Acta Crystallographica Section D Biological Crystallography ISSN 1399-0047

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

 $\ensuremath{\mathbb{C}}$ 2014 International Union of Crystallography

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

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 gasfacing 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 Received 27 July 2014 Accepted 25 September 2014

PDB reference: GvpF, 4qsg

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 gvpgenes 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 (Haves et al., 1988; Walsby & Haves, 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 $6 \times$ His tag and overexpressed in *Escherichia coli* strain BL21 (DE3) harbouring the pKY206 plasmid (Novagen) using $2 \times$ YT culture medium (5 g NaCl, 16 g Bacto tryptone and 10 g yeast extract per litre) containing $30 \,\mu g \,m l^{-1}$ kanamycin and $5 \,\mu g \,m l^{-1}$ tetracycline. The cells were grown at $37^{\circ}C$ to an $A_{600 \text{ nm}}$ of 0.8. Expression of the recombinant protein was induced with 0.2 mM isopropyl β -D-1-thiogalactopyranoside for a further 4 h at 37°C before harvesting. The cells were collected by centrifugation at 4000g for 10 min and resuspended in 40 ml lysis buffer (200 mM NaCl, 20 mM Tris-HCl pH 8.0). After 30 min of sonication and centrifugation at 12 000g 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 mM NaCl, 20 mM Tris-HCl pH 8.0). The target protein was eluted with 300 mM imidazole, 200 mM NaCl, 20 mM Tris-HCl pH 8.0 and loaded onto a HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with 150 mM NaCl, 14 mM β -mercaptoethanol, 20 mM 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 $A_{600 \text{ nm}}$ 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 $A_{600 \text{ nm}}$ 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 $1 \ \mu l \ 10 \ mg \ ml^{-1}$ protein sample with $1 \ \mu l$ 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 singlewavelength 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 *REFMAC5* (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 μ mol 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 $A_{750 \text{ nm}} = 0.66-1.0$.

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 mM 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 mM Tris-HCl pH 7.5, 5 mM sodium cyanide six or seven times. The purified gas vesicles were stored in 6.3 mM ammonium bicarbonate containing 5 mM 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 ₂ 21
Unit-cell parameters	
$a = b (\tilde{A})$	81.92
<i>c</i> (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 \ddagger/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 ($^{\circ}$)	0.946
Mean <i>B</i> 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

[↑] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle | / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an observation and $\langle I(hkl) \rangle$ is the mean value for its unique reflection; summations are over all reflections. [‡] R factor = $\sum_{hkl} ||F_{obs}| - |F_{cabc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{cabc} are the observed and calculated structure-factor amplitudes, respectively. § R_{free} was calculated with 5% of the data, which were excluded from refinement. ¶ Root-meansquare deviation from ideal values. [†] \uparrow 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 G² 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 μ g protein or 176 μ g 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 mM Tris-HCl pH 6.8, 20% glycerol, 0.2 M dithiothreitol, 5% SDS, 0.02% bromophenol blue) and boiled for 10 min. After

research papers

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 μ m 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 mM NaCl, 10 mM 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 gas vesicles *Microcystis* 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 facedown 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 $(1 \times PBS; 10 \text{ m}M \text{ Na}_2\text{HPO}_4,$ 1.8 mM KH₂PO₄, 137 mM NaCl, 2.7 mM 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 G² 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.

research papers

		β1		η1	β2	β3		000000000000000000000000000000000000000	(1	β4	
MICrocystis		1 10	2	0	30	40	50	60)	70	80
Microcvstis		MTVGLYLYGI	FPEPIPDGL	VLOGIDNE	VHSEMII	GFSFLYSA	AHKEKYLA	SRRYLICHE	KVLETVMEA.	GFTTLL	LRFG
Nodularia		.MNFGLYLYGI	FPDIIPSSV	SITGLDGKE	VYSQVVI	GLTFLYSE	ASKEKYLA	SRRNLLTHE	KVLEETMQA.	GFHVLLE	LOFG
Nostoc		.MSSGLYLYGI	[FPDPIPETV	TLQ <mark>G</mark> LDSQI	VYSQIII)GFTFL <mark>YS</mark> E	AKQEKYLA	S R <mark>R N L I S</mark> H E	K <mark>VLEQAM</mark> HA.	GFRTLLF	PLR FG
Oscillatoria		. MDCGFYLYGI	FPTPGPCNL	NLKGLDQKI	VHSKNLI	OGFTFLYSQ	ATQEKYLA	SRRHLLSHE	KVLEEAMSA.	GFRTLLE	LRFG
Raphidiopsis	onaia	MSSNGLYLYGI	LPKPLAQDL	TLEGLDKQE	VKTODVN	IGFVFLYSH	ALQKQYLA	SRRNLLAHE	RVLEQVMQE.	GCRTLLI	LRFG
Arthrospira	lopsis	MSYSLYLYGI	LPKTKL	HLOGLDNOF	VYTHEIL	OFIFLYSE	AOOERYLA	SRRNLLGHE	KVLEAMMOE.	GYRTIL	LOFG
Synechococcus	3	NPHSALYLYGI	VAAPGPRHL	SVVGLDKQI	VQVHGLO	SSLAFLYSQ	AROERYLA	SRANLLAHE	AVLEKVMNE.	GHRALLE	LOFG
Trichodesmium	n	MDESKDESKNO	GLKGSNESKD	ELK <mark>GLD</mark> KEI	VKIQDVI	DEFAVLYSI	AKKERYLA	S R <mark>R N L I T</mark> H E	K <mark>VLE</mark> SAMEA.	GYRNLLP	MQFG
Pseudanabaena	L	MSLYLYAI	LQAENLDLIQNL	DLQGMNAQS	QFHALF	PFAIAYSE	AQQERYLA	SRANLITHE	TVLESLMKVI	DPHQAVPLE	LQFG
		β5 n2	α2 α	3	β6	α4		α.5			α.6
Microcystis		→ eée s	2000 20000	00000 -	1	► 200000	lee ee	000000	22222	000000000	2000
		90	100	110	12	20	130	140	150	160	
Microcystis		LVIKTWESVTE	EQLITPYKTQLKE	LFAKLSGQF	REVSIKIF	WDNQWELQ	AALESNPK	LKQERDAMM	IGKN <mark>L</mark> NMEEII	HIGQLIEAT	CVLR R
Nodularia		LVIKDWDAITI	COLIEPYKEQLHN	LFQKLAGQF	REVSVKIE	WDSKSELQ	AMMESNLA	LKQERDNME	GRKLSIEEVI	HIGQLIESN	ILSAR
Oscillatoria		MVIKTWEAVTE	COLTNPYOTOLKE	LFEKLTGLO	EVSVKIP	WETTVELO	ALMESNES	LKEKRDSLE	GOALSMDOII	OIGOMIEAZ	LEHR
Raphidiopsis		LVVKDWQEISN	VQLINPSEGQLYQ	LFEKLAGQF	EVSVKII	WDTKSELQ	MMMDSNPD	LKRQRDAME	GKNLSMDEVI	QIGQLIESN	ILMOR
Cylindrosperm	opsis	LVVKDWQEISN	NQLINPSQNQLHQ	LFEKLAGQF	REVSVKII	WDTKSELQ	MMIDSNPD	L KQQ <mark>R</mark> DAME	EGKN <mark>L</mark> SMDEVI	QIGQLIESN	ILMER
Arthrospira		LIVSDWDAVAK	KQLVNPYHERLKE	LFVKLEGNF	EVGVKVF	WNETAELN	ALMEENQD	LRERRDRLE	EGQPLSMDEVI	AIGREIESN	VILDR
Synechococcus	3	LVVSGWEQVER	CDLVQPRLADLLA	LERLEGKE	EVGVKVF	WDPEQELK	LGLEENPA.	LKARRDEMA	AGAPLGLDAVV	EIGRALEQI	LEQR
Pseudanabaena		LVVENWEDVOF	RDLLIPYETKLKE	LIHNLIGKE	EVSVKLF	WNOTEELN	LAVAENAG	LRORREALV	GKVLSMDEAI	AIGOELESA	AIEER
									_		
			η3 β7		β8	η4	α7	β	9		
Microcystis		170 1		2 (210	220	TT		T	
Mignogustia		KODITOVERD		DMEDDAT							
Nodularia		KOSVIEVESNE	CINPLATEVIDGE	LMTEDMIYN	TAFLIPW	VESESKEGE	TVEATDIK	FGDR LRI	I RYN <mark>HFTAPY</mark> T	FAOLAEP.	
Nostoc		KESIIQVFFDE	LKPLADEVIESD	PMTEDMIYN	AAFLIPW	VENESIFSQ	QVESIDHK	FDER. LRI	I R Y N N F T A P Y T	FAQIS	
Oscillatoria		KETVVQAFRNE	E <mark>LNYLAVEVIE</mark> SA	MLREGMIYN	JAAF <mark>LI</mark> PW	VESEPKFGE	TVEEIDLK	FGE <mark>RLR</mark> I	I R Y N <mark>N F</mark> T A P Y <mark>T</mark>	FAQLS	
Raphidiopsis		KEGVIQVFREE	LQPFAKDVVENE	PMMEEMIYN	JAAFLIPW	VDAESAFSD	RVESIDKQ	FGDR. LRI	I R Y N <mark>N F</mark> T A P Y <mark>T</mark>	FAQLSPS.	
Cylindrosperm	opsis	KEAVIQVEREC	2LQPFAKDVVENE	PMMEEMIYN	AAFLIP	VDAESAFSD	RVESTDRQ	FGDR.LRI	RYNNFTAPYT	FAQLSPS.	
				SMTDLMTVN		MNCHFKHCO	KVEETDNV		DVNNETADVN		
Synechococcus	3	RORIAOTETAA	ALSPLASDRVEGD	SMTDLMIYN LLTENMAYN	IAAYLIPV IGSFLINV	VNSEEKFSQ VEDEPVFAO	KVEEIDNY KVEELDOT	FPKRLRI FOGRLRI	I R Y N <mark>N F</mark> T A P Y <mark>N</mark> I R Y N N F T A P Y <mark>N</mark>	FTQLD FVKL	
Synechococcus Trichodesmium	5 1	RQRIAQTETAP KQNIIEIFQET	ALSPLASDRVEGD LNKMAIEVIENE	SMTDLMIYN LLTENMAYN VQTEK <mark>M</mark> IYN	NAAYLIPW N <mark>GSFLIN</mark> W NAAY <mark>LI</mark> PW	VNSEEKFSQ VEDEPVFAQ VDQEEDFGE	KVEEIDNY KVEELDQT KVETIDSK	FPKRLRI FQGRLRI LCE <mark>R</mark> GNFT <mark>I</mark>	I RYN <mark>NF</mark> TAPY <mark>N</mark> I RYN <mark>NF</mark> TAPY <mark>N</mark> I RYN <mark>SF</mark> TAPY <mark>N</mark>	FTQLD FVKL FARIRQQD	
Synechococcus Trichodesmium Pseudanabaena	3 1 1	RQRIAQTFTA KQNIIEIFQET QQVIITAFLNT	I PLALDYVENE LSPLASDRVEGD LNKMAIEVIENE LKPLSHEYAEGE	SMTDLMIYN LLTENMAYN VQTEKMIYN LLTESMIYN	IAAYLIPW IGSFLINW IAAYLIPW IGSFLIDW	VNSEEKFSQ VEDEPVFAQ VDQEEDFGE VDKEPEFAK	KVEEIDNY KVEELDQT KVETIDSK AVEALDQQ	FPKRLRI FQGRLRI LCERGNFTI FENRLRI	LRYN <mark>NF</mark> TAPY <mark>N</mark> LRYN <mark>NFTAPYN</mark> LRYN <mark>SF</mark> TAPY <mark>N</mark> LRYN <mark>DF</mark> TAPY <mark>N</mark>	FTQLD FVKL FARIRQQD FVKVDRD.	
Synechococcus Trichodesmium Pseudanabaena	3 1 1	ENELITAFODO RORIAQTETA KONIIEIFOEI QOVIITAFLNI	LIPLALDYVENE ALSPLASDRVEGD LNKMAIEVIENE LKPLSHEYAEGE	SMTDLMIYN LLTENMAYN VQTEKMIYN LLTESMIYN	IAAYLIPW IGSFLINW IAAYLIPW IGSFLIDW	VNSEEKFSQ VEDEPVFAQ VDQEEDFGE VDKEPEFAK	KVEEIDNY KVEELDQT KVETIDSK AVEALDQQ	FPKRLRI FQGRLRI LCERGNFTI FENRLRI	IRYNNFTAPYN IRYNNFTAPYN IRYNSFTAPYN IRYNDFTAPYN	FTQLD FVKL FARIRQQD FVKVDRD.	
Synechococcus Trichodesmium Pseudanabaena	5 1 1	ENEIITAFQDC RQRIAQTFTAF KQNIIEIFQEI QQVIITAFLNI	LIPLALDYVENE ALSPLASDRVEGD LNKMAIEVIENE LKPLSHEYAEGE	SMTDLMIYY LLTENMAYY VQTEKMIYY LLTESMIYY	IAAYLIPW IGSFLINW IAAYLIPW IGSFLIDW (a)	VNSEEKFSQ VEDEPVFAQ VDQEEDFGE VDKEPEFAK	KVEEIDNY KVEELDQT KVETIDSK AVEALDQQ	FPKRLRI FQGRLRI LCERGNFTI FENRLRI	IRYNNFTAPYN IRYNNFTAPYN IRYNSFTAPYN IRYNDFTAPYN	FTQLD FVKL FARIRQQD FVKVDRD.	
Synechococcus Trichodesmium Pseudanabaena	5 1 1	ENEIITAFODO RORIAOTETAF KONIEIFOEI QOVIITAFLNI	LSPLASDRVEGD LSPLASDRVEGD LNKMAIEVIENE LKPLSHEYAEGE	SMTDLMIYN LLTENMAYN VQTEKMIYN LLTESMIYN	AAYLIPW IGSFLINW IAAYLIPW IGSFLIDW (a) (5	VNSEEKFSQ VEDEPVFAQ VDQEEDFGE VDKEPEFAK	KVEEIDNY KVEELDQT KVETIDSK AVEALDQQ	FPKRLRI FQGRLRI LCERGNFTI FENRLRI	K RYNNFTAPYN RYNNFTAPYN RYNSFTAPYN RYNDFTAPYN	TQLD FVKL FARIRQQD FVKVDRD.	β7
Synechococcus Trichodesmium Pseudanabaena	Micro	RQRIAQTFTAF KQNIIEIFQET QQVIITAFLNT	ALS PLASD RVEGD LNKMAIEVIENE LKPLSHEVAEGE	SMTDLMIYN LLTENMAYN VQTEKMIYN LLTESMIYN α4	$\begin{array}{c} \text{AAYLIPW}\\ \text{IGSFLINW}\\ \text{IGSFLIDW}\\ \text{IGSFLIDW}\\ (a)\\ \alpha 5\\ 000000 \end{array}$	VNSBEKFSQ VEDEPVFAQ VDQEEDFGE VDKEPEFAK	KVEEIDNY KVEELDQT KVETIDSK AVEALDQQ	FPKRLRI FQGRLRI LCERGNFTI FENRLRI	I RYNNFTAPYN I RYNNFTAPYN I RYNSFTAPYN I RYNDFTAPYN 26 20000000000000000000000000000000000	FTQLD FVKL FARIRQQD FVKVDRD. Λ η3 00000000	β7
Synethococcus Trichodesmium Pseudanabaena	Micro	ENELITAPOLA RORIAQTETAK KONIIEIFQET QOVIITAPLNT		SMTDLMIYN LLTENMAYN VQTEKMIYN LLTESMIYN 00000000 130	$\begin{array}{c} \text{AAYLIPW}\\ \text{IGSFLINW}\\ \text{IGSFLIDW}\\ (a)\\ (a)\\ 000000\\ 14 \end{array}$	VNSBEKFSQ VEDEPVFAQ VDQEEDFGE VDXEPEFAK VDKEPEFAK	KVEEIDNY KVEELDOT KVETIDSK AVEALDOO 159	FPKRLRI FQGRLRI LCERGNFTI FENRLRI 2000000000000000000000000000000000000	I RYNNFTAPYN I RYNNFTAPYN I RYNSFTAPYN I RYNDFTAPYN 44 44 44 46 00000000000000000000000000	η3 00000000 180	<u>β7</u>
Synethococcus Trichodesmium Pseudanabaena	Micr.	ENELITAPOLA RORIAQTFTAF KONIIEIFOET QOVIITAPLNT		a4	AAYLIPW GSFLINW AAYLIPW GSFLIDW (a) α5 <u>000000</u> 14 εNPKLKQE	VNSBEKFSQ VEDEPVFAQ VDQEEDFGE VDXEPEFAK VDKEPEFAK QQ ? RDAMMGKNL	KVEEIDNY KVEEIDSK AVEALDOO 150 MEEIIH	FPKRLRI FQGRLRI LCERGNFTI FENRLRI 0000000000 160 11GQLIEATV	EXYNNFTAPYN EXYNNFTAPYN EXYNSFTAPYN CYNSFTAPYN COUDEDCOUDE 2000000000000000000000000000000000000	η3 00000000 189 RQQD NHRAQ	
Artifospira Synechocccus Trichodesmium Pseudanabaena	Micro Nodu Nost	ENGLIATOR RQRIAQUETAA KONIIEIPOET QOVIITAELN1 Docystis Docystis Laria Doc	β6 β6 β6 β6 β6 β6 β6 β6 β6 β6	CALQANES CALQANES CALQANES CALQANES CALQANES CALQANES CALQANES	AAYLIP GSFLIN AAYLIP (<i>a</i>) (<i>a</i>) (<i>a</i>) 000000 14 NPKLKQE NLALKQE NLALKQE	VNSBEKFSQ VED B V FAQ VDQE ED FGE VDKE PEFAK Q Q RDAMMGKNL RDAMMGKNL RDAMMGKNL	KVEEIDNY KVEELDQT KVEIDSK AVEALDQQ 150 NMEEITH SIEEVIH	FPKRLRI FQGRLRI LCERGNFTI FENRLRI 000000000 160 160 11GQLIESNL 11GQLIESNL	RYNNFTAPYN RYNNSFTAPYN RYNSFTAPYN RYNDFTAPYN 20202020202020 179 /LRRKODIIOVF .SARKOSVIEVY	TCLD FVKL FARIRQQD FVKVDRD. ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	β7 EVIE EVID EVID
Artifospira Synechococcus Trichodesmium Pseudanabaena	Micro Micro Nodu Nost	ENGLIATAGUE RORIAOTETAA KONIIEIFOE QOVIITABLNT Docystis Laria Docustis Laria Docustis	β6 β6 β2 μα	a4 COLOCALES COLOCALES CALCANES CAELQAMMES CAELQAMMES CAELQAMMES CAELQAMMES	AAYLIPW GSFLINW IGSFLIDW (<i>a</i>) (<i>a</i>) 000000 14 NPKLKQE HQDLKQK NPSLKEK	NNSBEKFSQ PDDPVFAQ DQEEDFGE NDKBPEFAK Q RDAMMGKNL RDAMMGKNL RDQMEGKAL RDQHEGKAL	KVEELDOT KVETIDSK AVEALDOO 150 NMEEITH SIEEVIH SMEEVIH SMEDITQ	FPKRLRI FQGRLRI LCERGNFTI FENRLRI 16000000000000000000000000000000000000	E Y NN F TA P YN E YNN F TA P YN E YN S FTA P YN E YND F TA P YN 6 200000000000000000000000000000000000	TCLD FVKL FARIRQQD FVKVDRD. NR 180 RDQLNHRAQ SMELNPLAT FDELKPLAT FDELKPLAY	β7 EVIE EVID EVIE EVIE
Architospira Synechococcus Trichodesmium Pseudanabaena photosynthetic	Micro Micro Nodu Nost. Sci: Raph	ENELITACOLO RQRIAQUETARA KONIIEIEQET QUVIITAELNT Ocystis Laria Do Llatria diopsis	β6 β6 β6 β6 β6 β6 β6 β6 β6 β6	SMTDLMIL LLTENMAX VQTEKMIY LLTESMIY 20000000 330 WELQAALES SELQAMMES SELQAMMES SELQAMMES SELQAMMES SELQAMMES	AAYLIPW GSFLINW AAYLIPW GSFLIDW (<i>a</i>) α5 <u>000000</u> 14 NPKLKQE NDALKQK NPSLKEK NPSLKEK	NISEBERES EDDEPUE EDDEPUE EDDEPUE EDDEPERA (DKEPERAK Q Q Q Q RDAMMGKNL RDAMMGKNL RDAMMGKNL RDAMGGRKL RDAMGGRKL RDAMGGRKL	KVEELDQT KVETIDSK AVEALDQO 150 NMEEITH SIEDVIH SMEQVIH SMDQVIC	FPKR .LRI FQGR .LRI FENR .LRI 1000000000 160 1160 1160 1160 1160 116	E Y NN PT A P YN E YNN PT A P YN E YNN S PT A P YN E YN S PT A P YN C YN S PT A PYN C YN S PT A PYN A PYN C YN S PT A PYN C	TQLD FVKL FVKL FVKLDRD. FVKVDRD. FVKVDRD. ***********************************	β7 EVIE EVIE EVIE EVIE EVIE DVVE
photosynthetic bacteria	Micro Micro Nodu Nosci Raph Cylin Arth	ENELITACOO RQRIAQUSTALA KONIIEISOEJ QOVIITAELN7 Docystis Docystis Latia Doc Llatoria ddiopsis ddrospermopsis rospira	β6 β2 β2 μ3 β2	CAL CAL CAL CAL CAL CAL CAL CAL	AAYLIP (GSFLIN (A) (C) (C) (C) (C) (C) (C) (C) (C) (C) (C	NISEBEKESQ DDDEPDFGE DDEPDFGE VDKEPEFAK P RDAMMGKNL RDNMBGKAL RDSNEGKAL RDAMGKAL RDAMGKNL	XVEEIDNY XVETIDSK XVETIDSK AVEALDQO 150 NMEEITH SMEEVIH SMEEVIH SMDQTIQ SMDEVIG SMDEVIG SMDEVIG	FPKE.LLRI FQGR.LLRI LCERGNFTI LCERGNFTI JCOLOCOCOCO 160 IIGQLIESNI IIGQLIESNI IIGQLIESNI IIGQLIESNI IIGQLIESNI IIGQLIESNI	K K K K K K K K K K K K K K	TOLD FARIRQQD VKVDRD. VVVDRD. 180 RDQLNHRAQ SNELNPLAT FDELKELAPLAT RNELNVLAV REELQPFAK REDQLQFFAK	β7 EVIE EVIE EVIE DVVE DVVE DVVE
photosynthetic bacteria	Micr Micr Nodu Nosti Raph Cyliz Arth Syne	ENELITACUC RQRIAQTETAR KONIIEIRQEI QQVIITAELN1 Docystis Laria Doc Llatoria Ldiopsis ndrospermopsis cospira chococcus	β6 β6 β6 β6 β6 β6 β6 β6 β6 β6	SMTDLMILX LLTENMAX VQTEKMIX LLTESMIX 20202020 130 20020200 130 20020200 130 20020200 130 20020200 20020200 20020000 2002000000 200200	AAYLIP (GSFLIN (A) (C) (C) (C) (C) (C) (C) (C) (C) (C) (C	NNGEBEKESQ EDDEPVEAQ DDEPDEDEGE NDKEPEFAK P RDAMMGKNL RDAMMGKNL RDAMEGKAL RDAMEGKNL RDAMEGKNL RDAMEGKNL RDAMEGKNL	KVEELDAY KVEELDAT KVETIDSK AVEALDQO 150 NMEEITH SMEEVTH SMEEVTH SMDEVTC SMDEVTC SMDEVTC SMDEVTC SMDEVTC	FPKR.LLRI FQGR.LLRI LCERGNFTI FENN.LLRI 160 1100011ENL 110011ENL 110011ENL 110011ENL 110011ENL 110011ENL 110011ENL 110011ENL	RYNN FTAPYN RYNN SFTAPYN RYN SFTAPYN RYND FTAPYN 200000000000 170 YLRRKODIIOVF SARKOSVIEV LLSRKESTVVOAK MCRKESVIOVF MERKEAVIOVF LDRENEIINAF	TQLD FARIRQQD FVKUDRD. NRQUNHRAQ SNELNFLAF FDELKELAPLAR RNELNYLAV REELQFFAK REQUQFFAK REQUQFFAK REQUQFFAK	β7 EVIE EVIE EVIE DVVE DVVE DVVE DYVE DYVE
photosynthetic bacteria	Micr. Micr. Nodu Nost. Osci Raph Cyli Arth Syne Tric	ENGLIATAGU RQRIAQTETAR KONIIEIPOE QOVIITAELN7 DOCYSTIS DOCYSTIS Laria DC Llatoria Llatoria Llatoria Llatoria Chocoscus nodesmium Decosos	β6 β6 β2 μ Δ μ Δ μ Δ μ Δ μ Δ μ Δ μ Δ μ Δ μ Δ μ	SMTDLMILY LLTENMAXY VQTEKMIY LLTESMIY 20000000 130 20000000 130 20000000 130 20000000 20000000 20000000 20000000 2000000	AAYLIP GSFLINK GSFLINK (a) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	NISEBERES EDDEPUER EDDEPUER EDDEPUER EDDEPUER EDEPERAK Q Q Q RDAMMGKNL RDAMGGKNL RDAMGGKNL RDAMGGKNL RDAMGGKNL RDAMGGKNL RDAMGGKNL	KVEELDQT KVEELDQT KVETIDSK AVEALDQQ 150 NMEEITH SHEVTH SMDQTIQ SMDEVTQ SMDEVTQ SMDEVTQ SMDEVTQ SMDEVTQ SMDEVTQ SMDEVTQ	FPKR.LLRI FQGR.LLRI LCERGNFTI FENR.LLRI 10000000000 160 IIGQLIESNL IIGQLIESNL IIGQLIESNL IIGQLIESNL IIGRALESLL IIGRALESL	RYNN FTAPYN RYN SFTAPYN RYN SFTAPYN RYN SFTAPYN 60 20000000000000000000000000000000000	TQLD FVKL FVKLQDRD. VKVDRD. 1800 RDQLNHRAQ SNELNYLAV REELQFAK QDGLIPLAV REELQFAK QDGLIPLAS QCFFAK QDGLIPLAS QCFFAK	β7 EVIE EVIE EVIE DVVE DVVE DRVE DRVE EVIE EVIE
photosynthetic bacteria	Micr Nodu Nosti Osci Raph Cyli Arth Syne Tric Pseu	ENELITACUO RQRIAQUTTAP KQNIIEISQET QQVIITASLNT boystis Laria boystis Latoria Ldiopsis ddrospermopsis cospira chococcus nodesmium lanabaena lictyon	β6 β6 β6 β6 β6 β6 β6 β6 β6 β6	CALINE AND	AAYLIP GSFLINK GSFLINK (G) (G) (G) (G) (G) (G) (G) (G) (G) (G)	NIGHE KHSQ IDD I PUFAQ IDD I PUFAQ IDD I PUFAQ IDD I PUFAC IDD I	KVEEIDNY KVEEIDQT KVETIDSK AVEALDQT NMEBIT SIEEVTH SIEEVTH SMDEVIC SMDEVIC SMDEVIC SMDEVIC SMDEVIC SMDEVIC SMDEVIC SMDEVIC SMDEVIC	FPKB.LLRI FQGR.LLRI LCERGMFTI FENR.LLRI GQLQQQQQQ 160 160 11GQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGREISNI UGREISNI UGREISNI UGQKIEQGM	E Y NN PT A P YN E YN N PT A P YN E YN S PT A P YN C YN S PT A P YN C YN D FT A P YN C YN D FT A P YN C YN D FT A PYN C YN D F	TQLD FARIRQQD VKVDRD. VKVDRD. 180 RDQL0101 SNELP LAT FDELK PLAD FDELK PLAD REELQ PPAK REQLQ PPAK REQLQ PPAK QEGLIFLAL TAALS PLAS QETLINKMAI LUNTLK PLSH	β7 EVIE EVIE DVVE DVVE EVIE EVIE EVIE EVIE
photosynthetic bacteria	Micr. Nodu. Nost. Osci: Raph Cylii Arth Syne: Trici Pseu. Pseu.	ENELITACOO ENELITACON RCRIACTETA KONIIEISOE QUVIITAELN Ocystis Laria Oc Llatoria dicospis drospermopsis rocospira cospira cospira lanabaena lictyon lacabacter	β6 β2 μ μ μ μ μ μ μ μ μ μ μ μ μ μ μ μ μ μ μ	CALL TENMARY VQTEKMIN LLTESMIN COLOCOCO 130 WELQAALES CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS MIELSLLEN FEELNLAVAE FEELNLAVAE FRIYENIQW	AAYLIP GSFLINK GSFLINK (A) (a) (a) (a) (b) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	NIGHE KHSQ DD DFVFAQ DD FDFGE VD KEPFGE VD KEPFFAK PEDAMMGKNL RDNMEGRKL RDNMEGRKL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDAMGGKAL RDAMGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL RDAMGGKAL	KVEEIDNY KVEEIDOT KVETIDSK AVEALDQO 150 NMEEITH SIBEVTH SMEEVTH SMDQTIC SMDEVTA SMDEVTA SMDEVTA SMDEVTA SMDEVTA SMDEVTA SMDEVTA SMDEVTA SMDEVTA	FPKB.LLRI FOGR.LLRI LCERGNFTI LCERGNFTI LCERGNFTI ICQLIENT IIGQLIENT IIGQLIENT IIGQLIENT IIGQLIENT IIGQLIENT IIGREIENT IIGREIENT IIGREIENT	K VIN PTAPYN E VIN PTAPYN E VIN SFTAPYN C VIN SF	TQLD FARIRQQD VKVDRD. VKVDRD. SNELING SNELINGA SNELINGA FDELKELAD RNELNVLAV REELQFFAK REELQFFAK REELQFFAK REELQFFAK QDGLIFLAL TLALSFIAE LDILSFIAE LSRIRFHWH	β7 EEVIE EVIE EVVE DVVE EVVE EVVE EVVE EVV
photosynthetic bacteria	Micr. Nodu Nost. Osci Raph Cyli Arth Syne Tric Pseu Pelo Octaa Geob	ENGLIATORO RQRIAQTETAR KONIIEISQEY QQVIITAELNY Docystis Laria Doc Llatoria ddiogsis ndrospermopsis cospira chococcus nodesmium lanabaena lictyon lecabacter acter	β6 2.3 β2 2.3 β6 2.3 β6 2.3 β6 2.3 β6 2.3 β2 2.3 β6 2.3 β6 2.3 β6 2.3 β6 2.3 β2 3.3 β2 3.4	SMTDLMIL LLTENMAX VQTEKMIN LLTESMIN 20202020 130 202020200 130 202020200 20202020 20202020 2020 202020 200 200 200 200 200 200 200 200 200 200 200 20	AAYLIP GSFLINK GSFLINK (A) (a) (a) (a) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	NISEBERS DD B PUFAQ DD B PUFAQ DD B PUFAC ND B PUFAC P R DAMMGKNL R DAMMGKNL R DAMGKNL R DAMGKAL R DAMGKNL R D	KVEELDOT KVEELDOT KVETIDSK AVEALDQO 150 NMEEITH SMEEVIH SMEEVIH SMDEVIC	FPKB.LLRI FQGR.LLRI LCERGNFTI FENN.LLRI MIGQLIENL IGQLIENL IGQLIENL IGQLIENL IGQLIENL IGQLIENL IGQLIENL IGQLIENL IGQLIENN IGQLIENN IGRALBOLL IGQKIEQGM IGQLIESNI IGRALBOLL IGQKIEQGM	RYNN FTAPYN RYNN SFTAPYN RYN SFTAPYN RYND FTAPYN 20000000000000 170 YLRRKODIIOVF JSARKOSVIEV LLRRKODIIOVF MERKETVVOAF MCRKECVIOVF MERKEAVIOVF LLDRENEIINTR MCRKECNIOVF MERKEANICVF SERCOVICTAR	TQUD FARIRQQD VKVDRD. 180 RDQLNHRAQ SNELNPLAT FDELKVLAD FDELKVLAD FDELKVLAD RNELNVLAV QDGLIPLAL TAALSPLAS QEGLIPLAL LDTLKVLSPLAS LSRIRPHVH LDKLKPJSP	β7 EVIE EVIE EVIE DVVE DVVE DVVE EVIE EVIE
photosynthetic bacteria	Micr. Nodu Nost. Osci: Raph Cyli: Arth Syne. Tric Pseu. Octaa Geob. Koril Desu.	ENELITACUO RQRIAQUTTAL KQNIIEIFQET QQVIITAFLMT bocystis latoria diopsis drospermopsis rospira codesmium lanabaena inctyon lacabaeter scter jacter fotomaculum	β6 β6 β6 β6 β6 β6 β6 β6 β6 β6	CALINE AND	AAYLIP GSFLINK GSFLINK (G) (G) (G) (G) (G) (G) (G) (G) (G) (G)	NISE EK ES Q DD E P VF A Q DD E P VF A Q VD K P P F A K VD K P P F A K RDAMMG K NL RDAMMG K K L RDAMG K K L RDAMG K K L RDAMG K NL RDAMG K N	KVEELDAY KVEELDQT KVETIDSK AVEALDQQ 150 NMEEITH SMEEVIH SMEEVIH SMDEVIC SMDEVIC SMDEVIC SMDEVIC SMDEVIC SMDEVIC SMDEVIC SMDEVIC BAVVE DGSYVERIC DGSYVERIC DGSYDENIC QSTYLARMC	FPKBLRI FQGRLRI LCERGNFTI FENRLRI 1000000000 160 11GQLIESNI 11GQLIESNI 11GQLIESNI 11GQLIESNI 11GQLIESNI 11GQLIESNI 11GRALEQLI 11GQKIESNI 11GQKIESNI 11GQKIESNI 11GQKIESNI 11GQKIESNI 11GQKIESNI 11GQKIESNI 11GQKIESNI 11GQKIESNI 11GQKIESNI	RYNN FTAPYN RYN SFTAPYN RYN SFTAPYN RYN SFTAPYN A A A A A A A A A A A A A A A A A A A	TQLD FVKLJC FVKLVDRD. VKVDRD. 180 RDQL0002 180 RDQL0002 180 RDQL0002 180 SNBLIP LAT FDELK FLAD FDELK FLAD FDELK FLAD REQLQ FPAK RQLQ FPAK RQLQ FPAK RQLQ FPAK QGTLN KMAI LNTLK FLSH LSRIR HVH LD FLSF IAE LSRIR HVH LD KLKF ISR YEDLQ FVKL	β7 EEVIDE EVVID EVVID EVVIE EVVE DVVVE EVVE E
htterotrophic	Micr. Nodu Nodu Nosci Raph Cyli Arth Syne Pelo Ccta Gocta Gocta Gocta Soci Losu Sacci	ENGLITACOU RQRIAQUTTAF KQNIIEISQET QQVIITAFLN7 Doystis Laria Do Llatoria ddiopsis ddrospermopsis cospira hhococcus nodesmium lanabaena inctyon lecabacter acter socter socter ifotomaculum haropolyspora	β6 12 9 LASD RV EGD NKMAIEVIENE KPS LSHEVAEGE 12 0 REVSIKIFWDN (REVSIKIFWDN (REVSVKIFWDS) REVSVKIFWDS (REVSVKIFWDS) REVSVKILWDT (REVSVKILWDT) REVSVKILWDT (REVSVKILWDT) REVSVKILWDT) REVSVKILWDT (REVSVKILWDT) VSVKJFWNQ (REVSVKILWDT) VSVKJFWNQ (REVSVKILWDT) VSVKJFWNQ (REVSVKILWDT) VSVKJFWNQ (REVSVKILWDT) VECSVKJFWNQ (REVSVKILWDT) VECSVKJFWNQ (REVSVKILWD) VECSVKJFWNQ (REVSVKILWD) VECSVKJFWNQ (REVSVKILWD) VECSVKJFWND (REVSVKILWD) VECSVKJFWND (REVSVKILWD) VECSVKJFWND (REVSVKI VECSVKJFWND (REVSVKI VFFOR (REVSVKI VECSVKJFWD) REVSVKJFWD (REVSVKI REVSV	SMTDLMIC LLTENMAX VQTEKMIN LLTESMIN SCOLOGO 130 WELQAALES SELQAMMES SELCAMMES SELCAMES SELC	AAYLIP GSFLINK GSFLINK (GSFLID (GSFLID (GSFLID)	NISEBEKESQ DDEPEDEGE DDEPEGE VDKEPEGE VDKEPEGK RDAMMGKNL RDDMBGKL RDDMBGKL RDDMBGKL RDAMGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGKL RDAMGKL RDAMGGKL RDAMGKL	KVEEIDNY KVEELDQT KVEELDQT KVEELDQT SOUDDON SO	FPKB.LLRI FOGR.LLRI LCERGMFTI LCERGMFTI LCERGMFTI ICQLIESCU IIGQLIESCU IIGQLIESCU IIGQLIESCU IIGQLIESCU IIGQLIESCU IIGQLIESCU IIGREIESCU IIGREIESCU IIGREIESCU IIGREKESCU	EYNN FTAPYN EYNN STAPYN EYNN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN SARK STU SARK STU SARK STU SARK STU SER KESTU SER STU SER STU SE	TQLD FVKLJ FVKLJQRD. VKVJDRD. VKVJDRD. SNELPVLJ SNELPVLJ SNELPVLJ SNELPVLJ REELQPFAK REELQPFAK REELQPFAK REELQPFAK REELQPFAK TAALSPLAS QBGLIFLAL TAALSPLAS QBGLIFLAL STARPHVH YEDLRSVCV VKSLGAAAV	β7 EVIIE EVVIE EVVIE EVVIE EVVIE EVVIE EVVIE EVVIE EVVIE EVVIE EVVIE EVVIE EVVIE EVVIE EVVIE EVIE
heterotrophic bacteria	Micr. Nodu Nost. Osci: Raph Cyli: Arth Syne Tric: Pseu Pelo Octa Geob Kori! Desu Sacci Sacci	ENELITACOO ENELITACON RCRIACTTAL KONIIEISOET QUVIITAFLN7 Docystis Latia Joc Llatoria dicospermopsis rospira cospira Latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira latoria dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicosport	β6 β2 βLASD VENE β2 βLASD RVEGD MNMAIEVIENE 12 9 LASD VENE REVSVER R	CALL TENMAX VQTEKMIN LLTESMIN VQTEKMIN LLTESMIN CALQALES (SELQAMMDS) (SELQAMDS) (SELQAMS	AAYLIP GSFLINK GSFLINK (a) (a) (a) (a) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	NIGHE KHSQ DD DFVFAQ DD FVFAQ DD FFFAK DFFFAK PFFAK PFFAK PFFAK PFFAK PFFAK PFFAK PFFAK PFFAK PFFAK PFFAK PFFAK PFFAK PFFAK PFFAC	KVEELDQT KVEELDQT KVEELDQT KVEELDQT SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ	FPKB.LLRI FPGR.LLRI LCERGNFTI LCERGNFTI LCERGNFTI GOLIEBAU IGQLIEBAU IGQLIEBAU IGQLIESAU IGQLIESAU IGQLIESAU IGQLIESAU IGQLIESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU IGQELESAU	K K K K K K K K K K K K K K	TQUD FARIRQQD VKVDRD. VVVDRD. 1800 SNELNPLAT FDELKELAD RNELNVLAV REELQFFAK REELQFFAK REELQFFAK REELQFFAK REELQFFAK LDTLKPLSH LDTLKPLSH LDTLKPLSH LDTLKPLSH LDTLKPLSH VVELLGAAAV VVRSLGAAAV VVRSLGAAAV	β7 EVVIE EVVIE EVVIE EVVIE DVVE DVVE EVIE EVI
photosynthetic bacteria	Micr. Nodu Nosti Syne Tric Pseu Octa Geob Koril Desu Sacci Sacci Strej Micr.	ENELITACOO ENELITACON RORIACYTRAF KONIIEISOET QOVIITAFLN7 Docystis Laria Doc Llatoria dicospira Llatoria dicospira Dicospira Llatoria dicospira Dicospira Bicabacter Socher Lictonaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter Lictomaculum Larotacter	β6 120 120 REVSIKIFWDN. REVSIKIFWDN. REVSVKIFWDS. REVSVKIFWDS. REVSVKIFWDS. REVSVKIFWDS. REVSVKIFWDS. REVSVKIFWDS. REVSVKIFWDS. REVSVKIFWD. REVSVKIFWD. REVSVKIFWD. REVSVKIFWD. TO TO TAMANA REVSVKIFWD. TO TO TAMANA TO TAMANANA TAMAN	CALINE ALL CALLS AND ALL CALL	GSFLINK GSFLINK GSFLINK (GSFLINK (GSFLINK (GSC)	NISEBEKESQ DDEPFAQ DDEPFERE VDKEPFEAK PEFAK PEFAK RDAMMGKNL RDAMGKAL RDAMGKAL RDAMGKAL RDAMGKAL RDAMGKAL RDAMGKAL RDAMGKAL RDALGQPL RDALGQPL RDALGQPL RDALGQPL RDALGQPL RDALGQPL RDALGQPL RDALGQPL RDALGQPL RDSLKQSV KGKILSVPA REALVGKVL RDSLKQSV KGKILSVPA RDSLKQSV KGKILSVPA RDSLKQSV KGKILSVPA RDSLKQSV KGKILSVPA RDSLKQSV KGKILSVPA RDSLKQSV KGKILSVPA RDSLKQSV KGKILSVPA RDSLKQSV KGKILSVPA RDSLKQSV KGKILSVPA RDSLKQSV RDSLKV RD	KVEELDOT KVEIDOT KVETIDSK AVEALDQT 50 NMEEITH SMEEVIH SMEEVIH SMEEVIH SMDEVIC	FPKB.LLRI FPGR.LLRI LCERGNFTI FENR.LLRI GOLTENT FENR.LLRI ICQLIENT ICQLIENT ICQLIENT ICQLIENT ICQLIENT ICQLIENT ICQLIENT ICQLIENT ICQLIENT ICQLIENT ICQLIENT	RYNN FTAPYN RYNN SFTAPYN RYN SFTAPYN RYNDFTAPYN 60 200000000000 170 YLRRKODIIOVF JSARKOSVIEV LLRRKOVIOVF LLRKETVVOAF LDRENEIIOVF LDRENEIITAF LERVENIIOTTA SERSTEYVE SSAKEADSOD	TQLD FARIRQQD VKVDRD. VKVDRD. SNELNPLAT SNELNPLAT FDELKPLAT TAALSPLAS QETLNNMAI TAALSPLAS QEGLIPPAK REELQPFAK REELQPFAK REELQPFAK TAALSPLAS QEGLIPPAL VEDLRGVCV VEDLRGVCV VEDLRGVCV	β7 EEVIEE DVVVEE EVVVVVEE EVVIVE DDVVVVEE EVIE EVI
httirospira Synechocccus Trichodesmium Pseudanabaena photosynthetic bacteria heterotrophic bacteria	Micr. Nodu Nost. Osci Raph Cylin Syne Trici Syne Trici Scob Koril Desu Sacci Strej Micr.	ENELITACUO ENELITACUNE RELATION KONIIEIECE QUVIITAELNI Ocystis Laria Oc Llatoria Ldiopsis Adrospermopsis cospira chococcus ocdesmium lanabaena lictyon lacabacter sacter fotomaculum haropolyspora otomia	β6 120 120 120 120 120 120 120 120	CALL TENMARY VQTEKMIN LLTESMIN (QTEKMIN	A YLIP GSFLINK GSFLIN GSFLIN (G) (G) (G) (G) (G) (G) (G) (G) (G) (G)	NIGHE KHSQ DD D PVFAQ DD PVFAQ VD K PFGE VD K PFGE RDAMMGKNL RDAMMGKNL RDAMMGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDAMGKNL RDALVGKV	KVELIDNY KVELDQT KVETIDSK AVEALDQT SQCQQQQ 150 NMEBIT SSEEVIH SIEEVIH SKEEVIH SMDEVIQ SMDEVIG	FPKB.LLRI FPGR.LLRI LCEGGMFTI LCEGGMFTI LCEGGMFTI ICQLIESCU IGQLIESCU IGQLIESCU IGQLIESCU IGQLIESCU IGQLIESCU IGQLIESCU IGQLIESCU IGQLIESCU IGREISCU IGREISCU IGREISCU IGREISCU IGREISCU IGREISCU IGENUNDAL LGELSCU LGCWVADAL LGENUNDAL LGCUVMAN LGCUVMAN LGELINTY	E Y NN P TA P YN E Y NN P TA P YN E YN S F TA P YN C YN S F TA P YN S F TA P YN C YN S F TA P YN S F TA P YN C YN S F TA P	TQLD FVKLJ FVKLJQRD. VKVJQRD. 180 RDQL0101 SNELP 10 FDELK FLAD FDELK FLAD FDELK FLAD FDELQ FPAK REQLQ PFAK REQLQ PFAK REQLQ PFAK REQLQ FFAK REQLQ FFAK LS FLAS DGGL 11 EAL TAALS FLAS QETLINKMAI LD FLS FLAS QETLINKAI LS FLAS VEDLRGYCV VERFAS CTA VDRFAS CTA	₿7 EEVIIE EEVIIE EEVIVE EEVVVVVE EEVVIE EEVVVVVE EEVIE EEVVVVVE EEVIE EEVVIE EEVIE EEVVIE
heterotrophic bacteria	Micr. Nodu Nost. Nost. Syne Tric: Pseu Pelo Ccta Geob. Koril Desu Sacci Stre Micr.	ENGLITACOUR ENGLIACTINA KONIIEISOF QUVIITAFLN7 Doystis Laria Dodesmium Larobacter Laropolyspora Dolderia Atia Dolderia	β6 12 9 LASD XV END S PLASD XV END S PLASD XV END L PLAND XV END B PLASD XV END 12 0 REVS XI END REVS XI FWDS. REVS XV IFWDS. REVS	SMTDLMIL LLTENMAX VQTEKMIN LLTESMIN SCOLOGO 130 WELQAALES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES (SELQAMMES) (SELQAMMES (SELQAMMES) (SELQAMMES (SELQAMMES) (SELQAMMES (SELQAMMES) (SELQAMMES) (SELQAMMES) (SELQAMMES) (SELQAMMES) (SELQAMES)	A YLIP GSFLINK GSFLINK GSFLINK (a) (a) (b) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	NISE EKESQ EDE FVEAQ DE FVEAQ DE FFAK DE FFAK PEAMMGKNL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMGCKL RDAMLCQ RESLQOVAE RESLQOVAE RESLQOVAE RESLQOVAE RESLQOVAE RDAMLALGG RDAMLALGG RDAMLALGG RDAMLALGG RDAMLALGG RDAMLALGG RDAMC	KVELIDNY KVELDQT KVELDQT KVETIDSK AVEALDQQ iso NMEETH SIEEVTH SHEEVTH SMEEVTH SMDQTIC SMDDVIC SMDDVIC SMDEVIA SMDEVIA SMDEVIA SMDEVIA DKIYQRVK DGSYYERIQ DKIYQRVK DGSYYERIQ DKIYQRVK DGSYYERIQ DKIYQRVK DASHARVR DASHSKIR DASRSKER AVTRGDKIE	FPKB.LLRI FPGR.LLRI LCERGNFTI LCERGNFTI LCERGNFTI LCERGNFTI LCERGNFTI LCERGNFTI SECON LCCERGN LCCL LCELSSNI LCRNFESNI LCRNFESNI LCRNFESNI LCRNFESNI LCRNFESNI LCRNFESNI LCRNFESNI LCRNFESNI	EXIN PTAPYN EXIN STAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CHYN CYMC CHYN	TQLD FARIRQQD VKVDRD. VKVDRD. SNELPVRD. SNELPVRD. SNELPVRD. FDELKPLAD FDELKPLAD FDELKPLAD FDELKPLAD FDELKPLAL TAALSPLAS QBGLIPFAK REELQPFAK REELQPFAK REELQFFAK TAALSPLAS QBGLIPFAK TAALSPLAS USTLKPIST LSRIRPHVH LDKLKPIST LSRIRPHVH LDKLKPIST LSRIRPHVH LDKLKPIST VVSRLGAAAV VVSRLGAAAV VVSRLGAASCTA VVDRFASCTA	β7 EVVIIE EEEEDVVVE EEEEDVVVE EEEEDVVVE EEVISTSSSVD DDVYVIAEISSSVD DDVYVIAEISSSVD DDVYVIAEISSSVD DDVYVIAEISSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSVD EEVISTSSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSS EIVISTSSSS EIVISTSSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSSS EIVISTSSSS EIVISTSSSS EIVISTSSS EIVISTSSS EIVISTSSSS EIVISTSSS EIVISTSSS EIVISTSSSS EIVISTSSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISTSSS EIVISSS EIVISTSSSS
photosynthetic bacteria	Micr. Nodu. Nost. Syne: Trick Syne: Trick Syne: Trick Syne: Ser: Micr. Burk Ser: Halo	ENELITACODE ENELITACONE RCRIACTETA KONIIEISOET QUVIITAFLN7 Docystis Latia Docystis Latia Docystis Latia Dococcus Dideosmium Lanabaena Lanabaenabaena Lanabaena Lanabaenabaenabaena Lanabaenabaenabaenabaenabaenabaenabaenab	β6 β2 βLASD V VENE β PLASD	CA CA CA CA CA CA CA CA CA CA CA CA CA C	GSFLINK GSFLINK GSFLINK (GSFLINK (GSFLINK (GSC)	ANGEBEKESQ DD DFVFAQ DD DFVFAQ DD FEDFGE VDKEPEFAK PERAK RDNMEGKL RDNMEGKL RDSNEGKAL RDAMEGKNL RDAMEGKNL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDAMEGKAL RDSLKGQSV KGKILSSKPE RDSLKGQSV KGKILSSKPE RDSLKGQSV KGKILSSKPE RDSLKGQSV KGKILSSKPE RDSLKGQSV KGKILSSKPE RDSLKGQSV KGKILSSKPE RDSLKGQSV KGKIGSKPE KGKIGSKPE	KVELIDOT KVELDOT KVELDOT KVETIDSK AVEALDQO 150 NMEETH SMEETH SMEETH SMEEVIH SMEVIG SMDEVIG	FPKB.LLRI FPGR.LLRI LCERGNFTI FENR.LLRI UCERGNFTI FENR.LLRI UCQUICOLUS UCQUICOLUS UIGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIESNI UGQLIMVAAI UGQLIMDAI UGQLIMDAI UGRIIEXAI UGRIIEXAI UGRIIEXAI UGRIIEXAI UGRIIEXAI UGRIIEXAI UGRIIEXAI UGRIIEXAI UGRIIEXAI UGRIIEXAI UGRIIEXAI	K K K K K K K K K K K K K K	TQUCCCC TARLOCK TAR	β7 IIIDE EVVIVE EVVIVE EEDVVVE EEDVVVE EEDVVE EEDVVE EEDVVE EEDVVE EEDVVE EEDVIA KRSSVD EEDIS
heterotrophic bacteria	Micr. Nodu Nost. Osci. Raph Cyli. Arth Syne Tric. Peloo Cota Geob Korii Sacc. Strey Micr. Micr. Strey Strey Haloo Haloo	ENELITACUO ENELITACUNE RELATION KONIIEIESE QUVIITASLM? Ocystis Latria oc Llatoria Ldiopsis drospermopsis rospira codesmium Latoria Ldiopsis drospermopsis rospira codesmium Latoria Ldiopsis drospermopsis rospira codesmium Latoria Ldiopsis drospermopsis rospira Latoria Ldiopsis codesmium Latoria Latori	β6 β6 β2 μ Δ Δ Δ Υ ΥΝΕ μ κμα Ι Ε Υ Ι ΕΝΕ μ κμα ι ΕΝΕ κ μ γ Δ ε κ κ μ γ	CALL TENNAR VQTEKMIN LLTESMIN COUCLOS ISO WELQAALES CSELQAMMES CSELQAMMES CSELQAMMES CSELQAMMES CSELQAMMES CSELQAMMES CSELQMMIDS TAELNALMEE QELKLGEE EQELKLGEE EVENHIQE NHI JOELKKLEN TEVENHIGEVITE EVFHEIGET HI IDELKKE EVFHEIGET HI IDELKKE VEVFHEIGET DALLEVUTE PVLNEVLE EVFLRDLEE TPVLRDLEE	GSFLINK GSFLINK GSFLINK (G) (G) (G) (G) (G) (G) (G) (G) (G) (G)	UDE EKESQ DDE EVEAQ DDE EDEGE VDK EPEFAK VDK EPEFAK RDAMMGKNL RDAMMGKNL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL REALVGKV RESLVGVA RESLVGV	KVELIDNY KVELDQT KVETIDSK AVEALDQT SQCQQQQ SQCQQQQQ SQCQQQQQQ SQCQQQQQQQQ SQCQQQQQQQQ	FPKB.LLR FPGR.LLR LCEGGNPT FENRLR LCEGGNPT FENRLR LCEGGNPT FENRLR LCEGGNPT FENRLR CGGNPT IGQLISSNI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI LGELSNE LGCLVMDAT LGELVNEAV LGCUVMAT LGELIANTI LGSRFRVL LGSRFRVI LGSRFRVI LGSRFRVI LGSRFRVI	RYNN FTAPYN RYNN SFTAPYN RYN SFTAPYN RYN SFTAPYN RYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN LRRK OLLOW JERK STEI LORENELLOY MCRK EGWLOYF JERK EAVIOYF MCRKE CAUNON SERSTEY	TQLD FVKLJ FVKLVDRD. VKVDRD. NKUDRD. SNELINEAT SNELINELAT FDELKELAD SNELINELAT FDELKELQEFAK REQLQEFAK REQLQEFAK QGLIPLAL TAALSELAS QETLINKMAI LNTLKELSEN QETLINKMAI LUNILKEISEN USEN USEN VERGUGEFASCTA VERLGEFASCTA VERLGEFASCTA VERLGEFASCTA	β7 UIDEEEEBEBEB EVVIVEVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV
htterotrophic bacteria	Micr. Nodu Nosti Saph Cyli: Arth Syne Trici Pseu Desu Sacci Stre Micr. Pseu Desu Sacci Stre Micr. Micr. Trici Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Sacci Stre Micr. Micr. Not. Sacci Stre Micr. Micr. Micr. Sacci Stre Micr. Micr. Micr. Micr. Sacci Stre Micr. Micr	ENELITACOO ENELITACON RORIACTETA KONIIEISOE Ocystis Laria Docystis Laria Docustis Latria Latoria Lictyon Latoria Lictyon Lacoria Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Lictyon Licty	β6 1200 1	SMTDLMIL LLTENMAX VQTEKMIN LLTESMIN SCOLOGO 130 WELQAALES SELQAMMES SELQAMMES SELQAMMES SELQAMMES SELQAMMES SELQMMIS SELQMMIS SELQMMIS SELQMMIS SELQMMIS SELQMIS SELQMMIS SELQMIS SELQMIS SELQMIS SELQMIS SELQMIS SELQMIS SELQMIS SELQMIS SELQMIS SELQMIS SELQMIS SELQAMES SELQAMES SELQAMES SELQAMES SELQAMES SEL SEL SEL SEL SEL SEL SEL SEL SEL S	A YLIP GSFLINK GSFLINK AYLIP GSFLIN (a) 000000 14 NPKLKQE NDDLKQ NDD NDD NDD NDD NDD NDD NDD ND NDD ND N	NISE BEKESQ DD BEVFAQ DD BEDFGE VDK BPEFAK VDK BPEFAK RDAMMGKNL RDAMMGKNL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDAMGGKL RDALVGKU RDELLVGKU RESLQUVAE RESLQUVAE RESLQUVAE RESLQUVAE RESLQUVAE RDAILGSKALGG LPEAIGRR.	KVELIDNY KVELDQT KVELDQT KVETIDSK AVEALDQQ iso NMEEITH SIEEVTH SIEEVTH SMEEVTH SMDQIIC SMDQIC SMDQIC SMDQVIC SMDQVIC SMDQVIC SMDQVIC SMDQVIC SMDQVIC SMDQVIC DGSYYERIC DGSYYERIC DGSYYERIC DGSYYERIC DGSYYERIC DGSYYERIC DGSYYERIC SMDQVIC DGSYYERIC SMDQVIC DGSYYERIC SMDQVIC DGSYYERIC SMDQVIC DGSYYERIC SMDQVIC DGSYYERIC SMDQVIC DGSYYERIC SMDQVIC DGSYERIC SMDQVIC 	FPKB.LLRI FPKGR.LLRI LCERGMFTI LCERGMFTI LCERGMFTI LCERGMFTI LCERGMFTI LCERGMFTI LCERGMFTI FBNR.LLRI IGQLIESNI IGQLIESNI IGQLIESNI IGQLIESNI IGREISSNI IGREISSNI IGREISSNI IGREISSNI LGERWADAL LGERWADAL LGERWADAL LGELWMAN LGELWAN LGELWAN LGELWAN LGELWAN LGELMANTY LGELWAN	EYNN PTAPYN EYNN STAPYN EYNN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN	TQLD FVKLJ FVKLJQDD. VKVDRD. 1800 REOGINHRAQ SINELNPLAT FDELKPIAN REELQPFAK REQLQPFAK REQLQPFAK REQLQPFAK REQLQPFAK REQLQPFAK REQLQPFAK REQLQPFAK REQLQPFAK REQLQPFAK REQLQPFAK SINELNPLAT DGLIPFAK TAALSPLAS VUDGFASCAA VUDGFASCAA VUDRFASCAA VUDRFASCAA VUDRFASCAA VUDRFASCAA VUDRFASCAA VUDRFASCAA VUDRFASCAA VUDRFASCAA VUDRFASCAA VUDRFASCAA	β7 IIIDEEEE VIVIVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV
heterotrophic bacteria haloarchaea	Micr. Nodu Nosti Osci Raph Cyli Arth Syne Trici Pseu Octa Geob Kori Desu Sacci Stre Burki Serr. Halo Halo Halo	ENELITACOUR ENELITACON RCRIACTTAL KONIIEISOET QOVIITATIN Laria Docystis Laria Docystis Laria Docystis Laria District Laria Docystis Docystis Laria Docystis Laria Docystis Laria Docystis Laria Docystis Docystis Laria Docystis Laria Docystis Docystis Laria Docystis Docyst	β6 12 9 LASD Y VENE S PLASD Y VENE S PLASD Y VENE L S PLASD Y VENE 12 9 LASD Y VENE R S VALLENE R V S V VI L V VALLENE R V S V VI F VDS. I R V S V VI F WDS. I R V S V VI F WDS. I R V S V VI F WDS. I R V S V VI L WDT. I R V S V VI L WDT. I R V S V V F WNE T. 7 R V S V V F WNE T. 7 R V S V V V F W NE T. 7 R V S V V V F W NE T. 7 R V S V V V F W NE T. 7 R V S V V V S T. 7 V E L G V V V S T. 7 V E L G V V I I T S	SMTDLMIL LLTENMAX VQTEKMIX LLTESMIX VQTEKMIX COCCOCCC SGLQAMMES GELQAMMES GELQAMMES GELQAMMES GELQAMMES CAELQALMES CAELQALMES CAELQALMES CAELNAVAM TRIVENICA MIFGEVLRE FELNLAVAM FEVLNEVLTE FELNLAVAM FFIYLNICK SHICKLEN FFILLANNE FFILLANNE FFILLANNE FFILLANNE FFILLANNE FFILLANNE FFILLEKK FVLMEVLTE FVLMEVLTE SAVDVLFTEG	A YL IP GSFLINK GSFLINK GSFLINK (a) (a) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	NISEBEKESQ DD B PVFAQ DD B PVFAQ ND B PVFAQ VD K PPFAK ND K PFFAK ND K PFFAK RD NMEGRKL RD NMEGRKL RD NMEGKNL RD AMEGKNL RD AMEGKNL RD AMEGKNL RD AMEGKNL RD AMEGKNL RD AMEGKNL RD AMEGKNL RD AMEGKNL REALVGKVL REAL	KVELIDNY KVELIDQT KVETIDSK AVEALDQT SOUCOOCOO SOUCOOCOO SOUCOOCOO SOUCOOCOO SUBJECT SU	FPKB.LLRT FPGR.LLRT LCERGNFTI LCERGNFTI LCERGNFTI FENR.LLRT UCQUICOLUE IGQUIESNI IGQUI	EYNN FTAPYN EYNN STAPYN EYNN SFTAPYN CRYN SFTAPYN CYN SFTAPYN CYN SFTAPYN CYN SFTAPYN CYN CYN SFTAPYN CYN SFTAPYN CYN	TQLD FARIRQQD VKVDRD. VKVDRD. VKVDRD. SNELDVLAT FDELKPLAD FDELKPLAD FDELKPLAD FDELKPLAD TFDELKPLAN FDELKPLAN FDELKPLAN FDELKPLAN FDELKPLAN FDELKPLAN FDELKPLAN TALSPLAS QDGLIFLAL SRIR HVH LDKLKFISR VVBLASCAA VVBLASCAAN	β7 UIIDEE EVVIVVVEE EVVVVEE EVVVVEE EVVVVEE EVVVEE EVVVEE EVVVEE EVVE
heterotrophic bacteria haloarchaea	Micr Nodu Nost Osci Raph Cyli Arth Syne Seob Kori Sacc Stre Micr Mycol Sacc Stre Micr Halo Halo Halo Halo	ENELITACOO ENELITACON RCRIACTTAL KONIIEISOET QOVIITALLNT Docystis Laria Docustis Laria Dococcus Dilatoria dicospermopsis rospira Dicoccus Dianabaena lictyon danabaena dicabacter dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis rospira dicospermopsis dicospermopsis rospira dicospermopsis dico	β6 β2 βLASD XVENE SPLASD XV	CALL TENMAX VQTEKMIX LLTESMIX VQTEKMIX LLTESMIX COCOCOCO 130 WELQAALES CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMMDS CSELQAMDS CSELCAMDS CSELC	GSFLINK GSFLINK GSFLINK GSFLINK (a) (a) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	ANGEBEKESQ EDDEPVEAQ DDEPVEAQ DDEPEFAK PEFAK PEFAK PEFAK PEFAK PEFAK PEFAK PEFAK PEFAK RDAMEGKNL	KVEEIDNY KVEEIDQT KVETIDSK AVEALDQO 150 NMEEITH SIENTH SIENTH SMEETH SMEETH SMEETH SMDQIIQ SMDEVIG SMDEV	FPKB.LLRI FPGR.LLRI LCERGNFTI FENN.LLRI UCERGNFTI FENN.LLRI UCQUICQUIC IGQUIESNI UCQUI	RYNN FTAPYN RYNN SFTAPYN RYN SFTAPYN RYN SFTAPYN RYN SFTAPYN RYN SFTAPYN RYN SFTAPYN RYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN CRYN SFTAPYN SFTAPYN SFTAPYN LRRK STAPYN LLRRK STU LLRKK STU SFTAPYN MCRK SFTAPYN LLRRK STU SFTAPYN LLRRK STU SFTAPYN SFTAPYN LLRRK STU SFTAPYN SFTAPYN LLRRK STU SFTAPYN S	TQUCUD FARIRQQD VKVDRD. VKVDRD. SNELNPLAT SNELNPLAT FDELKELNPLAT FDELKELAD FNELNVLAV COGGLIPLAL TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS QETLINKAI TAALSPLAS COGGLISELSI TAQLSSUS TAQUS TAQLSS	β7 IIIDEEE VIIIDEEE VVVIVE VVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVVE VVVE VVVVE VVVE VVVE VVVE VVVVE VVVVE VVVE VVVE VVVE VVVE VVVE VVVE VVVE VVVE VVVE VVE
heterotrophic bacteria	Micr. Nodu Nosti Syne Trici Syne Trici Syne Trici Socta Gotta Serri Micr. Pseu Secta Serri Micr. Secta Serri Haloo Haloo Haloo Haloo	ENELITACUO ENELITACUNE RERIACTERA KONIIEIECE QUVIITAELNY Ocystis Laria oc Llatoria diopsis adrospermopsis cospira cospira cospira cospira cospira cospira cospira cospira codesmium lanabaena lictyon lecabacter sacter fotomaculum haropolyspora otoerium colderia atia sacterium for acterium for	β6 120 120 120 120 120 120 120 120	SMTDLMIC LLTENMAX VQTEKMIX LLTESMIX SCOLOGO 130 WELQAALES SELQAMMES SELQAMMES SELQAMMES SELQAMMES SELQAMMIS SELQMMIDS SELQMMIDS SELQMMIDS SELQMMIDS SELLAVME VELALMES VELCALMES SELQMMIDS SELENAVAE WIFGEVLRE WIFGEVLRE WIFGEVLRE WIFGEVLRE VIEVENEIGET SALLESVTE SALLESVTE SALLESVTE STELLESVTE SALLESV	A YLIP GSFLINK GSFLINK GSFLID (G) 000000 14 NPKLKQE NDDLKQQ NDD NDC NDQL ND ND ND ND ND ND ND ND ND ND ND ND ND	NISE EK ES Q DD D PVF AQ DD D PVF AQ DD PF DF GE VD K PF FAK VD K PF FAK RDAMMSKNL RDAMMSKNL RDAMMSKNL RDAMSKKL RDAMSKKL RDAMSKKL RDAMSKKL RDAMSKKL RDAMSKKL RDALVSKVL RDAMSKKL RDALVSKVL RDALVSKVL RDALVSKVL RDALVSKVL RDALVSKVL RDSLLKOSVP HQEITRKHL RDSLLKOSVP HQEITRKHL RDSLLKOSVP HQEITRKHL RDSLLKOSVP HQEITRKHL RDSLLKOSVP HQEITRKHL RESLQOVAE	KVELIDNY KVELIDQT KVETIDSK AVEALDQT SQCQQQQ 150 NMEBIT SIESVIH SIESVIH SIESVIH SMDEVIC	FPKB.LLRI FPGR.LLRI LCEGGMFTI FENRLRI LCEGGMFTI FENRLRI UCCGGMFTI FENRLRI UCCGGMFTI FENRLRI UCCGMFT IGQLISSNI UCCLIESN	E Y NN PT A P YN E YN N S T A P YN E YN S F T A P YN C YN C YN C YN C YN C YN D T A YN C YN C YN C YN D T A YN C YN C YN C YD A F YN C YN C YN C YN C YN D YN C YN C YN C YN C YN D YN C YN C YN C YN C YN D YN C YN C YN C YN C YN C YN C YN D YN C YN C YN C YN C YN C YN C YN	TQLD FVKLJ FVKLJRQ.D VKVJRD.J. 180 RDQUUDD 180 RDQUUDD 180 RDQUUNHRAQ SNELP LAT FDELK PLAD FDELK PLAD FDELK PLAD FDELK PLAD FDELQ PFAK REQUQ FFAK REQUQ FFAK REQUQ FFAK REQUQ FFAK REQUQ FFAK UDGLI LAL TALLS PLAS QETLINKAI LUNTLK PLSH QETLINKAI LUNTLK PLSH QETLINKAI SIGNA VDGFASCIA VDGFASCIA VDGFASCIA SIGNA SIG	β7 IIIIE DE BE BE BE BE VIVIVINANT VIVINAN
heterotrophic bacteria haloarchaea	Micr. Nodu Nosti Sari Pseu Pelo Koril Desu Sacci Strej Micr. Pseu Sacci Strej Micr. Nathar Sacci Strej Haloo Haloo Halao	ENELITACUNE RQRIAQUTTAFA KQNIIEIFQET QQVIITAFLN7 Doystis Laria Do Llatoria ddiopsis ddrospermopsis rospira hhococcus nodesmium lanabaena hhococcus nodesmium lanabaena ilictyon lecabacter acter sacter sacter idcomaculum ictyon lecabacter acter sacter into a piora basterium ferax rubrum nuadratum kalicoccus soccus pranum inema Diforma fivar	β6 12 9 LASD RV EGD S PLASD RV EGD M KMAIEVIENE M KMAIEVIENE 12 9 LSHEYAEGE 12 9 REVSIKIFWDN (REVSIKIFWDN (REVSVKIFWDN (REVSVKIFWD (REVSVKIF	SMTDLMIL LLTENMAX VQTEKMIN LLTESMIN SCOLOGO SCOLOGO SCOLOANS SCELQAMMES SCELQAMES	AAYLIP GSFLINK GSFLINK AYLIP GSFLIN (a) (a) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	NISE EKESQ EDE FVFAQ DDE FVFAQ DDE EDFGE VDKEPEFAK VDKEPEFAK RDAMMGKNL RDDMBGKL RDDMBGKL RDDMBGKL RDAMGGKNL RDAMGGKNL RDAMGGKNL RDAMGGKNL RDAMGGKNL RDAMGGKNL RDAMGGKNL RDAMGGKNL RDALGQPL RDELLQGVAE RESLQOVAE RESLQOVAE RESLQOVAE RESLQOVAE RESLQOVAE RDALGGR RESLQOVAE RDALGGR RDALGGR COVAE RDALGGR RDALGGR COVAE RDALGGR RESLQOVAE RDALGGR RDALGGR RDALGGR COVAE RDALGGR RDALGR RDALGGR RDALGGR RDA	KVERIDNY KVELDQT KVELDQT KVETIDSK AVEALDQQ 150 NMEETH SIEEVTH SMEEVTH SMDQTIC SMDBVIC SMDBVIC SMDBVIC SMDEVIC SMDEVIC DSYJERT SYJERT SYJE	FPKB.LLRI FPKGR.LLRI LCERGMFTI LCERGMFTI LCERGMFTI LCERGMFTI LCERGMFTI LCERGMFTI FBNR.LLRI UCQLIESNI UCQLI	E Y NN PT A P YN E YN N S T A P YN E YN S F T A P YN CRYN C F T A P YN C F T A F T A P YN C F T A F T A P YN C F T A F	TQLD FVKL FVKLQQD VKVDRD. VKVDRD. SNELPVC SNELPVLAT FDELKPLAT FDELKPLAT FDELKPLAT REELQPFAK REELQPFAK REELQPFAK REELQPFAK REELQPFAK REELQPFAK REELQPFAK REELQPFAK SNELPVLAT DGGLIFLAL TAALSPLAS VCDGGLIFLAL TAALSPLAS VVDGFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA VVDRFASCAA SUCCOMPACTOR SUCCOM	β7 VIDE BE BE VIDE VIDE VIDE VIDE VIDE VIDE VIDE VIDE
heterotrophic bacteria haloarchaea	Micr. Nodu Nosti Nosti Sarch Syne Trici Pseu Octa Sacci Stre Micr. Pseu Octa Sacci Stre Micr. Nodu Sacci Sac	ENELITATO ENELITATO RELACTORAL KONIIEISOF QUVIITATIN Laria ocystis Laria ocospira llatoria dicospermopsis ocospira dicospermopsis cospira lanabaena lictyon lecabacter acter becabacter acter bictyon lecabacter acter bictyon lecabacter acter bictyon lecabacter acter bictyon lecabacter acter bictyon lecabacter acter bictyon lecabacter acter bictyon lecabacter acter bictyon lecabacter acter bispora boacterium coccus pranum lictyon lictyon lecabacter acter bispora boacterium lictyon lictyon lictyon lictyon lecabacter acter bispora boacterium lictyon	β6 12 9 LASD Y VENE S PLASD Y VENE S PLASD Y VENE NEMAI E VIENE L S PLASD Y VENE REVS VIENE REVS VIENE RE	SMTDLMIC LLTENMAX VQTEKMIN LLTESMIN SCOLOGO SCOLOGO SCOLOANMES CALQANMES CALQAMES CALQAMMES CALQAMES CALQ	AAYLIP GSFLINK GSFLINK GSFLINK (a) (a) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	NIGHE KHSQ DD D PVPAQ DD D PVFAQ DD PFFAK DD PFFAK PFGE RD AMMGKNL RD AMGKNL RD AMGGKAL RD AM	KVELIDNY KVELDQT KVETIDSK AVEALDQT SO SSE SSE SSE SSE SSE SSE SSE SSE SSE	FPKB.LLRI FPGR.LLRI LCERGNFTI FENR.LLRI UCERGNFTI FENR.LLRI UCQLIEST IGQLIEST IGQLIEST IGQLIEST IGQLIEST IGQLIEST IGQLIEST IGQLIEST IGQLIEST IGREI IGREIEST IGREI IGREIEST IGREI IGR	RYNN FTAPYN RYNN SFTAPYN RYN SFTAPYN RYN SFTAPYN RYN SFTAPYN RYN DFTAPYN CRYN SFTAPYN RYN DFTAPYN LRYR SFTAPYN LRRK COLLOW LLRRK COLLOW LLRRK COLLOW LLRRK STIOV LLRRK STIOV STION STICK STION STICK S	TQLD FARIRQQD VKVDRD. VKVDRD. VKVDRD. SNELNPLAT FDELKELAD FDELKELAD FDELKELAD FDELKELAD FDELKELAD FDELKELAD FDELKELAD FDELKELAD FDELKELAD COCCUP FDELKELAD FDELKELAD UR TALSPIA USTIL VINTLKPISH LD ILSPIA USTIL VINTLKPISH LD ILSPIA USTIL VINTLKPISH USTIL VINTLKPISH USTIL VINTLKPISH USTIL VINTLKPISH USTIL VINTLKPISH USTIL SNIR VVRSLGAAAV VVRSLGAAAV VVRSLGAAAV VVRSLGAAAV VVRSLGAAAV VVRSLGAAAV VVRSLGAAAV VVRSLGAAAV VVRSLGAAAV VDRISSUSI TAQUSELSUSI TAQUSELSVSI TAQUSELSVSI CORELPEITAA	β7 VIJUE EEUVIVE EEUVIVE EEUVIVE EEUVIVE EEUVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE EEVIVE KRATSRA AAVEN NEFRE KANNE KA

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 $\alpha 4$, $\alpha 5$ 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/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 $72 \times 34 \times 35$ 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 $\alpha 1-\alpha 7$ and strands $\beta 1-\beta 9$ (Fig. 1*c*). According to the SCOP protein classification, GvpF belongs to the $\alpha+\beta$ 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 ($\beta 1-\beta 5$) is sandwiched by two layers of α -helices ($\alpha 1-\alpha 3$; Fig. 1b). It adopts a $\beta-\alpha-\beta-\alpha$ topology [($\beta 1-\beta 3$)- $\alpha 1-(\beta 4-\beta 5)-(\alpha 2-\alpha 3)$; Fig. 1c]. The central five-stranded β -sheet, which is divided into two parts by helix $\alpha 1$, consists of a core three-stranded motif adopting an antiparallel topology ordered $\beta 1-\beta 3-\beta 2$ and two small sequential strands ($\beta 4$ and $\beta 5$) (Fig. 1b). Owing to the long $\beta 1-\beta 2$ loop, the core three-stranded motif forms a split β -loop- β motif mostly similar to

the split $\beta - \alpha - \beta$ motif, which consists of a variation of the ordinary $\beta - \alpha - \beta$ 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 $\alpha 1$ and the other is comprised of consecutive helices $\alpha 2$ and $\alpha 3$, flank either side of the β -sheet (Fig. 1b). 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 $\alpha + \beta$ class in the SCOP database and thus represents a novel fold.

The C-domain exhibits a two-layer sandwich architecture in which strands $\beta 6-\beta 9$ assemble into a single four-stranded antiparallel β -sheet, whereas helices $\alpha 4-\alpha 7$ gather on one side (Figs. 1b and 1c). 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 $\beta 6-(\alpha 4-\alpha 6)-\beta 7$ and the other of $\beta 8-\alpha 7-\beta 9$. Generally, the doublestranded 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 ($\alpha 4$, $\alpha 5$ and the N-terminal segment of $\alpha 6$), representing a modified ferredoxin fold. In particular, helices



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.

 α 4 and α 5 are approximately perpendicular to each other and antiparallel to the curved long helix α 6.

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 (α 1) and Leu234



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 μ g, lane 2) and cell lysates (30 μ g, 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 μ g, lane 2) and cell lysates (208 μ g, 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.

(C-tail), Leu63 (α 1) and Ala263 (C-tail), Pro76 (β 4) and Pro237 (C-tail), Pro76 (β 4) and Ala241 (C-tail), and Leu104 (α 3) and Leu243 (C-tail). Besides, the interface is further fixed by two hydrogen bonds, one of which is donated by the sidechain carboxyl group of Glu60 (α 1) 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 (α 1) 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 (β 7) and Phe240 (C-tail) and between Leu203 (β 8) 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 (β 6) and the main-chain carboxyl group of Asp233 (C-tail), between the side-chain carboxyl group of Glu113 (β 6) and the main-chain amide group of Thr239 (C-tail) and between the side-chain carboxyl group of Glu113 (β 6) 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. 3a).

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. 4a), 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;

research papers

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 formimidovl transferasecyclodeaminase (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-methylglutarylcoenzyme 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 ($\alpha 2$ and $\alpha 3$). 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 Ndomain has a $\beta 1-\beta 2-\beta 3-\alpha 1-\beta 4-\beta 5-\alpha 2-\alpha 3$ topological connection (Figs. 1*b* and 1*c*), distinct from the top hit with a $\beta 1-\alpha 1-\beta 2-\beta 3-\beta 4-\alpha 2-\beta 5$ 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 Cdomain 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. 4b), the C-domain has an extension region, making it a modified ferredoxin fold (Fig. 4b).

The extension region represents a topology of helix hairpins and a coiledcoil-like fold, which is composed of three helices: helices $\alpha 4$ and $\alpha 5$ in addition to the N-terminal segment of helix $\alpha 6$ (Met150–Lys168). Helices $\alpha 4$ and $\alpha 5$ are kinked at Asn133 by approximately 90°. Helix $\alpha 5$ is antiparallel to the long helix $\alpha 6$, which is bent at Lys168 (Figs. 4b and 4c).

A DALI search using this extension region enabled us to identify a highly homologous segment of Saccharomyces cerevisiae ubiquinolcytochrome c reductase complex 17 kDa protein (Fig. 4c), 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. 3b), 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. 5a and 5b). 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. 5a and 5b). The outer wall represents regularly spaced ribs running nearly perpendicular to the long axis of the vesicle (Fig. 5b). 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. 5c, 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. 5d, 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. 6a). 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. 6b), 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. 6c). 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. 6d). Overall, we demonstrated that GvpF is indeed a constitutive component of the gas vesicle and is most likely to be localized on the gasfacing 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. D66, 213-221.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D67, 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. D66, 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. D66, 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.* D67, 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. D67, 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.