Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structure of the gas vesicle protein GvpF from the cyanobacterium *Microcystis aeruginosa*

Bo-Ying Xu, Ya-Nan Dai,* Kang Zhou, Yun-Tao Liu, Qianqian Sun, Yan-Min Ren, Yuxing Chen and Cong-Zhao Zhou*

Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Heifei, Anhui 230027, People's Republic of China

Correspondence e-mail: daiyanan@mail.ustc.edu.cn, zcz@ustc.edu.cn

Received 27 July 2014
Accepted 25 September 2014

PDB reference: GvpF, 4qsg

Gas vesicles are gas-filled proteinaceous organelles that provide buoyancy for bacteria and archaea. A gene cluster that is highly conserved in various species encodes about 8–14 proteins (Gvp proteins) that are involved in the formation of gas vesicles. Here, the first crystal structure of the gas vesicle protein GvpF from *Microcystis aeruginosa* PCC 7806 is reported at 2.7 Å resolution. GvpF is composed of two structurally distinct domains (the N-domain and C-domain), both of which display an $\alpha+\beta$ class overall structure. The N-domain adopts a novel fold, whereas the C-domain has a modified ferredoxin fold with an apparent variation owing to an extension region consisting of three sequential helices. The two domains pack against each other *via* interactions with a C-terminal tail that is conserved among cyanobacteria. Taken together, it is concluded that the overall architecture of GvpF presents a novel fold. Moreover, it is shown that GvpF is most likely to be a structural protein that is localized at the gas-facing surface of the gas vesicle by immunoblotting and immunogold labelling-based tomography.

1. Introduction

The gas vesicle containing a gas-filled space is one of the most extraordinary organelles and subcellular structures found in prokaryotic organisms. Over 150 species of prokaryotes, including at least five phyla of bacteria and two phyla of archaea, contain gas vesicles (Walsby, 1994). This unique organelle is best known to occur in aquatic microorganisms such as aquatic anoxyphototrophic bacteria, cyanobacteria and halophilic archaea. Owing to the fact that gas vesicles are gas-filled subcellular structures, they can increase the buoyancy of aquatic bacteria by lowering the density of cells. Bacteria can thus float towards air–liquid interfaces in aqueous environments, enabling their positioning in optimal light and oxygen conditions for growth and subsequent niche colonization (Walsby, 1972, 1994).

Gas vesicles are rigid, hollow and light-refractile proteinaceous structures (Walsby & Buckland, 1969; Jones & Jost, 1971; Walsby, 1994). Mature gas vesicles take the shape of a cylindrical or spindle shell with conical end-caps with general dimensions of 45–200 nm in width and 100–2000 nm in length (Bowen & Jensen, 1965; Jost, 1965, Walsby, 1994). The diameters of gas vesicles are species-specific, whereas their lengths vary greatly even in a single cell. Gas vesicles are formed by a single wall layer only 2 nm thick (Walsby, 1994). The inner surface of the gas vesicle wall is hydrophobic, whereas the outer surface is more hydrophilic. The gas vesicle wall is freely permeable to gases but is impermeable to liquid water (Walsby, 1969, 1982). Gas vesicles irreversibly collapse upon exposure to pressure that exceeds a critical value, which

varies from 0.09 MPa to greater than 1 MPa in different organisms and correlates inversely with the width of the gas vesicle (Hayes & Walsby, 1986; Walsby, 1994; Dunton & Walsby, 2005).

Gas vesicle formation involves 8–14 different Gvp proteins, which are encoded by a *gvp* gene cluster that has been found in a variety of bacteria and archaea (Pfeifer, 2012). For instance, the cyanobacterium *Microcystis aeruginosa* PCC 7806, which is often responsible for the seasonal algal bloom in polluted freshwater environments, has 12 *gvp* genes designated *gvpA*, *gvpC*, *gvpN*, *gvpJ*, *gvpX*, *gvpK*, *gvpF*, *gvpG*, *gvpV* and *gvpW* involved in gas vesicle synthesis. Ten of these *gvp* genes are organized in two operons, $gvpA_IA_{II}A_{III}CNJX$ and *gvpKFG*, whereas *gvpV* and *gvpW* are individually expressed (Mlouka *et al.*, 2004; Dunton & Walsby, 2005). Gas vesicle walls are exclusively made up of proteins (Walsby & Buckland, 1969; Jones & Jost, 1970; Walsby, 1994), the dominant constituent of which is a 7–8 kDa hydrophobic protein called GvpA (Walsby, 1994). It forms 'ribs' (with an inter-rib distance of 4.6 nm) that run nearly perpendicular to the long axis of the vesicle (Hayes *et al.*, 1986; Buchholz *et al.*, 1993; Walsby, 1994). In addition, a second dominant structural component, GvpC, is located on the outer surface of the gas vesicle walls and strengthens the resistance of the gas vesicles to pressure (Hayes *et al.*, 1988; Walsby & Hayes, 1988; Englert & Pfeifer, 1993; Walsby, 1994; Dunton *et al.*, 2006). Notably, another conserved protein, GvpF, has been shown to be a structural component of gas vesicles from *Halobacterium* sp. strain NRC-1 (Shukla & DasSarma, 2004). However, the other structural components besides GvpA and GvpC in cyanobacterial gas vesicles remain uncharacterized. Moreover, no three-dimensional structure of a Gvp protein from a gas vesicle-forming microbe has been reported to date.

In this study, we solved the crystal structure of GvpF from *M. aeruginosa* PCC 7806 at 2.7 Å resolution, representing the first structure of a Gvp protein. The overall structure of GvpF displays a monomer with two distinct domains (N-domain and C-domain) and a C-terminal tail (C-tail) which bridges the two domains. Both domains can be assigned to the $\alpha+\beta$ class but adopt different folds according to the criteria of the Structural Classification of Proteins (SCOP) database. Structural analysis revealed that the N-domain represents a novel fold and the C-domain consists of a modified ferredoxin fold. Moreover, immunoblotting analysis and immunogold labelling of purified gas vesicles and collapsed fragments confirmed that GvpF is a structural component of the gas vesicle and is most likely to be localized on the gas-facing surface of the wall.

2. Materials and methods

2.1. Cloning, overexpression and purification of GvpF

The coding region for GvpF (735 base pairs) was amplified from the genomic DNA of *M. aeruginosa* PCC 7806, cloned into a pET-28a-derived vector with an N-terminal 6×His tag and overexpressed in *Escherichia coli* strain BL21 (DE3) harbouring the pKY206 plasmid (Novagen) using 2×YT culture medium (5 g NaCl, 16 g Bacto tryptone and 10 g yeast extract per litre) containing 30 μg ml^{-1} kanamycin and 5 μg ml^{-1} tetracycline. The cells were grown at 37°C to an $A_{600\,nm}$ of 0.8. Expression of the recombinant protein was induced with 0.2 m*M* isopropyl β-D-1-thiogalactopyranoside for a further 4 h at 37°C before harvesting. The cells were collected by centrifugation at 4000*g* for 10 min and resuspended in 40 ml lysis buffer (200 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0). After 30 min of sonication and centrifugation at 12 000*g* for 30 min, the supernatant containing the soluble target protein was collected and loaded onto a nickel–nitrilotriacetic acid column (GE Healthcare) equilibrated with binding buffer (200 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0). The target protein was eluted with 300 m*M* imidazole, 200 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0 and loaded onto a HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with 150 m*M* NaCl, 14 m*M* β-mercaptoethanol, 20 m*M* Tris–HCl pH 7.0. Fractions containing the target protein were pooled and concentrated to 10 mg ml^{-1} for crystallization. The protein purity was assessed by SDS–PAGE and the protein sample was stored at −80°C.

2.2. Incorporation of selenomethionine into GvpF

Selenomethionine (SeMet)-labelled GvpF (Se-GvpF) was obtained by overexpressing *gvpF* in *E. coli* strain B834 (DE3) (Novagen) harbouring the pKY206 plasmid. A culture of transformed cells was inoculated into LB medium and incubated at 37°C. The cells were harvested when the $A_{600\,nm}$ reached 0.2 and were then washed twice with M9 medium. The cells were then cultured in SeMet medium (M9 medium with 25 mg l^{-1} L-SeMet and the other essential amino acids at 50 mg l^{-1}) to an $A_{600\,nm}$ of 0.6–0.8. The remaining steps of protein expression, purification and storage of Se-GvpF were the same as those for native GvpF.

2.3. Crystallization, data collection and processing

Screening for the crystallization conditions of GvpF was performed using the Crystal Screen, Crystal Screen 2, Index, Grid Screens and SaltRx kits (Hampton Research) with the hanging-drop vapour-diffusion method in 96-well plates at 16°C. The native crystals of GvpF were grown by mixing 1 μl 10 mg ml^{-1} protein sample with 1 μl reservoir solution [10%(*w*/*v*) polyethylene glycol 20 000, 0.1 *M* sodium citrate tribasic pH 5.6] and equilibrating against 0.5 ml reservoir solution. Crystals of Se-GvpF were obtained using the same method with the same reservoir as was used to obtain the native crystals. Typically, crystals appeared in 2–3 d and reached maximum size in one week. The crystals were transferred to cryoprotectant (reservoir solution supplemented with 60% sucrose) and flash-cooled in liquid nitrogen. Both native and SeMet-derivative data sets were collected from single crystals at 100 K in a liquid-nitrogen stream using an ADSC Q315r CCD detector (MAR Research, Germany) on beamline 17U at the Shanghai Synchrotron Radiation Facility (SSRF). All diffraction data were indexed, integrated and scaled with *iMosflm* (Battye *et al.*, 2011).

2.4. Structure determination and refinement

The crystal structure of GvpF was determined by the single-wavelength anomalous dispersion phasing method using data from a single SeMet-substituted protein crystal to a maximum resolution of 2.7 Å. *AutoSol* from *PHENIX* (Adams *et al.*, 2010) was used to locate the heavy atoms. The initial model was built automatically with *AutoBuild* in *PHENIX*. Refinement was carried out using the maximum-likelihood method as implemented in *REFMAC*5 (Murshudov *et al.*, 2011) as part of the *CCP*4 suite (Winn *et al.*, 2011) and rebuilding was carried out interactively using *Coot* (Emsley *et al.*, 2010). The final model was evaluated with *MolProbity* (Chen *et al.*, 2010) and *PROCHECK* (Laskowski *et al.*, 1993). Crystallographic parameters and data-collection statistics are listed in Table 1. All figures showing structures were prepared with *PyMOL* (DeLano, 2002).

2.5. *Microcystis* cultures

M. aeruginosa strain PCC 7806, obtained from the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB Collection), Wuhan, People's Republic of China, was grown in 1 l Erlenmeyer flasks half-filled with BG11 medium (Rippka *et al.*, 1979) at 30°C under constant illumination with cool-white fluorescent light of photon irradiance 30 µmol m^{-2} s^{-1}. The flasks were agitated on a rotary shaker (200 rev min^{-1}). The algae had a generation time of 14–18 h and were grown to a final density of $A_{750\ nm}$ = 0.66–1.0.

2.6. Gas vesicle isolation

Cell-free preparations of *M. aeruginosa* were made by penicillin treatment and osmotic lysis of the cells as described previously (Jones & Jost, 1970; Weathers *et al.*, 1977). Cells were collected by floatation and then concentrated by reverse osmosis against 1 *M* glycerol. The *Microcystis* cells were lysed with lysozyme (1 mg ml^{-1}; Sigma) in 5 m*M* sodium cyanide overnight. Gas vesicle fractions were enriched by collecting (with a Pasteur pipette) the white buoyant layer that accumulated at the top of cultures in which cells lysed spontaneously. The suspension containing gas vesicles was left standing until a new and more concentrated floating layer had accumulated, which was then collected as described above. Gas vesicles were then purified by repeated centrifugal floatation in 10 m*M* Tris–HCl pH 7.5, 5 m*M* sodium cyanide six or seven times. The purified gas vesicles were stored in 6.3 m*M* ammonium bicarbonate containing 5 m*M* sodium cyanide to prevent bacterial contamination and degradation (Powell *et al.*, 1991). The gas vesicle protein concentration was estimated by measuring the pressure-sensitive optical density (PSOD) at an absorbance wavelength of 500 nm as described previously, where the gas vesicle protein at 1 mg ml^{-1} has a PSOD of 20.8 (Walsby & Armstrong, 1979; Walsby, 1994; Belenky *et al.*, 2004; Dunton & Walsby, 2005).

Table 1
Crystal parameters, data collection and structure refinement of Se-GvpF.

Values in parentheses are for the highest resolution bin.

Data collection	
Wavelength (Å)	0.97915
Space group	$P3_221$
Unit-cell parameters	
$a = b$ (Å)	81.92
c (Å)	87.60
$\alpha = \beta$ (°)	90.00
γ (°)	120.00
Resolution range (Å)	50.00–2.70 (2.85–2.70)
Wilson B factor (Å^2)	43.73
Unique reflections	9485 (1376)
Completeness (%)	98.2 (98.1)
Anomalous multiplicity	4.4 (4.5)
Anomalous completeness (%)	98.8 (98.7)
$\langle I/\sigma(I)\rangle$	14.8 (4.5)
R_{merge}† (%)	11.2 (49.6)
Average multiplicity	8.7 (9.0)
Phasing statistics	
Correlation coefficient	0.68
Figure of merit	0.289
Structure refinement	
Resolution range (Å)	50.00–2.70
R factor‡/R_{free}§ (%)	23.1/27.3
No. of protein atoms	1982
No. of water atoms	17
R.m.s.d.¶, bond lengths (Å)	0.005
R.m.s.d., bond angles (°)	0.946
Mean B factor (Å^2)	45.3
Ramachandran plot††	
Poor rotamers (%)	3.24
Most favoured (%)	97.08
Additional allowed (%)	2.92
Outliers (%)	0
Clashscore (per 1000 atoms)	2.02
PDB code	4qsg

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an observation and $\langle I(hkl)\rangle$ is the mean value for its unique reflection; summations are over all reflections. ‡ R factor = $\sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. § R_{free} was calculated with 5% of the data, which were excluded from refinement. ¶ Root-mean-square deviation from ideal values. †† Categories were defined by *MolProbity*.

2.7. Electron microscopy

Carbon-coated copper grids (400-mesh) were immersed in the purified gas vesicles for 1 min and excess liquid was removed with filter paper. 2% uranyl acetate pH 4.0 was used to stain the gas vesicles. The gas vesicles were examined with a Tecnai G^2 F20 transmission electron microscope (FEI, USA) running at 200 kV voltage and 750 000× magnification. Images were taken using a CCD camera attached to the microscope. Gas vesicle dimensions were calculated with the *FEI Image* software.

2.8. Electrophoresis and immunoblotting analysis

Immunoblotting analysis of cell lysates and purified gas vesicles were performed as described previously (Shukla & DasSarma, 2004). Typically, cell lysates containing 208 µg protein or 176 µg purified gas vesicles were mixed with a final concentration of 2.5%(*w*/*v*) sodium dodecyl sulfate (SDS) and an equal volume of 2× sample-loading buffer (60 m*M* Tris–HCl pH 6.8, 20% glycerol, 0.2 *M* dithiothreitol, 5% SDS, 0.02% bromophenol blue) and boiled for 10 min. After

cooling to room temperature, the samples were briefly vortexed and electrophoresed on 15% SDS–PAGE for 100 min at 120 V at room temperature. The electrophoretically resolved protein bands were electroblotted onto a 0.45 µm pore Immobilon-NC nitrocellulose membrane (Millipore, Boston, Massachusetts, USA) for 2 h at 250 mA. After electroblotting, the membranes were blocked for 1 h with 5% skimmed milk in TBST buffer (150 m*M* NaCl, 10 m*M* Tris–HCl pH 7.5, 0.1% Tween-20) at 4°C overnight and washed six times for 5 min each with TBST buffer. The washed membrane was incubated for 1 h at room temperature under gentle shaking with anti-GvpF polyclonal antibodies diluted 1:3000. The membrane was then washed six times for 5 min each with TBST buffer and incubated with goat anti-rabbit secondary antibodies (Promega) labelled with alkaline phosphatase diluted (1:3000) in TBST buffer. For detection of the protein bands, the membrane was incubated in Thermo Lumi-Phos WB chemiluminescent substrate and detected using a Luminescent Image Analyzer ImageQuant LAS 4000 mini system (GE Healthcare Life Sciences).

2.9. Immunogold labelling of gas vesicles and automated tomography

Immunogold labelling of *Microcystis* gas vesicles was performed as described previously (Hayat, 1989; Buchholz *et al.*, 1993; Wyffels, 2001). Drops of purified intact or collapsed gas vesicles were placed on the carbon-coated copper grids for 15 min and excess liquid was removed with filter paper. The grids were incubated face-down for 1 h on drops of anti-GvpF polyclonal antisera diluted 100-fold. They were then drained with filter paper, washed six times with drops of phosphate-buffered saline (1×PBS; 10 m*M* Na_2HPO_4, 1.8 m*M* KH_2PO_4, 137 m*M* NaCl, 2.7 m*M* KCl pH 7.4) and floated for 1 h on a drop of 20-fold diluted 10 nm gold spheres conjugated to goat anti-rabbit secondary antibodies (Boster). The grids were drained and washed as described above with two final washes with drops of distilled water. The samples were finally negatively stained with 0.75% uranyl formate. Antibodies were diluted in PBS containing 1%(*w*/*v*) bovine serum albumin (Sigma). Collapsed gas vesicles were obtained by sonication for 2 min with an ultrasonic cleaner (75 W, 25 kHz). The samples were then also observed under the Tecnai G^2 F20 transmission electron microscope. Automated tomography was performed on one segment of immunogold-labelled collapsed gas vesicles. Tilt series were collected from −60 to 60° using the transmission electron microscope with a total dose rate of about 100 e Å^{-2}. Image stacks were aligned and reconstructed using *IMOD* (Kremer *et al.*, 1996;

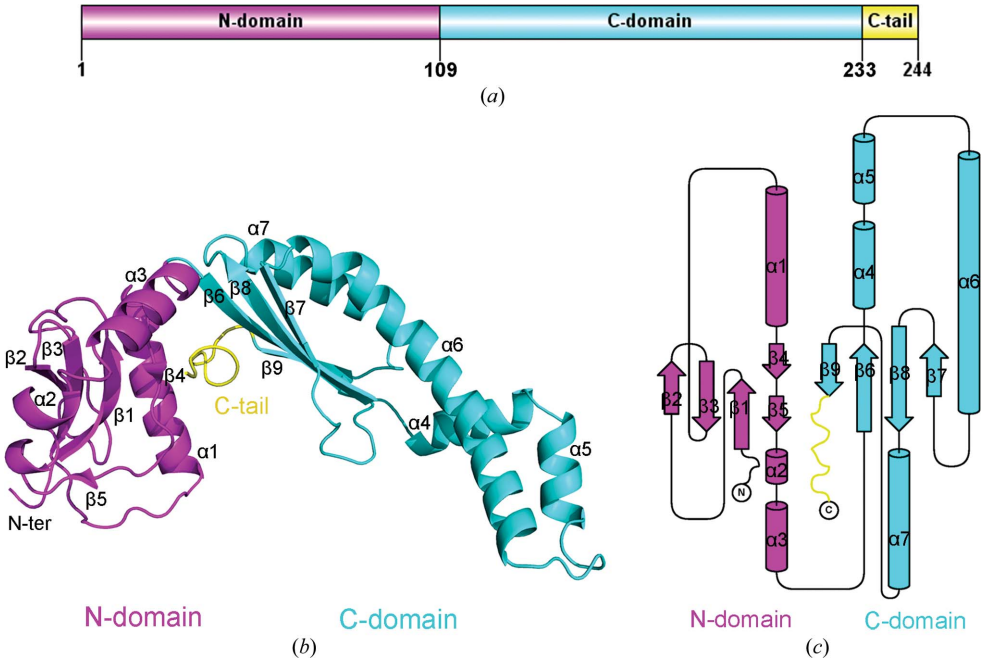


Figure 1
Overall structure of GvpF. (*a*) Organization of GvpF. Three segments of GvpF drawn by *Domain Graph* v.1.0 (Ren *et al.*, 2009). The N-domain, C-domain and C-tail are shown in magenta, cyan and yellow, respectively. (*b*) A schematic representation of GvpF. The secondary-structural elements are labelled sequentially. (*c*) A topology diagram of GvpF.

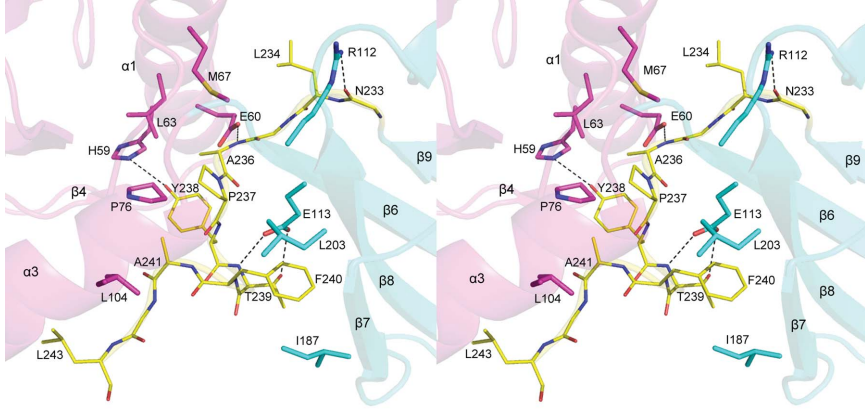


Figure 2
Stereographic representation of the interactions between the N/C-domain and the C-tail. The residues from the C-tail are shown as thin sticks in yellow and the residues from the N- and C-domains are shown as bold sticks in magenta and cyan, respectively. Black dotted lines denote polar interactions. The backbone of the protein is shown as a semitransparent cartoon.

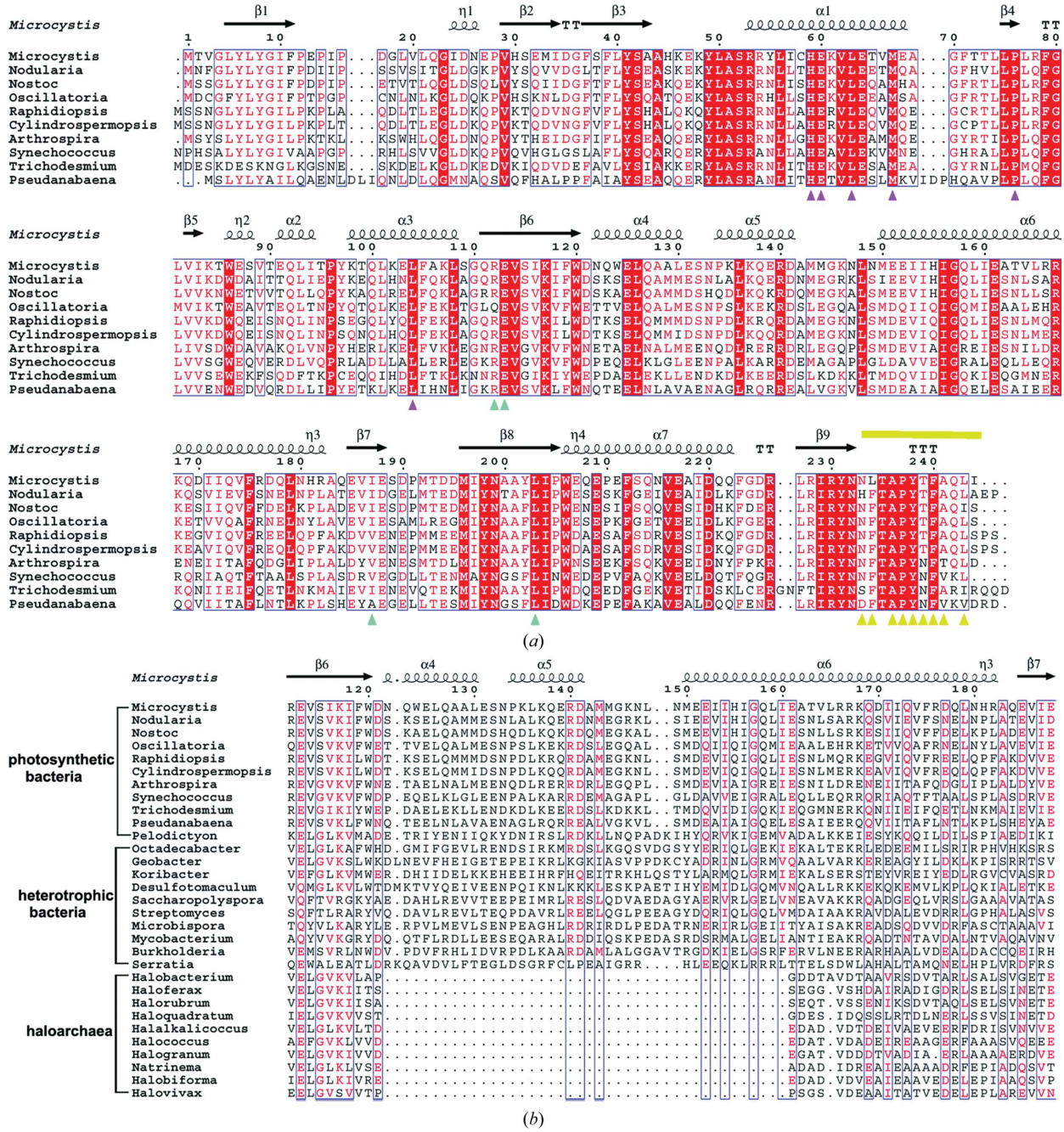


Figure 3
Multiple sequence alignment. (*a*) Sequence alignment of cyanobacterial GvpF proteins. The C-tail is labelled as a yellow line and residues involved in interactions between the N/C-domain and the C-tail are labelled. Magenta triangles indicate residues from the N-domain, cyan triangles indicate residues from the C-domain and yellow triangles refer to residues from the C-tail. (*b*) Sequence alignment of the extension region (helices α4, α5 and a segment of α6) of GvpF from gas vesicle-forming microbes (labelled as photosynthetic bacteria, heterotrophic bacteria and haloarchaea, respectively). The multiple sequence alignment was performed using *MultAlin* (http://multalin.toulouse.inra.fr/multalin/multalin.html) and *ESPript* 3.0 (http://espript.ibcp.fr/ESPript/cgibin/ESPript.cgi). All sequences were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov). The GvpF sequences are (NCBI accession Nos. are given in parentheses) from *Microcystis aeruginosa* PCC 7806 (WP_002747917.1), *Nodularia spumigena* (WP_006197964.1), *Nostoc* sp. PCC 7120 (NP_486288.1), *Oscillatoria* sp. PCC 6506 (WP_007355102.1), *Raphidiopsis brookii* D9 (WP_009342595.1), *Cylindrospermopsis raciborskii* (WP_006276398.1), *Arthrospira* sp. PCC 8005 (WP_006626108.1), *Synechococcus* sp. JA-2-3B′a 2-13 (YP_477775.1), *Trichodesmium erythraeum* IMS101 (YP_722032.1), *Pseudanabaena iceps* PCC 7429 (WP_009625330.1), *Pelodictyon phaeoclathratiforme* BU-1 (YP_002018656.1), *Octadecabacter arcticus* 238 (YP_007702338.1), *Geobacter uraniireducens* Rf4 (YP_001231580.1), *Koribacter versatilis* Ellin345 (YP_591476.1), *Desulfotomaculum acetoxidans* DSM 771 (YP_003192403.1), *Saccharopolyspora erythraea* NRRL 2338 (YP_001103230.1), *Streptomyces himastatinicus* ATCC 53653 (WP_009712796.1), *Microbispora* sp. ATCC PTA-5024 (ETK35504.1), *Mycobacterium* sp. MCS (YP_639499.1), *Burkholderia oklahomensis* (WP_010109833.1), *Serratia* sp. ATCC 39006 (WP_021013988.1), *Halobacterium* sp. NRC-1 (NP_395757.1), *Haloferax mediterranei* ATCC 33500 (YP_006349392.1), *Halorubrum vacuolatum* (CAA69884.1), *Haloquadratum walsbyi* DSM 16790 (YP_657542.1), *Halalkalicoccus jeotgali* B3 (YP_003737528.1), *Halococcus thailandensis* (WP_007739251.1), *Halogranum salarium* (WP_009366043.1), *Natrinema altunense* (WP_007109510.1), *Halobiforma lacisalsi* (WP_007143587.1) and *Halovivax ruber* XH-70 (YP_007284645.1). The secondary-structural elements of GvpF are displayed above the alignment. Highly conserved residues are coloured red.

Mastronarde, 1997) and the final pixel size of the tomogram is 7.55 Å per pixel. The tomograms were then visualized and segmented with *Amira* (FEI, USA).

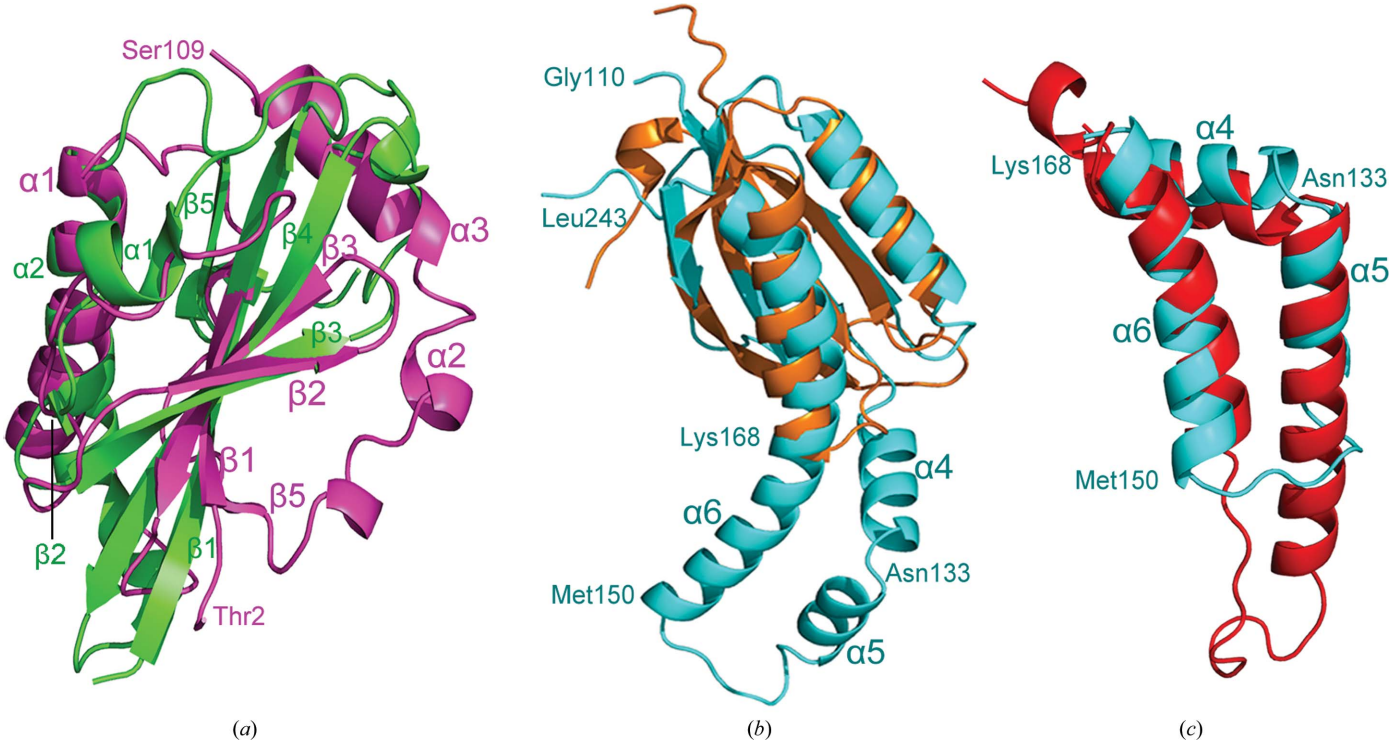
3. Results and discussion

3.1. Overall structure of GvpF

The structure of GvpF was solved and refined to 2.7 Å resolution. The crystals belonged to space group $P3_221$, with one molecule in the asymmetric unit. The overall dimensions of the molecule are approximately 72 × 34 × 35 Å. GvpF consists of two structurally distinct domains (N-domain and C-domain) and a tail at the C-terminus (C-tail) (Figs. 1*a* and 1*b*). The secondary-structural elements are sequentially labelled as helices α1–α7 and strands β1–β9 (Fig. 1*c*). According to the SCOP protein classification, GvpF belongs to the α+β class, in which α-helices and β-strands are largely segregated.

The N-domain has a three-layer architecture in which a meander β-sheet (β1–β5) is sandwiched by two layers of α-helices (α1–α3; Fig. 1*b*). It adopts a β–α–β–α topology [(β1–β3)–α1–(β4–β5)–(α2–α3); Fig. 1*c*]. The central five-stranded β-sheet, which is divided into two parts by helix α1, consists of a core three-stranded motif adopting an antiparallel topology ordered β1–β3–β2 and two small sequential strands (β4 and β5) (Fig. 1*b*). Owing to the long β1–β2 loop, the core three-stranded motif forms a split β–loop–β motif mostly similar to the split β–α–β motif, which consists of a variation of the ordinary β–α–β unit with a third antiparallel strand inserted between the two parallel β-strands (Orengo & Thornton, 1993). The two layers of α-helices, in which one layer is comprised of helix α1 and the other is comprised of consecutive helices α2 and α3, flank either side of the β-sheet (Fig. 1*b*). It is widely accepted that proteins are classified into the same fold if they have the same major secondary structures in the same arrangement and with the same topological connections (Murzin *et al.*, 1995; Hubbard *et al.*, 1997, 1999; Lo Conte *et al.*, 2000). The N-domain adopts a topology different from any other folds grouped into the α+β class in the SCOP database and thus represents a novel fold.

The C-domain exhibits a two-layer sandwich architecture in which strands β6–β9 assemble into a single four-stranded antiparallel β-sheet, whereas helices α4–α7 gather on one side (Figs. 1*b* and 1*c*). Specifically, the C-domain represents a double-stranded crossover motif containing two hairpin ribbons (Richardson, 1981; Orengo & Thornton, 1993; Efimov, 1994; Zhang & Kim, 2000), one of which consists of β6–(α4–α6)–β7 and the other of β8–α7–β9. Generally, the double-stranded crossover motif is defined as the ferredoxin fold (Orengo & Thornton, 1993; Efimov, 1994; Zhang & Kim, 2000). Compared with the classic ferredoxin fold proteins, the C-domain has an extension region comprising three consecutive helices (α4, α5 and the N-terminal segment of α6), representing a modified ferredoxin fold. In particular, helices



(*a*) (*b*) (*c*)

Figure 4
Structural comparisons. Superpositions of (*a*) the N-domain (magenta) of GvpF with *Thermus thermophilus* hypothetical protein TTHA0061 (PDB entry 2ebe, green), (*b*) the C-domain (cyan) of GvpF with *T. thermophilus* 30S ribosomal protein S6 (PDB entry 2j5a, orange) and (*c*) the extension region (cyan) of GvpF with *Saccharomyces cerevisiae* ubiquinol–cytochrome *c* reductase complex 17 kDa protein (PDB entry 1p84, red). The relevant residues of GvpF are labelled separately.

α4 and α5 are approximately perpendicular to each other and antiparallel to the curved long helix α6.

Further structural analysis revealed that there is no direct interaction between the N- and C-domains of GvpF. However, the two domains pack against each other *via* interactions with the C-tail (Asn232–Leu243; Figs. 1*b* and 2). Deletion of the C-tail leads to expression of the protein as inclusion bodies (data not shown), indicating that the C-tail is indispensable for the folding and/or stability of GvpF. The interfaces flanking the C-tail are composed of a network of hydrophobic interactions and hydrogen bonds (Fig. 2). In detail, the interface formed by the N-domain and the C-tail is mainly stabilized by hydrophobic interactions between Met67 (α1) and Leu234 (C-tail), Leu63 (α1) and Ala263 (C-tail), Pro76 (β4) and Pro237 (C-tail), Pro76 (β4) and Ala241 (C-tail), and Leu104 (α3) and Leu243 (C-tail). Besides, the interface is further fixed by two hydrogen bonds, one of which is donated by the side-chain carboxyl group of Glu60 (α1) and the main-chain amide group of Ala236 (C-tail) and the other of which is formed between the side-chain amino group of His59 (α1) and the side-chain phenolic hydroxyl group of Tyr238 (C-tail). The interactions between the C-domain and the C-tail consist of hydrophobic interactions between Ile187 (β7) and Phe240 (C-tail) and between Leu203 (β8) and Phe240 (C-tail). In detail, the side-chain benzene ring of Phe240 is sandwiched between the side-chain atoms of Ile187 and Leu203. The interface is further stabilized by three hydrogen bonds which are formed between the side-chain amino group of Arg112 (β6) and the main-chain carboxyl group of Asp233 (C-tail), between the side-chain carboxyl group of Glu113 (β6) and the main-chain amide group of Thr239 (C-tail) and between the side-chain carboxyl group of Glu113 (β6) and the side-chain hydroxyl group of Thr239 (C-tail). Multiple sequence alignment revealed that the residues contributing to the formation of the interfaces are highly conserved in GvpF from diverse species of cyanobacteria (Fig. 3*a*).

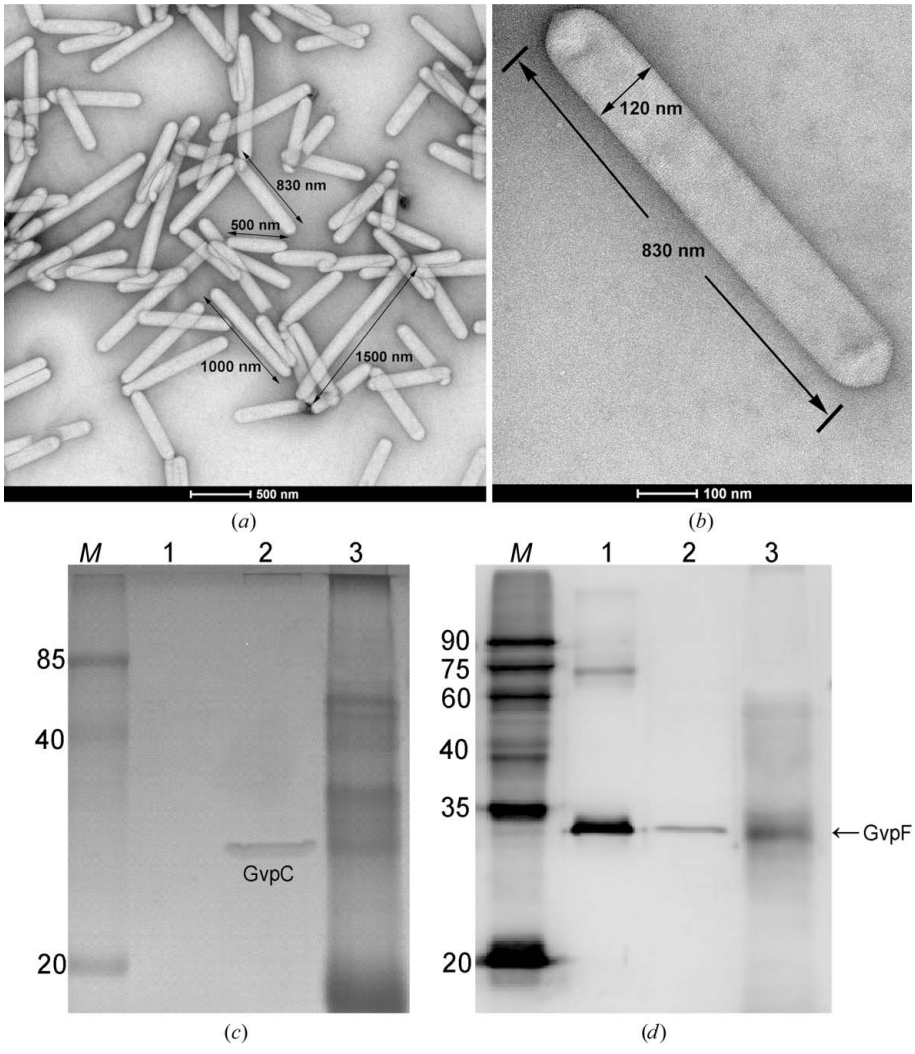


Figure 5
The purified gas vesicles and identification of GvpF. (*a*) Electron micrographs of highly purified intact gas vesicles from *M. aeruginosa* PCC 7806 and (*b*) the dimensions of a gas vesicle. Intact gas vesicles were negatively stained with uranyl acetate. The diameter and different lengths of gas vesicles are labelled. (*c*) SDS–PAGE analysis. Recombinant GvpF (5 ng, lane 1), purified gas vesicles (1.5 µg, lane 2) and cell lysates (30 µg, lane 3) of *M. aeruginosa* were electrophoresed. GvpC (25.153 kDa) is labelled in lane 2. (*d*) Identification of GvpF by Western blotting. Recombinant GvpF (5 ng, lane 1), purified gas vesicles (176 µg, lane 2) and cell lysates (208 µg, lane 3) were electrophoresed, transferred to the membrane and probed with GvpF antibodies. The antibody-amplified bands of GvpF in lanes 1–3 are marked GvpF. Prestained protein standards are displayed in the lanes marked *M* and their molecular masses are indicated in kDa.

3.2. Structural comparison

A *DALI* search was performed using the overall structure of GvpF to gain more structural information. However, the output only enabled us to identify proteins or domains similar to the individual N-domain or C-domain. As a consequence, the two distinct domains were separately input into the *DALI* search.

The *DALI* search using the N-domain gave 29 hits covering eight unique proteins with a *Z*-score higher than 3.5. The top hits include the hypothetical protein TTHA0061 from *Thermus thermophilus* (PDB entry 2ebe; *Z*-score of 5.2; r.m.s.d. of 2.5 Å over 71 C^{α} atoms; RIKEN Structural Genomics/Proteomics Initiative, unpublished work; Fig. 4*a*), followed by the *E. coli* cation-efflux system protein CusA (PDB entry 4dnr; *Z*-score of 3.8; r.m.s.d. of 3.1 Å over 74 C^{α} atoms; C.-C. Su, F. Long & E. Yu, unpublished work) and *Vibrio cholera* O395 acylphosphatase (PDB entry 4hi1;

Z-score of 3.8; r.m.s.d. of 2.7 Å over 62 C^{α} atoms; Nath *et al.*, 2014). The other hits included formimidoyl transferase–cyclodeaminase (PDB entry 2pfd; B. K. Poon, X. Chen, M. Lu, N. K. Vyas, F. A. Quiocho, Q. Wang & J. Ma, unpublished work), nitrogen-responsive transcription factor NrpR (PDB entry 2qyx; Wisedchaisri *et al.*, 2010), engineered protein OR494 (PDB entry 4pww; Northeast Structural Genomics Consortium, unpublished work), 3-hydroxy-3-methylglutaryl-coenzyme A reductase (PDB entry 1t02; Tabernero *et al.*, 2003) and RNA-directed RNA polymerase (PDB entry 4e76; Mosley *et al.*, 2012). Structural comparison showed that in addition to the split β–loop–β motif there are three unique features in the N-domain compared with these top hits. Firstly, the N-domain has two additional helices (α2 and α3). Secondly, the N-domain represents a three-layer sandwich arrangement, whereas the top hit has a two-layer sandwich architecture (Fig. 4*a*). Thirdly, the N-domain has a β1–β2–β3–α1–β4–β5–α2–α3 topological connection (Figs. 1*b* and 1*c*), distinct from the top hit with a β1–α1–β2–β3–β4–α2–β5 topological diagram (Fig. 4*a*). Altogether, we conclude that the N-domain of GvpF adopts a novel fold.

In a comparison of the C-domain with other protein structures deposited in the PDB, *DALI* generated 24 hits for ten unique proteins with a *Z*-score higher than 7.0. Structural comparison revealed that all of these hits and the C-domain adopt a consensus ferredoxin fold. However, compared with the top hit, which is the 30S ribosomal protein S6 from *T. thermophilus* (PDB entry 2j5a; *Z*-score of 8.1; r.m.s.d. of 2.7 Å over 86 C^{α} atoms; Olofsson *et al.*, 2007; Fig. 4*b*), the C-domain has an extension region, making it a modified ferredoxin fold (Fig. 4*b*).

The extension region represents a topology of helix hairpins and a coiled-coil-like fold, which is composed of three helices: helices α4 and α5 in addition to the N-terminal segment of helix α6 (Met150–Lys168). Helices α4 and α5 are kinked at Asn133 by approximately 90°. Helix α5 is anti-parallel to the long helix α6, which is bent at Lys168 (Figs. 4*b* and 4*c*).

A *DALI* search using this extension region enabled us to identify a highly homologous segment of *Saccharomyces cerevisiae* ubiquinol–cytochrome *c* reductase complex 17 kDa protein (Fig. 4*c*), which is also called the subunit 8 protein and referred to as the 'hinge protein', consisting of a bent hairpin helix. Subunit 8 protein is in close contact with cytochrome c_1 and is thought to be essential for proper complex formation between cytochromes *c* and c_1 (Zhang *et al.*, 1998; Iwata *et al.*, 1998). Therefore, we assumed that the extension region of GvpF may participate in protein–protein interactions during the formation of gas vesicles.

As is well known, gas vesicles are ubiquitously found in photosynthetic bacteria, heterotrophic bacteria and haloarchaea (Walsby, 1994; Pfeifer, 2012). However, we noted that the extension region only occurred in photosynthetic bacteria (such as *Microcystis*, *Nostoc* and *Pelodictyon*) and heterotrophic bacteria (*Octadecabacter*, *Desulfotomaculum* and *Serratia*) but not in haloarchaea (*Halobacterium*, *Haloferax* and *Halorubrum*) (Fig. 3*b*), implying that this extension region is species-specific to some degree.

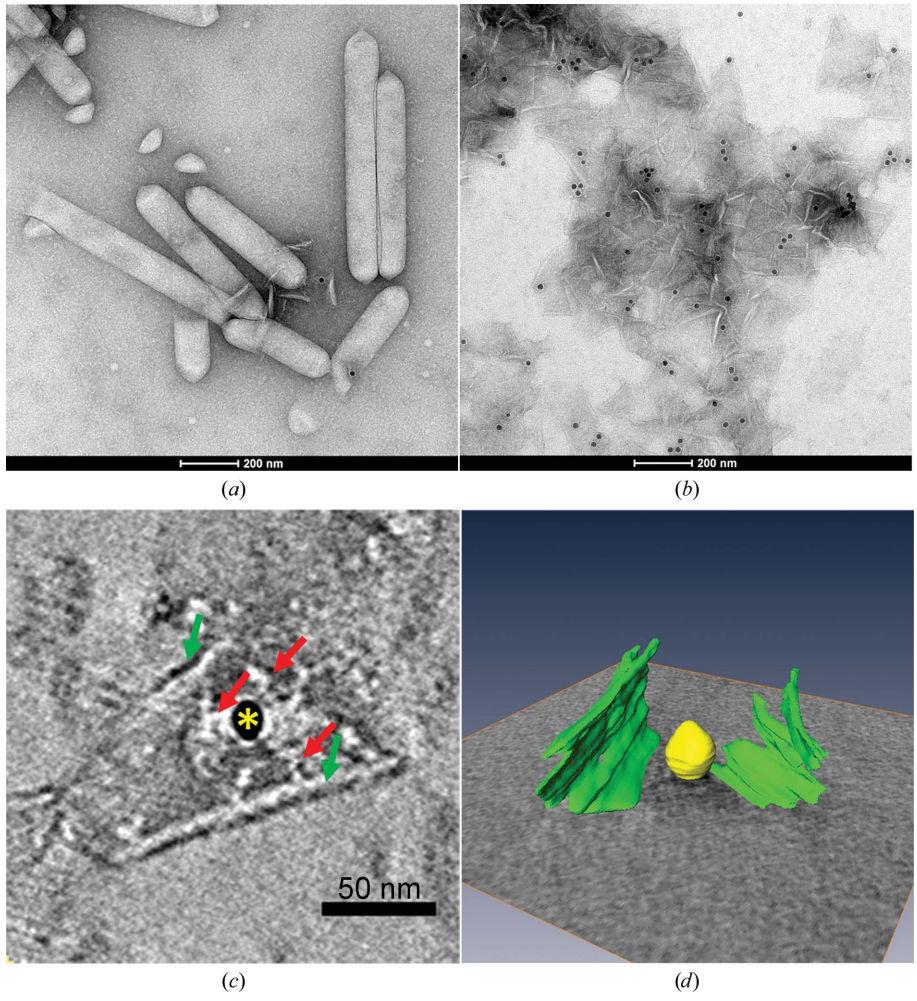


Figure 6
Immunogold localization of GvpF on the gas vesicle wall and electron tomograms of a gold particle labelled on collapsed gas vesicles. (*a*) Intact gas vesicles and (*b*) collapsed segments of gas vesicles probed with anti-GvpF antibodies. The solid black dots with a diameter of 10 nm are gold particles. (*c*) A tomogram slice showing the location of a gold particle on one segment of the collapsed gas vesicles. Red arrows represent putative antibodies conjugated to a gold particle, which is labelled with a yellow asterisk. Green arrows point to the edges of collapsed gas vesicles. (*d*) Three-dimensional rendering of the tomograms. Green regions represent two visible parts of the collapsed gas vesicle segment and the grey plane refers to the location of invisible parts. The gold particle is shown as a yellow sphere.

3.3. GvpF is a structural component of the gas vesicle

The purified gas vesicles with high purity isolated from *M. aeruginosa* PCC 7806 are intact, as shown by electron micrographs (Figs. 5*a* and 5*b*). They adopt a cylindrical shape with conical end caps. The gas vesicles have a similar diameter of about 120 nm and variable lengths from 500 to 1500 nm (Figs. 5*a* and 5*b*). The outer wall represents regularly spaced ribs running nearly perpendicular to the long axis of the vesicle (Fig. 5*b*). Notably, a previous electron micrograph at a higher resolution showed that the ribs in haloarchaeal gas vesicles adopt apparent turns of a shallow spiral (Offner *et al.*, 1998). The second dominant structural component GvpC of purified gas vesicles could be detected by SDS–PAGE (Fig. 5*c*, lane 2) and further confirmed by mass spectrometry (data not shown). Immunoblotting analysis with anti-GvpF antiserum showed the presence of GvpF in the purified gas vesicles and cell lysates (Fig. 5*d*, lanes 2 and 3), suggesting that GvpF is indeed a structural component of gas vesicles.

3.4. Localization of GvpF in the gas vesicle

To precisely locate GvpF on the gas vesicle, immunogold labelling was applied to both intact and collapsed gas vesicles. The gold particles did not attach to intact gas vesicles that had been incubated with anti-GvpF antibodies (Fig. 6*a*). However, when intact gas vesicles were sonicated into collapsed segments, the gold particles showed a distinctly greater tendency to aggregate and a somewhat high density of gold labelling was observed (Fig. 6*b*), indicating that the epitopes of GvpF are exposed to the gas-facing surface of the gas vesicle. Moreover, automated tomography of immunogold labelling was performed to offer direct photographic evidence. An electron tomogram slice shows the location of the labelled gold particle on one segment of collapsed gas vesicles (Fig. 6*c*). In addition, three-dimensional rendering of the tomograms visualized the location of the gold particle, which is embraced by two visible parts of the segment (Fig. 6*d*). Overall, we demonstrated that GvpF is indeed a constitutive component of the gas vesicle and is most likely to be localized on the gas-facing surface.

We are grateful to the beamline staff of the Shanghai Synchrotron Radiation Facility (SSRF) for technical help during X-ray data collection. This work was supported by the Chinese National Natural Science Foundation (Grant Nos. 31370757 and 31070652) and the Ministry of Education of China (Grant No. 20133402110023 and the Program for Changjiang Scholars and Innovative Research Team in University).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Belenky, M., Meyers, R. & Herzfeld, J. (2004). *Biophys. J.* **86**, 499–505.
Bowen, C. C. & Jensen, T. E. (1965). *Science*, **147**, 1460–1462.
Buchholz, B. E., Hayes, P. K. & Walsby, A. E. (1993). *J. Gen. Microbiol.* **139**, 2353–2363.
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.* D**66**, 12–21.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Dunton, P. G., Mawby, W. J., Shaw, V. A. & Walsby, A. E. (2006). *Microbiology*, **152**, 1661–1669.
Dunton, P. G. & Walsby, A. E. (2005). *FEMS Microbiol. Lett.* **247**, 37–43.
Efimov, A. V. (1994). *FEBS Lett.* **355**, 213–219.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Englert, C. & Pfeifer, F. (1993). *J. Biol. Chem.* **268**, 9329–9336.
Hayat, M. A. (1989). *Colloidal Gold: Principles, Methods, and Applications*. New York: Academic Press.
Hayes, P. K., Lazarus, C. M., Bees, A., Walker, J. E. & Walsby, A. E. (1988). *Mol. Microbiol.* **2**, 545–552.
Hayes, P. K. & Walsby, A. E. (1986). *Br. Phycol. J.* **21**, 191–197.
Hayes, P. K., Walsby, A. E. & Walker, J. E. (1986). *Biochem. J.* **236**, 31–36.
Hubbard, T. J. P., Ailey, B., Brenner, S. E., Murzin, A. G. & Chothia, C. (1999). *Nucleic Acids Res.* **27**, 254–256.
Hubbard, T. J. P., Murzin, A. G., Brenner, S. E. & Chothia, C. (1997). *Nucleic Acids Res.* **25**, 236–239.
Iwata, S., Lee, J. W., Okada, K., Lee, J. K., Iwata, M., Rasmussen, B., Link, T. A., Ramaswamy, S. & Jap, B. K. (1998). *Science*, **281**, 64–71.
Jones, D. D. & Jost, M. (1970). *Arch. Mikrobiol.* **70**, 43–64.
Jones, D. D. & Jost, M. (1971). *Planta*, **100**, 277–287.
Jost, M. (1965). *Arch. Mikrobiol.* **50**, 211–245.
Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. (1996). *J. Struct. Biol.* **116**, 71–76.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lo Conte, L., Ailey, B., Hubbard, T. J. P., Brenner, S. E., Murzin, A. G. & Chothia, C. (2000). *Nucleic Acids Res.* **28**, 257–259.
Mastronarde, D. N. (1997). *J. Struct. Biol.* **120**, 343–352.
Mlouka, A., Comte, K., Castets, A. M., Bouchier, C. & Tandeau de Marsac, N. (2004). *J. Bacteriol.* **186**, 2355–2365.
Mosley, R. T., Edwards, T. E., Murakami, E., Lam, A. M., Grice, R. L., Du, J., Sofia, M. J., Furman, P. A. & Otto, M. J. (2012). *J. Virol.* **86**, 6503–6511.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Nath, S., Banerjee, R. & Sen, U. (2014). *Biochem. Biophys. Res. Commun.* **450**, 390–395.
Offner, S., Ziese, U., Wanner, G., Typke, D. & Pfeifer, F. (1998). *Microbiology*, **144**, 1331–1342.
Olofsson, M., Hansson, S., Hedberg, L., Logan, D. T. & Oliveberg, M. (2007). *J. Mol. Biol.* **365**, 237–248.
Orengo, C. A. & Thornton, M. J. (1993). *Structure*, **1**, 105–120.
Pfeifer, F. (2012). *Nature Rev. Microbiol.* **10**, 705–715.
Powell, R. S., Walsby, A. E., Hayes, P. K. & Porter, R. (1991). *J. Gen. Microbiol.* **137**, 2395–2400.
Ren, J., Wen, L. P., Gao, X. J., Jin, C. J., Xue, Y. & Yao, X. B. (2009). *Cell Res.* **19**, 271–273.
Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–339.
Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. (1979). *J. Gen. Microbiol.* **111**, 1–61.
Shukla, H. D. & DasSarma, S. (2004). *J. Bacteriol.* **186**, 3182–3186.
Tabernero, L., Rodwell, V. W. & Stauffacher, C. V. (2003). *J. Biol. Chem.* **278**, 19933–19938.
Walsby, A. E. (1969). *Proc. R. Soc. London Ser. B*, **173**, 235–255.
Walsby, A. E. (1972). *Bacteriol. Rev.* **36**, 1–32.
Walsby, A. E. (1982). *J. Gen. Microbiol.* **128**, 1679–1684.

Walsby, A. E. (1994). *Microbiol. Rev.* **58**, 94–144.
Walsby, A. E. & Armstrong, R. E. (1979). *J. Mol. Biol.* **129**, 279–285.
Walsby, A. E. & Buckland, B. (1969). *Nature (London)*, **224**, 716–717.
Walsby, A. E. & Hayes, P. K. (1988). *J. Gen. Microbiol.* **134**, 2647–2657.
Weathers, P. J., Jost, M. & Lamport, D. T. (1977). *Arch. Biochem. Biophys.* **178**, 226–244.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wisedchaisri, G., Dranow, D. M., Lie, T. J., Bonanno, J. B., Patskovsky, Y., Ozyurt, S. A., Sauder, J. M., Almo, S. C., Wasserman, S. R., Burley, S. K., Leigh, J. A. & Gonen, T. (2010). *Structure*, **18**, 1512–1521.
Wyffels, J. T. (2001). *Microsc. Microanal.* **7**, 66.
Zhang, Z., Huang, L., Shulmeister, V. M., Chi, Y.-I., Kim, K. K., Hung, L.-W., Crofts, A. R., Berry, E. A. & Kim, S.-H. (1998). *Nature (London)*, **392**, 677–684.
Zhang, C. & Kim, S.-H. (2000). *J. Mol. Biol.* **299**, 1075–1089.