer interactions. These structural differences suggest a broad range of potential functions for this group of proteins. The identification of a C-terminal zinc-finger domain in two of the structures suggests one possible role for this class of proteins as transcriptional regulators. The NTD together with the CTD might serve as part of the nucleic acid binding surface and/or serve as a signal-sensing domain for the binding of unknown effectors. The absence of a zinc-finger domain in some PF02663 homologs, such as B8FYU2\_DESHY, provides some evidence for involvement in alternate processes. Further biochemical and biophysical studies should yield valuable insights into the relationship between structure and function for this interesting group of proteins.

Additional information about the proteins described in this study is available from TOPSAN (Krishna *et al.*, 2010) at <http://www.topsan.org/explore?PDBid=2glz> for B8FYU2\_DESHY, <http://www.topsan.org/explore?PDBid=2gvi> for Q9HJ63\_THEAC and <http://www.topsan.org/explore?PDBid=3d00> for Q2LQ23\_SYNAS.

This work was supported by the NIH, National Institute of General Medical Sciences, Protein Structure Initiative grant U54 GM074898. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL). The SSRL is a national user facility operated by Stanford University at the SLAC National Accelerator Laboratory on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences). *D. halfniense* DCB-2 was a gift from Drs Tamara Cole and Jim Tiedje, Michigan State University, East Lansing, Michigan, USA. Genomic DNA from *T. acidophilum* DSM1728 (ATCC No. 25905D) was obtained from the American Type Culture Collection (ATCC). *S. aciditrophicus* SB was a gift from Professor Michael J. McInerney, University of Oklahoma, Norman, Oklahoma, USA. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

#### References

Baglivo, I., Russo, L., Esposito, S., Malgieri, G., Renda, M., Salluzzo, A., Di Blasio, B., Isernia, C., Fattorusso, R. & Pedone, P. V. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 6933–6938.

- Bertram, P. A., Schmitz, R. A., Linder, D. & Thauer, R. K. (1994). *Arch. Microbiol.* **161**, 220–228.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). *Nucleic Acids Res.* **31**, 3497–3500.
- Chiswell, B. P., Zhang, R., Murphy, J. W., Boggon, T. J. & Calderwood, D. A. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 20677–20682.
- Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.
- Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J. & Perrakis, A. (2004). *Acta Cryst.* **D60**, 2222–2229.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cruikshank, D. W. J. (1999). *Acta Cryst.* **D55**, 583–601.
- Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). *Nucleic Acids Res.* **32**, W615–W619.
- Deane, J. E., Ryan, D. P., Sunde, M., Maher, M. J., Guss, J. M., Visvader, J. E. & Matthews, J. M. (2004). *EMBO J.* **23**, 3589–3598.
- Dey, S., Burton, R. L., Grant, G. A. & Sacchettini, J. C. (2008). *Biochemistry*, **47**, 8271–8282.
- Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
- Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. (2006). *Nucleic Acids Res.* **34**, W116–W118.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Finn, R. D., Tate, J., Mistry, J., Coghill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L. & Bateman, A. (2008). *Nucleic Acids Res.* **36**, D281–D288.
- Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M. & Mackay, J. P. (2007). *Trends Biochem. Sci.* **32**, 63–70.
- Grishin, N. V. (2001). *Nucleic Acids Res.* **29**, 1703–1714.
- Hallam, S. J., Putnam, N., Preston, C. M., Detter, J. C., Rokhsar, D., Richardson, P. M. & DeLong, E. F. (2004). *Science*, **305**, 1457–1462.
- Hochheimer, A., Hedderich, R. & Thauer, R. K. (1998). *Arch. Microbiol.* **170**, 389–393.
- Hochheimer, A., Linder, D., Thauer, R. K. & Hedderich, R. (1996). *Eur. J. Biochem.* **242**, 156–162.
- Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
- Jeffries, C. M., Graham, S. C., Stokes, P. H., Collyer, C. A., Guss, J. M. & Matthews, J. M. (2006). *Protein Sci.* **15**, 2612–2618.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Kabsch, W. (2010a). *Acta Cryst.* **D66**, 125–132.
- Kabsch, W. (2010b). *Acta Cryst.* **D66**, 133–144.
- Kleywegt, G. J., Zou, J.-Y., Kjeldgaard, M. & Jones, T. A. (2001). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 353–356. Dordrecht: Kluwer Academic Publishers.
- Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. (2008). *Proteins*, **71**, 982–994.
- Krishna, S. S., Majumdar, I. & Grishin, N. V. (2003). *Nucleic Acids Res.* **31**, 532–550.
- Krishna, S. S., Weekes, D., Bakolitsa, C., Elsliger, M.-A., Wilson, I. A., Godzik, A. & Wooley, J. (2010). *Acta Cryst.* **F66**, 1143–1147.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Lao-Sirieix, S. H., Nookala, R. K., Roversi, P., Bell, S. D. & Pellegrini, L. (2005). *Nature Struct. Mol. Biol.* **12**, 1137–1144.
- Laskowski, R. A. (2009). *Nucleic Acids Res.* **37**, D355–D359.
- Lesley, S. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11664–11669.
- Leslie, A. G. W. (1992). *Int CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Liu, Y. & Whitman, W. B. (2008). *Ann. NY Acad. Sci.* **1125**, 171–189.
- Martin, J. L. (1995). *Structure*, **3**, 245–250.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J. (2007). *Acta Cryst.* **D63**, 32–41.
- Misra, S. & Hurlley, J. H. (1999). *Cell*, **97**, 657–666.
- Qi, Y. & Grishin, N. V. (2005). *Proteins*, **58**, 376–388.
- Santarsiero, B. D., Yegian, D. T., Lee, C. C., Spraggon, G., Gu, J., Scheibe, D., Uber, D. C., Cornell, E. W., Nordmeyer, R. A., Kolbe, W. F., Jin, J., Jones, A. L., Jaklevic, J. M., Schultz, P. G. & Stevens, R. C. (2002). *J. Appl. Cryst.* **35**, 278–281.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1937–1940.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* **D55**, 849–861.
- Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W. & Hedderich, R. (2008). *Nature Rev. Microbiol.* **6**, 579–591.

- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* **D55**, 191–205.
- Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
- Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
- Vorholt, J. A., Vaupel, M. & Thauer, R. K. (1996). *Eur. J. Biochem.* **236**, 309–317.
- Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
- Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
- Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
- Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst.* **D60**, 1833–1839.
- Ye, J., Cho, S. H., Fuselier, J., Li, W., Beckwith, J. & Rapoport, T. A. (2007). *J. Biol. Chem.* **282**, 34945–34951.