

Acta Crystallographica Section D Biological Crystallography ISSN 0907-4449 Editors: E. N. Baker and Z. Dauter

# Structure of a thermophilic cyanobacterial $b_6 f$ -type Rieske protein

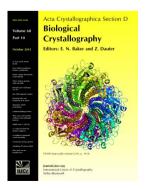
### Sebastian Veit, Kazuki Takeda, Yuichi Tsunoyama, Dorothea Rexroth, Matthias Rögner and Kunio Miki

Acta Cryst. (2012). D68, 1400-1408

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see http://journals.iucr.org/services/authorrights.html



Acta Crystallographica Section D: Biological Crystallography welcomes the submission of papers covering any aspect of structural biology, with a particular emphasis on the structures of biological macromolecules and the methods used to determine them. Reports on new protein structures are particularly encouraged, as are structure–function papers that could include crystallographic binding studies, or structural analysis of mutants or other modified forms of a known protein structure. The key criterion is that such papers should present new insights into biology, chemistry or structure. Papers on crystallographic methods should be oriented towards biological crystallography, and may include new approaches to any aspect of structure determination or analysis. Papers on the crystallization of biological molecules will be accepted providing that these focus on new methods or other features that are of general importance or applicability.

## Crystallography Journals Online is available from journals.iucr.org

### research papers

Acta Crystallographica Section D Biological Crystallography

ISSN 0907-4449

#### Sebastian Veit,<sup>a</sup> Kazuki Takeda,<sup>b</sup> Yuichi Tsunoyama,<sup>c</sup> Dorothea Rexroth,<sup>a</sup> Matthias Rögner<sup>a</sup>\* and Kunio Miki<sup>b</sup>\*

<sup>a</sup>Plant Biochemistry, Ruhr University Bochum, 44780 Bochum, Germany, <sup>b</sup>Department of Chemistry, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan, and <sup>c</sup>Division of Biology, Radioisotope Research Center, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan

Correspondence e-mail: matthias.roegner@rub.de, miki@kuchem.kyoto-u.ac.jp

© 2012 International Union of Crystallography

# Structure of a thermophilic cyanobacterial $b_6 f$ -type Rieske protein

The 'Rieske protein' PetC is one of the key subunits of the cytochrome  $b_6 f$  complex. Its Rieske-type [2Fe-2S] cluster participates in the photosynthetic electron-transport chain. Overexpression and careful structure analysis at 2.0 Å resolution of the extrinsic soluble domain of PetC from the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1 enabled in-depth spectroscopic and structural characterization and suggested novel structural features. In particular, both the protein structure and the positions of the internal water molecules unexpectedly showed a higher similarity to eukaryotic PetCs than to other prokaryotic PetCs. The structure also revealed a deep pocket on the PetC surface which is oriented towards the membrane surface in the whole complex. Its surface properties suggest a binding site for a hydrophobic compound and the complete conservation of the pocket-forming residues in all known PetC sequences indicates the functional importance of this pocket in the cytochrome  $b_6 f$  complex.

#### 1. Introduction

Rieske proteins are essential subunits of all classes of cytochrome *bc*-type complexes and play a key role in photosynthetic and respiratory electron-transport processes (see, for example, Link, 2001). The Rieske protein PetC is a subunit of the cytochrome  $b_6f$  complex and contains a 2Fe–2S cluster with a midpoint redox potential of about +300 mV (see, for example, Bernát & Rögner, 2011; Baniulis *et al.*, 2008; Kallas, 2011; Crofts *et al.*, 2000). PetC transfers one electron from plastoquinol to the cytochrome *f* subunit of the complex. Participation of the Rieske 2Fe–2S cluster in the accompanying proton transfer has also been suggested (Crofts *et al.*, 2000).

As in other cyanobacteria, the cytochrome  $b_6 f$  complex in Thermosynechococcus elongatus BP-1 (T. elongatus) participates in both respiratory and photosynthetic electron transport as the central intersection point between the two processes. Its significance for regulation is also indicated by the fact that multiple petC genes have been reported for several cyanobacterial strains (Schneider et al., 2002, 2004; Yan & Cramer, 2003; for a review, see Schneider & Schmidt, 2005). This multiplicity (up to four petC genes) enables adaptation to stress conditions such as high light (Tsunoyama et al., 2009), low oxygen (Summerfield et al., 2008) or the high demand for reduction equivalents during nitrogen assimilation (Schneider & Schmidt, 2005; Arnold, 2001). In contrast, T. elongatus contains only a single petC gene coding for PetC1, in analogy to all photosynthetic eukaryotes (Schneider et al., 2004). Structurally, all PetC proteins consist of an extrinsic soluble domain which is connected to an N-terminal Received 10 May 2012 Accepted 31 July 2012

PDB Reference: TePetC, 3azc

Printed in Singapore - all rights reserved

transmembrane helix by a flexible hinge. Although crystal structures of both prokaryotic and eukaryotic cytochrome  $b_6 f$ complexes have been analyzed up to about 3 Å resolution (Baniulis et al., 2009; Yamashita et al., 2007; Kurisu et al., 2003; Stroebel et al., 2003), this resolution is not sufficient to visualize details such as the positions of water molecules, which may be important for modelling functional mechanisms on a molecular level. In contrast, high-resolution structures (1.9 Å) are available for the soluble domain of cytochrome ffrom both prokaryotic and eukaryotic sources (Carrell et al., 1999; Chi et al., 2000; Martinez et al., 1996). This high resolution revealed a buried chain of five water molecules with possible relevance for proton translocation. In the case of PetC, a 1.83 Å resolution structure is available of the extrinsic C-terminal domain of spinach Rieske protein (Carrell et al., 1997). However, a high-resolution structure of cyanobacterial PetC is still lacking.

Here, we report detailed spectroscopic and crystallographic analyses of the soluble domain of PetC from the thermophilic cyanobacterium *T. elongatus* (Nakamura *et al.*, 2002; Yamaoka *et al.*, 1978) which has been heterologously expressed. The high resolution of these crystallographic data revealed the positions of structurally important internal water molecules and enabled their comparison with the corresponding water molecules in the eukaryotic structure from spinach. We also present for the first time a putative ligand-binding pocket on the surface of PetC from *T. elongatus*. Conservation of this pocket in all other available PetC structures suggests a new regulatory mechanism for the cytochrome  $b_6 f$  complex.

#### 2. Materials and methods

# 2.1. Cloning, expression and purification of recombinant TePetC

The gene fragment encoding the soluble domain (residues 54-180) of the Rieske protein of T. elongatus (TePetC) was amplified from T. elongatus genomic DNA by PCR using the primers ATCTGTGGTCTCTGCGCGCGGGTTGCCAAGG-ATGCTTT and ATCGCCGGTCTCTTATCATTACGTCC-ACCAAGGTTCTTTG to generate BsaI sites at the 5'- and 3'-termini. Using these BsaI restriction sites, the PCR fragment was inserted into pASK-IBA37plus (IBA GmbH) to generate the respective expression plasmid. The recombinant protein was designed to have 16 additional residues (MAS-RGSHHHHHHIEGR) at the N-terminus with a His<sub>6</sub> tag and a factor Xa restriction site. Transformed Escherichia coli cells (Rosetta, Novagen) were grown at 310 K in LB medium containing 100 µg ml<sup>-1</sup> ampicillin to an optical density  $(OD_{600})$  of 0.6. Overexpression was then induced by 200 µg l<sup>-1</sup> anhydrotetracycline and after 4-5 h cultivation cells were harvested by centrifugation (4000g for 30 min and 277 K). All further steps were performed at 277 K. Cell pellets were suspended in buffer A (50 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 100 mM NaCl), pelleted by centrifugation (see above) and suspended in buffer B (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8.0, 300 mM NaCl, 10 mM imidazole) supplemented with protease inhibi-

tors [1:100; protease-inhibitor cocktail (EDTA-free), Nacalai Tesque] and lysozyme  $(1 \text{ mg ml}^{-1})$ . After incubation for 30 min on ice, the cells were lysed by sonication. The cell lysate was centrifuged (14 000g for 30 min at 277 K) and the filtered supernatant was loaded onto an Ni-NTA column (Qiagen; 2.5 ml Ni-NTA resin per litre of culture) equilibrated with buffer B. A dark brown colour of the column indicated binding of recombinant Rieske protein to the Ni-NTA resin. After washing with 25 ml buffer C (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8.0, 300 mM NaCl, 20 mM imidazole) per litre of resin, bound protein was eluted with buffer D (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8.0, 300 mM NaCl, 250 mM imidazole), which was immediately replaced by buffer F (10 mM Tris-HCl pH 7.3) via ultrafiltration (Amicon Ultra 10K). As several trials failed to remove the His<sub>6</sub> tag by digestion with factor Xa protease, Histagged protein variants were used for all experiments.

Recombinant TePetC was further purified by gel-filtration chromatography using a Superdex 75 column (GE Healthcare) equilibrated with buffer E (10 mM Tris–HCl pH 7.3, 150 mM NaCl). The main peak fractions containing monomeric TePetC were collected and the buffer was replaced by buffer F (see above). The concentrated protein solution (54 mg ml<sup>-1</sup> protein concentration) was frozen in liquid nitrogen and stored at 193 K.

#### 2.2. Spectroscopic measurements

The absorption spectra of oxidized [by addition of 5 mM K<sub>2</sub>Fe(CN)<sub>6</sub>] or reduced (by addition of 5 mM DTT) TePetC (0.7 mg ml<sup>-1</sup>) in buffer *F* were recorded at room temperature (298 K) in a V-630 spectrometer (Jasco).

Circular-dichroism (CD) spectra were recorded in a quartz cuvette with 1.0 mm path length at 293–373 K between 190 and 250 nm in 1 nm increments (bandwidth of 2 nm) using a J-715 CD spectropolarimeter (Jasco). The spectra represent averages of three scans using TePetC (~0.24 mg ml<sup>-1</sup>) in 5 mM potassium phosphate buffer pH 7.5. The thermal melting point ( $T_m$ ) of the protein was estimated by plotting the CD data at 222 nm against temperature followed by fitting to a Boltzmann sigmoidal curve between 348 and 367 K.

#### 2.3. Crystallization and structure determination

The sitting-drop vapour-diffusion method was used with 54 mg ml<sup>-1</sup> TePetC in buffer *F* and reservoir solution consisting of 18%(*w*/*v*) PEG 3000, 9%(*v*/*v*) 2-propanol, 18.5%(*v*/*v*) glycerol and 90 m*M* Na HEPES pH 8.0 in a 2:1 ratio. Crystallization drops were equilibrated against 500 µl reservoir solution at 293 K, yielding hexagonal crystals (maximal dimensions  $0.5 \times 0.5 \times 0.3$  mm) within one week. All diffraction experiments were performed on BL41XU at SPring-8 (Harima, Japan) using an MX225HE detector (Rayonix), a beam size of 50 × 50 µm and an original photon flux of 4 × 10<sup>11</sup> photons s<sup>-1</sup> at the sample position. Data collection was performed using 0.65 Å X-rays with reduction of the initial flux density to 1/25 by an aluminium attenuator of 3.0 mm thickness using helium gas at 40 K for cooling. 180 frames of diffraction images were collected with an oscillation

#### Table 1

Crystallographic and refinement statistics.

Values in parentheses are for the highest resolution shell.

Crystal data	
Space group	P65
Unit-cell parameters (Å)	a = 91.3, c = 43.3
Resolution range (Å)	25-2.0 (2.13-2.00)
Reflections (total/unique)	98082/12341
Multiplicity	7.9 (5.6)
Completeness (%)	87.4 (57.1)
$\langle I/\sigma(I) \rangle$	32.2 (2.1)
$R_{\text{merge}}$ † (%)	7.6 (47.1)
Refinement	
Residues in asymmetric unit	133
Heterogeneous molecules	$1 \times 2$ Fe–2S, $98 \times H_2O$
Total atoms	1134
$R_{\text{work}}$ $\ddagger/R_{\text{free}}$ (%)	19.7/23.6
R.m.s.d. bonds (Å)/angles (°)	0.009/1.5
Average B factor $(Å^2)$	
Protein	57.3
Ligands [2Fe–2S]	48.4
Water/Wat202	66.6/40.6
Ramachandran statistics of $\varphi/\psi$ angles	
Residues in favoured regions	126 [96.2%]
Residues in allowed regions	4 [3.1%]
Outliers	1 [0.8%]
PDB code	3azc

 $\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl). \ddagger R_{\text{work}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|. \$ R_{\text{free}}$  was calculated using 5% of the reflections that were not included in the refinement as a test set.

#### Table 2

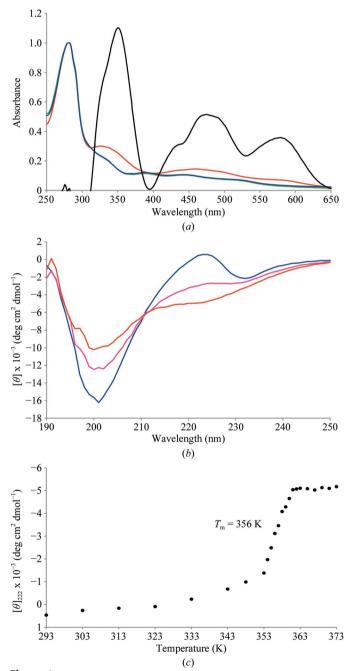
Crystallographic statistics of data sets collected at two different wavelengths near the iron K absorption edge.

Values in parentheses are for the highest resolution shell.

	High energy	Low energy
Data collection		
Wavelength (Å)	1.730	1.750
X-ray energy (eV)	7166	7085
f'' (e <sup>-</sup> ) (S/K/Ca/Fe)	0.7/1.3/1.6/4.0	0.7/1.3/1.6/0.5
Crystal data		
Space group	$P6_{5}$	$P6_5$
Unit-cell parameters (Å)	a = 91.5, c = 43.0	a = 91.5, c = 43.1
Resolution range (Å)	30-2.5 (2.59-2.50)	30-2.5 (2.59-2.50)
Reflections (total/unique)	66060/6965	67058/7086
Multiplicity	9.5 (9.6)	9.5 (9.5)
Completeness (%)	95.8 (76.7)	97.4 (83.0)
$\langle I/\sigma(I)\rangle$	54.9 (13.1)	57.5 (14.2)
$R_{\rm merge}$ † (%)	10.8 (24.0)	10.2 (22.3)

†  $R_{\text{merge}} = \sum_{hkl} \sum_{i} |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_{i} I_i(hkl).$ 

angle of  $1.0^{\circ}$  and an exposure time of 0.5 s per frame. The absorption dose was estimated with *RADDOSE* (Murray *et al.*, 2004) and additional data sets for the detection of anomalous signal were collected at wavelengths of 1.73 and 1.75 Å. Processing and scaling of the diffraction data sets was performed using the *HKL*-2000 software package (Otwinowski & Minor, 1997). The crystals belonged to space group *P*6<sub>5</sub>, with unit-cell parameters a = b = 91.3, c = 43.3 Å. The crystal structure was solved by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010) in the *CCP*4 suite (Winn *et al.*, 2011) using the coordinates of the Rieske domain of the *b*<sub>6</sub>*f* complex from *Mastigocladus laminosus* (PDB entry 2e74; Yamashita *et al.*, 2007) as the search model. Starting from an initial model autotraced using the *ARP/wARP* program (Perrakis *et al.*, 1999), the output structure was manually improved and refined using *XtalView* and *CNS* (McRee, 1999; Brünger *et al.*, 1998). The final structure was validated with *RAMPAGE* (Lovell *et al.*, 2003). Data-collection and refinement statistics are shown in Tables 1 and 2. Isoelectric points



#### Figure 1

Spectroscopic analyses of TePetC. (a) UV–Vis spectra for nontreated (0.7 mg ml<sup>-1</sup>; green line), artificially reduced (blue line) and oxidized samples (red line); scaling is according to the 280 nm absorbance of the nontreated sample, with the difference spectra ( $A_{ox} - A_{red}$ ) shown as black lines (with a tenfold magnification). (b) UV (190–250 nm) region of the CD spectra at 293 K (blue), 356 K (pink) and 373 K (red). (c) Temperature-dependent unfolding of TePetC according to CD at 222 nm. The thermal melting point,  $T_{m}$ , was determined by a Boltzmann sigmoidal curve fitted to data between 348 and 367 K.

(pI) were calculated using the *Compute pI/Mw* tool of the ExPASy server (Gasteiger *et al.*, 2005). Figures of molecular models were prepared using the *PyMOL* program (DeLano, 2002). Coordinates and structure factors for TePetC have been deposited in the Protein Data Bank (PDB entry 3azc).

#### 3. Results and discussion

# 3.1. Purification and spectroscopic characterization of TePetC

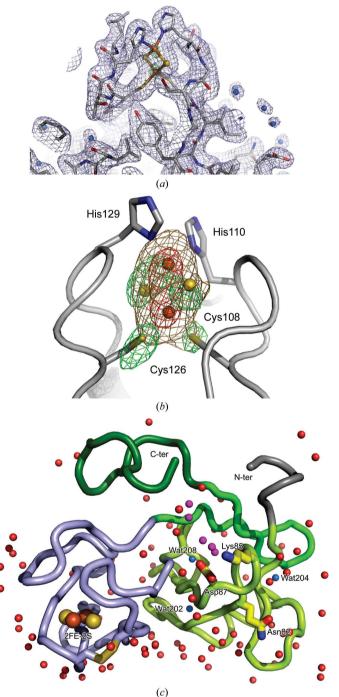
Heterologous overexpression of the soluble part of PetC from T. elongatus (TePetC) in E. coli and subsequent purification by affinity and size-exclusion chromatography yielded about 50 mg per litre of E. coli culture. The concentrated protein solution was deep red-brown, which is typical for ironsulfur proteins. Redox changes were accompanied by characteristic changes in the absorption spectrum, with maxima at 351, 475 and 579 nm in the oxidized minus reduced difference spectrum (Fig. 1a; Link, 2001). The fact that the spectrum of the nontreated sample was almost identical to the spectrum of the artificially reduced sample indicates that the purified Rieske proteins were mainly in the reduced state. From CD spectra (Fig. 1b) recorded at different temperatures a  $T_{\rm m}$  of 356 K was calculated (Fig. 1c), indicating the remarkable stability of the protein structure. The UV CD spectrum of TePetC, with a maximum at 224 nm and two minima at 200 and 232 nm (Fig. 1b), is atypical of proteins which consist predominantly of  $\beta$ -sheets. However, similar CD spectra have been reported for other small  $\beta$ -sheet-dominated proteins (see, for example, Knapp et al., 1998). These features may be caused by aromatic side chains, which have been shown to strongly affect UV CD spectra (Freskgård et al., 1994).

The construction of a truncated TePetC variant without the transmembrane domain and hinge region but with a His tag fused to the amino-terminus enabled large-scale heterologous overexpression of this redox-active protein. In contrast to previous reports (Holton *et al.*, 1996; Schneider *et al.*, 2000; Gubernator *et al.*, 2003), our system does not require an additional refolding process or reconstruction of the iron-sulfur cluster; this reduces the likelihood of artifacts in protein folding considerably.

#### 3.2. Crystal structure

Data collection was performed under a low X-ray dose with cryocooling at 40 K to avoid X-ray-radiation-induced bias (see, for example, Yano *et al.*, 2005; De la Mora *et al.*, 2012) in the TePetC structure. The dose per frame and the total dose were estimated to be  $5.5 \times 10^2$  and  $1.0 \times 10^5$  Gy, respectively, with the latter being smaller than the Henderson limit (Henderson, 1990) by more than two orders of magnitude. Diffraction spots could be observed to 1.9 Å resolution along the  $a^*$  and  $b^*$  axes, but only to 2.3 Å resolution along the  $c^*$  axis. The resolution of the data set was determined to be 2.0 Å, while the highest resolution shell had a relatively low completeness owing to the anisotropic diffraction. However, the electron density was improved upon the inclusion of

diffraction data from the highest resolution shell. Finally, the  $R_{\text{work}}$  and  $R_{\text{free}}$  factors converged to 19.7% and 23.6%,

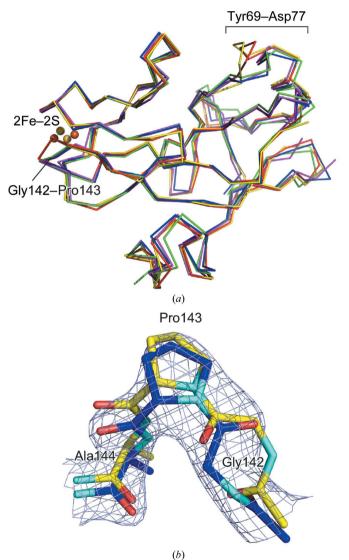


#### Figure 2

Structure of TePetC. (a)  $\sigma_A$ -Weighted  $2F_o - F_c$  map contoured at the  $1.0\sigma$  and  $10\sigma$  levels as a grey and a green mesh, respectively. (b) Anomalous difference Fourier map ( $\lambda = 1.73$  Å) at the  $3.0\sigma$  level (beige mesh) and the  $10\sigma$  level (red mesh); an additional map ( $\lambda = 1.75$  Å) is shown at the  $3.0\sigma$  level (green mesh). (c) Overall structure and subdomain architecture of TePetC: iron-binding subdomain (residues 105-149, light blue) and large subdomain (residues 54-104, light green; residues 150-165, green; residues 166-180, deep green), with additional residues at the N-terminus (grey). Water molecules are shown on the protein surface (red spheres), as internal molecules (blue) and as a buried water cluster in the putative ligand-binding pocket (pink; see §3.5). The side chains of Asn82, Lys85 and Asp87 in the protruding loop are shown as stick models.

Veit *et al.* • Thermophilic cyanobacterial *b*<sub>6</sub>*f*-type Rieske protein **1403 electronic reprint**  respectively (Table 1). The high-resolution electron-density map allowed the determination of exact conformations of the side chains and of the positions of water molecules (Fig. 2*a*).

The existence, positions and identities of heavy atoms was examined by anomalous difference Fourier maps which were calculated from data sets collected using X-rays of 1.73 and 1.75 Å wavelength, respectively (Table 2). These wavelengths correspond to high- and low-energy remotes of the K absorption edge of iron at 1.7433 Å. Strong densities were observed in the anomalous difference Fourier map at 1.73 Å in the vicinity of Cys108, His110, Cys126 and His129, while the corresponding map at 1.75 Å showed no strong density at these positions (Fig. 2b). However, weaker densities for S

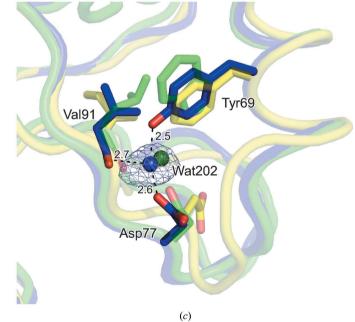


atoms could be observed in both data sets. These results indicate that the two heavy atoms which were coordinated by these residues and by two additional S atoms were unambiguously iron. We therefore conclude that the Rieske-type 2Fe–2S cluster is properly integrated into the protein from the *E. coli* cells without any additional treatment.

The soluble part of TePetC consists of an Fe–S clusterbinding subdomain and a large subdomain with three clusters of residues: amino acids 54–104, 150–165 and 166–180 (Fig. 2c). Among them, the C-terminal residues (166–180) are highly conserved in all  $b_6f$ -type Rieske homologues, while they are completely absent in other types of Rieske proteins. The secondary structure of TePetC is dominated by  $\beta$ -sheets, similar to other PetC1 proteins. However, the structures of the cyanobacterial and eukaryotic PetCs differ in detail, which may lead to functional and/or regulatory differences.

#### 3.3. Structural comparison of TePetC with other PetCs

Structural superposition of TePetC and other PetCs leads to root-mean-square deviation (r.m.s.d.) values of 0.7–1.4 Å for corresponding C<sup> $\alpha$ </sup> atoms (Fig. 3*a*). Unexpectedly, superpositions with SoPetC from *Spinacia oleracea* (spinach; PDB entry 1rfs; Carrell *et al.*, 1997) and CrPetC from the green alga *Chlamydomonas reinhardtii* (PDB entry 1q90; Stroebel *et al.*, 2003) revealed smaller r.m.s.d. values than those with other cyanobacterial PetCs, including MIPetC from the thermophilic cyanobacterium *Mastigocladus laminosus* (PDB entry 2e74;



#### Figure 3

Structural comparison with other PetC proteins. (a) Structures of PetC from *T. elongatus* (blue; this work), *S. oleracea* (green; PDB entry 1rfs; r.m.s.d. 0.7 Å; 65.6% residue identity), *C. reinhardtii* (magenta; PDB entry 1q90; r.m.s.d. 0.8 Å; 70.1% residue identity), *M. laminosus* (yellow; PDB entry 2e74; r.m.s.d. 1.2 Å; 68.5% residue identity) and *Nostoc* sp. PCC 7120 (red; PDB entry 2zt9; r.m.s.d. 1.1 Å; 68.5% residue identity) shown as C<sup> $\alpha$ </sup> models. Residue numbering is according to TePetC. (b) *Cis*-peptide bond between Gly142 and Pro143 of the TePetC structure (blue) superimposed with the corresponding *trans*-peptide bond of PetC from *M. laminosus* (yellow) as stick models. A  $\sigma_A$ -weighted  $2F_o - F_c$  map for the TePetC structure contoured at 2.0 $\sigma$  is shown as a grey mesh. (c) Internal water molecule (Wat202) coordinated by the side chains of Tyr69 (distance 2.5 Å) and Asp77 (2.6 Å) and the main-chain carbonyl of Val91 (2.7 Å). For comparison, PetC structures from a higher plant (*S. oleracea*; green) and from another cyanobacterium (*M. laminosus*; yellow) are shown. The OMIT map of Wat202 is contoured at 3.0 $\sigma$  as a grey mesh. Anomalous difference Fourier maps at long wavelengths ( $\lambda = 1.73$  and 1.75 Å) showed no electron-density peaks ( $4\sigma$  levels) at the position of Wat202.

Yamashita *et al.*, 2007). These differences in the r.m.s.d. values mainly arise from residues Tyr69–Asp77 and the fact that a *cis*-peptide bond occurs at Gly142-Pro143 exclusively in eukary-otic PetCs and TePetC (Fig. 3*b*).

Thermophilic and mesophilic PetC proteins show no apparent differences in the numbers of hydrogen bonds, hydrophobic interactions or disulfide bridges. However, the thermophilic PetCs from *T. elongatus* and *M. laminosus* have a more neutral pI than most PetCs from mesophilic cyanobacteria, indicating differences in the protein surface charge. It has been shown in previous studies that the exchange of only a few surface-exposed residues can result in considerably increased thermostability of a protein owing to electrostatic stabilization (Pace, 2000; Perl *et al.*, 2000). Additional salt

bridges (Kumar *et al.*, 2000; Taylor & Vaisman, 2010), longrange coulombic interactions (Grimsley *et al.*, 1999; Perl *et al.*, 2000) and the elimination of unfavourable electrostatic repulsion (Perl *et al.*, 2000) can contribute to such stabilization. As the crystal structures of the thermophilic TePetC and the mesophilic SoPetC show the same number of salt bridges, the latter two may contribute to the increased thermal stability of TePetC. We also identified additional internal water molecules in the TePetC structure (Wat202, Wat204 and Wat208) which are partly absent in the corresponding mesophilic proteins. In summary, both the presence of internal water molecules and the charge-altering mutation of specific surfaceexposed residues could considerably increase the  $T_{\rm m}$  of thermophilic PetC variants.

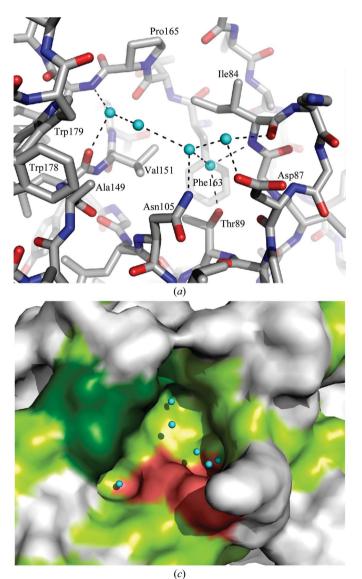
				ΔΔ* x* x	x x
<b>Te</b> BP-1	MAQVSG-MSDVP <mark>D</mark> MG <mark>RRO</mark> FM	L <mark>L</mark> TF <mark>G</mark> TITGTALGALYPVVK <mark>Y</mark>	FIPPASGGTG <mark>G</mark> -GAV <mark>AKD</mark> AL <mark>CN</mark> DIKV	SE <mark>YL</mark> AK <mark>H</mark> LP <mark>G</mark> DRS <mark>L</mark> A <mark>QGIKG</mark> D	P <b>TYVIV</b> TE <mark>D</mark> HQ <mark>I</mark> AN <mark>YGLNAV<mark>CTHLGC</mark>V<b>VP</b>W</mark> NVS 120
MI	MAQFTE-SMDVPDMG <mark>RRQ</mark> FMN		FIPPSGGAVGG-GTT <mark>AKD</mark> KL <mark>GN</mark> NVKV	SK <mark>FL</mark> ESHNAGDRV <mark>L</mark> VQGLKGD	P <b>TXIVVESKEAI</b> RD <mark>YGINAVCTHLGCVVP</mark> WNAA 120
Av ATCC 29413	MAQFSE-SVDVPDMGRRQFMN			SKF <b>L</b> ESHNVGDRTLVQGLKGD	PTYIVVESKEAITDYGINAVCTHLGCVVPWNAA 120
<b>C</b> ATCC 51142			FIPPSSGGAGG-GIT <mark>AKD</mark> AL <mark>GN</mark> DIMA		
C CCY0110				SN <mark>FL</mark> ANHNAGDRVLAQGLKGD	
C PCC 7424				SDFLATHNAGDRTLAQGLKGD	P <b>TYIVV</b> TEEKTIASYGLNAVCTHLGCVVPWNAS 120
N PCC 7120 O PCC 6506				SKFLESHNVGDRTLVQCLKGD SKFLAEHAVGDRVLAQCLKGD	
Sco PCC 7335				SKFLAEHAVGDRVLAQGLKGD AAFLEENSPGAKVLVQGLKGD	
Scy PCC 6803	MTOISG-SPDVPDLGRROFM		LIPPSSGGSGG-GVTAKDALONDVKA		
Cr	SEVPDMNKRNIMA		FVPPSSGGGGG-GQA <mark>/KD</mark> ALCNDIKA		
Cp	DDVPDMRKRNIM		FVPPSSGGGGG-GQA <b>KLALON</b> DIKA LVPPKAGGGAGAGQA <mark>KL</mark> AL <b>ON</b> DVKV		
Mp CCMP1545			FVPPGGGGGGGGGGGGGQP <mark>AK</mark> ALONDVKV		
Tp CCMP1335	EYYIDQNRRNLMN		FVPPSDDGGSG-TTI <b>AKD</b> AI <b>GN</b> EIFA		
At			FVPPGTGGGGG-GTPAKDALCNDVVA		
Ps			FVPPGSGGSGS-GLI <mark>AKD</mark> AI <mark>GN</mark> DILA		
So	DNVPDMQKRETLN	l <mark>l</mark> ll <mark>g</mark> alslptgymllpyas <mark>f</mark>	FVPPGGGAGTG-GTI <mark>AKD</mark> AL <mark>GN</mark> DVIA	AEWLKTHAPGDRT <mark>I</mark> T <mark>QGLKG</mark> D	PTYLVVESDKTLATFGINAVCTHLGCVV2FNAA 187
Zm	DRVPDMEKRKLMN	L <mark>L</mark> LL <mark>G</mark> AISLPTVGMVVPYGA <mark>F</mark>	FVPAGSGNGGG-GTY <mark>AKD</mark> KL <mark>GN</mark> DITV	EEWLNTHGPNDRTLA <mark>QGLKG</mark> D	P <mark>TYLVV</mark> EQDKTLAT <mark>YGINAVCTHLGC</mark> V <b>VP</b> WNGA 167
Av ATCC 29413	MDDTLNQLNPSMSRRQLL	FFTGAIVAATASAAIYPATKF	EMPPAESTDAEGGVLAKDKIGHPIPA	SQILV-QASGTRALIAGLAGE	PTYLTVREDGTLDPMGIVNNCTHLGCTFFWNPV 119
<b>C</b> ATCC 51142	MDNNLSFDQTSLSRRQLLN	FLTGSTVAVVAG <mark>SALYPAA</mark> KF	FTPPKET-GEGGAILAKDKLGNPIPA	SQILE-NPPETRALVAGLAGE	PTYLINTDANTLDGR <mark>BIVDNCTHLGCTFF</mark> WNET 118
C CCY0110	MENNISLDQASLSRRQLL	FLTGSTVAVVVGSALYPAAKF	FTPPKDI-GEGGAILAKDKLGNPIPA	SQIL <mark>A-NPP</mark> GTRALVAGLAGE	PTYLI <mark>VTD</mark> AN <mark>TI</mark> DGR <mark>GIVDNCTHLGCTFFW</mark> NEM 118
C PCC 7424	MENNLPIEN <mark>PS</mark> L <mark>SRRQLL</mark>			GQILA-NPAGTRALVAGLAGE	PTYLTVRDDGTLHPWGIVDNCTHLGCTFPWNEI 118
N PCC 7120	MDNSIPIESPSLSRRQLLN	FITGATVAVTAGAALYPAGKF		SQILA-EPPQTRALVAGLAGE	PTYLIVKEDHTLDRIGLVDNCTHLGCTFPWNPL 119
<b>O</b> PCC 6506	MDNTFPLDSPSISREQLL		LIPPTES-GAGGAIIAKUKLGNPIPA		PTYLINKODGAIDNIBIVDNCTHLGCTFFWNET 118
Sco PCC 7335	MDNALPLENPSLSRIQLMN			GQILV-EPPGTRALVAGIAGE	PTYLLVQDNISISKRGIVDNCTHLGCTFDWVPK 118
Scy PCC 6803	MDNTQAIAPPSYSRRQLL	FLAGTTVAVTASAGAYAMGKF	FVPPADKGGAG <mark>GGIIAKD</mark> VLGNPIPA	SQILA-EAPGTRALVAGLAGD	PTYLIVKEDGSIDSIGIVDSCTHLGCTFPWNGN 119
-					
<b>T</b> - DD 4		XX			
Te BP-1	E <mark>NK</mark> FICPCHGSQYDSTGKVVR	<b>IGPAP<mark>LSL</mark>ALV<mark>K</mark>ATVTED<mark>D</mark>K<mark>L</mark>V</b>			
МІ	ENKFKCPCHGSQYDETCKVIF	IGPAPL <mark>SL</mark> ADVKATVTEDDKTV IGPAPL <mark>SL</mark> ADCHATVQD-DNTV	LTP <mark>W</mark> TET <mark>DFRT</mark> GEK <mark>PWW</mark> V- 179		
MI Av ATCC 29413	ENKEK <mark>CPCHGSQY</mark> DETCKVIR E <mark>NKE</mark> K <mark>CPCHGSQYD</mark> ATCKVVR	IGPAELS ALVKATVTEDDKLV IGPAELS ALCHATVQD-DNIV IGPAEKS ALSHAKTEN-DKIV	LTPNTETDFRTGEK <mark>PWN</mark> V- 179 LTPNTETDFRTGEE <mark>PWN</mark> S- 179		
<b>MI</b> <b>A∨</b> ATCC 29413 <b>C</b> ATCC 51142	E <mark>NKFKCPCHGSQ¥DETCKVI</mark> F ENKFKCPCHGSQ¥DATCKVVF E <mark>DKFICPCHGSQ¥NAQCKVVF</mark>	IGPAPLSIALVKATVTED <mark>D</mark> KIV IGPAPLSIALCHATVQD-DNIV IGPAPKSIALSHAKTEN-DKIV IGPAPLSIALAHVNVSEDDKVV	LTPNTETDERT <mark>GEKPWW</mark> V- 179 LTPNTETDERTGEEPWWS- 179 ESENTETDERTEQD <mark>EWW</mark> A- 180		
MI Av ATCC 29413 C ATCC 51142 C CCY0110	ENK <mark>EKCECHGEQ</mark> IDETCKUTE ENKEKC <mark>PCHGEQ</mark> IDATCKUVE EDKEICPCHGEQINAQCKUVE EDKEICPCHGEQINAQCKUVE	IGPAELS <mark>H</mark> ALVKATVTED <mark>DKIV IGPAELSH</mark> ALCHATVQD-DNIV IGPAPKSHALSHAKTEN-DKIV IGPAPLSHALAHVNVSEDDKVV IGPAPL <mark>SH</mark> ALAHVNVTEG <mark>DKV</mark> L	LTPMTET FRTSCREWNW- 179 LTPMTET FRTSCEEPWNS- 179 FSENTET FRTEQDPNNA- 180 ESENTET FRTEQDPNNA- 180		
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424	ENKEK CPCHGE QVDETCKVT ENKEK CPCHGE QVDATCKVVE EDKET CPCHGE QVDATCKVVE EDKET CPCHGE QVNAQCKVVE EDKET CPCHGE QVNACCKVVE EDKET CPCHGE QVNSECKVVE	IGP. ELS ALV ATVTED <mark>K</mark> V IGP. ELS ALC ATVQD-DN V IGP. EKS ALS AKTEN-DKIV IGP. ELS ALA VNVSEDKV IGP. ELS ALA VNVTEGKV IGP. ELS ALV ADVTEDTV	LTPOTET FRTGER GER MAN - 179 ITPOTET FRTGEREMMS- 179 REATET FRTEQTEMMA- 180 REATET FRTEQTEMMA- 180 FTPOTET FRTNEDGEMMA- 180	At	Arabidopsis thaliana Anabaena variabilis
MI Av ATCC 29413 C ATCC 51142 C CCY0110	ENKEKCPCHGEQYDETCKVT ENKEKCPCHGEQYDATCKVVF EDKEICPCHGEQYNAOCKVVF EDKEICPCHGEQYNAOCKVVF EDKEICPCHGEQVNACKVVF EDKEICPCHGEQVDATCKVVF	IGPAELS <mark>H</mark> ALVKATVTED <mark>DKIV IGPAELSH</mark> ALCHATVQD-DNIV IGPAPKSHALSHAKTEN-DKIV IGPAPLSHALAHVNVSEDDKVV IGPAPL <mark>SH</mark> ALAHVNVTEG <mark>DKV</mark> L	LTPMTET FRTSCREWNW- 179 LTPMTET FRTSCEEPWNS- 179 FSENTET FRTEQDPNNA- 180 ESENTET FRTEQDPNNA- 180	Av	Anabaena variabilis
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120	ENKEKCPCHGEQYDETCKVT ENKEKCPCHGEQYDATCKVVF EDKEICPCHGEQYNAOCKVVF EDKEICPCHGEQYNAOCKVVF EDKEICPCHGEQVNACKVVF EDKEICPCHGEQVDATCKVVF	IGP. ELS ALVAATVTEDEK V IGP. ELS ALCAATVQD-DN V IGP. ELS ALS IAKTEN-DK V IGP. ELS ALA VNVSEDEK V IGP. ELS ALVAVVTEGEK L IGP. ELS ALVADVTEDEV V IGP. EKS ALS IAKTEN-DK V	TPUTET FRIGERENNV- 179 TPUTET FRIGERNNS- 179 TSETTET FRIEDENNS- 180 TSETTET FRIEDENNA- 180 TPUTET FRIEDENNA- 180 TPUTET FRIEDENNS- 179 TPUTET FRIEDENNS- 179		
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506	ENKIFK CPCHGS QVDETCKVI ENKIFK CPCHGS QVDATCKVVF EDKIFI CPCHGS QVNAQCKVVF EDKIFI CPCHGS QVNAQCKVVF EDKIFI CPCHGS QVNSEKVVF ENKIFI CPCHGS QVNNEKVVF ENKIFI CPCHGS QVNNEKVVF	IGP. PLS ALVAATVTEDDK V IGP. PLS ALCHATVQD-DN V IGP. PLS ALSHAKTEN-DK V IGP. PLS ALAHVNVEGDK V IGP. PLS ALAHVNVTEGDK V IGP. PLS ALAHVNVTEGDT V IGP. PLS ALSHAKTEN-DK V IGP. PLS ALAHATVAD-DK V	TPITET FRTGERENWV- 179 TPITETOFRTGEDENWS- 179 SEITETOFRTEQDENWA- 180 SEITETOFRTEQDENWA- 180 TPITETOFRTEQDENWA- 180 TPITETOFRTEEDENWS- 179 TPIKETOFRTGEAPNWS- 180	Av C Cp Cr	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335	ENKEK CPCHGS QVDETCKVT ENKEK CPCHGS QVDATCKVVF EDKET CPCHGS QVNAQCKVVF EDKET CPCHGS QVNAQCKVVF EDKET CPCHGS QVNAQCKVVF ENKEK CPCHGS QVDATCKVVF AGKEMCPCHGS QVDATCKVVF ENKEMCPCHGS QVNAECKVVF	IGP ELS ALVIATVTEDRK V IGP ELS ALCIATVQD-DN V IGP ELS ALSIAKTEN-DKIV IGP ELS ALAIVNVSEDRK V IGP ELS ALAIVNVEGDKV IGP ELS ALVIADVTEDDTV IGP ELS IALAIATVAD-DKV IGP ELS ALAIATVAD-DKV IGP ELS ALAIATVAD-DKV	TPHTET FRTGER WINV- 179 TPHTET FRTGER WINV- 179 SEINET FRTGER WINA- 180 SEINET FRTGER WINA- 180 TPHTET FRTDER WINA- 180 TSIDET FRTGER WINA- 180 TSIDET FRTGER WINA- 179 TPHRET FRTGER WINA- 179 STITET FRTDER WINA- 180	Av C Cp Cr Mi	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803	ENKEK CPCHGE QYDETEK VI ENKEK CPCHGE QYDATEK VV EDKEI CPCHGE QYDATEK VV EDKEI CPCHGE QYNAOCK VV EDKEI CPCHGE QYNAOCK VV ENKEK CPCHGE QYDATEK VV ENKEK CPCHGE QYDATEK VV ENKEK CPCHGE QYDATEK VV ENKEM CPCHGE QYDATEK VV	ICP ELS ALV ATVTEDRK V     ICP ELS ALC ATVQD-DN V     ICP ELS ALC ATVQD-DN V     ICP ELS ALA VNVEDDK V     ICP ELS ALA ALVADVEDD V     ICP ELS ALA ALVADVEDDVV	TTP TET FATGEKENNV- 179 TTP TET FATGEENNS- 179 SE TET FATGEENNS- 180 FTP TET FATGOLONA- 180 TTP TET FATGEENNS- 180 TTP TET FATNELONS- 180 TTP TET FATGEENNS- 180 TTP TET FATGEENNS- 180 TTN EET FATGENNA- 180 ST TET FATGEACNA- 180 ST TET FATGEACNA- 206	Av C Cp Cr MI Mp	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545	ENKEK CPCHGE QVDET CKVI ENKEK CPCHGE QVDAT CKVV EDKEI CPCHGE QVNAQCKVV EDKEI CPCHGE QVNAQCKVV EDKEI CPCHGE QVDAT CKVV ENKEI CPCHGE QVDAT CKVV ACKMOPCHGE QVDAT CKVV ENKEMCPCHGE QVNAE CKVV ENKEKCPCHGE QVNAE	ISP ALVATVTEDRK V     ISP ALVATVTEDRK V     ISP ALS ALCATVQD-DN V     ISP ALS ALS AKTEN-DKIV     ISP ALS ALS AKTEN-DKIV     ISP ALS ALA VNVSEDRK V     ISP ALS ALA ANVTEDDTK V     ISP ALS ALA ATVAD-DKIV     ISP ALS ALA ATVAD-DKIV     ISP ALS ALA ALVTODEK V     ISP ALS ALA ALVSVLEDZVIA     ISP ALS ALA ALVSVLEDZVIA     ISP ALA ALA ALVSVLEDZVIA	TP:TET:FATGER     179       TP:TET:FATGER     179       SE:TET:FATGER     179       SE:TET:FATGER     180       SE:TET:FATGER     180       TP:TET:FATGER     180       TP:TET:FATGER     180       TP:TET:FATGER     180       TP:TET:FATGER     180       TP:TET:FATGER     180       TN:DET:FATGER     180       TN:DET:FATGER     180       STITET:FATGER     180       STITET:FATGER     180       STITET:FATGER     206       FFT:TET:TET:MARK     206       FFT:TET:FATGER     219       SFT:TET:FATGER     219	Av C Cp Cr MI Mp N	Anabaona variabilis Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1335	ENKEK CPCHGE QVDETEK VI ENKEK CPCHGE QVDATCKVV EDKEI CPCHGE QVDATCKVV EDKEI CPCHGE QVDATCKVV EDKEI CPCHGE QVDATCKVV ENKEK CPCHGE QVDATCKVV ENKEK CPCHGE QVDATCKVV ENKEK CPCHGE QVDATCKVV ENKEK CPCHGE QVDATCKVV ENKEK CPCHGE QVDATCKVV ENKEK CPCHGE QVDECKVV ENKEK CPCHGE QVDECKVV	GP2 ELS   ALV ATVTEDEK   V     GP2 ELS   ALC   ATVQD-DN   V     GP2 ELS   ALS   AKTEN-DKIV   V     GP2 ELS   ALA   VNVSEDDK   V     GP2 ELS   ALA   VNVSEDDK   V     GP2 ELS   ALA   VNVTEDDK   V     GP2 ELS   ALA   VNVTEDDK   V     GP2 ELS   ALA   AVTEDDK   V     GP2 ELS   ALA   AUTEDDK   V     GP3 ELS   ALA   AUTEDDK   V     GP4 ELS   ALA   AUTEDK   V     GP	TP:TET:FATGEKENNV-179       TP:TET:FATGEENNS-179       SE:DET:FATGEENNS-180       SE:DET:FATGEENNS-180       TP:TET:FATGEENNS-180       TP:TET:FATGEENNS-180       TP:TET:FATGEENNS-180       TP:TET:FATGEENNS-180       TP:TET:FATGEENNS-180       TP:TET:FATGEENNS-180       TP:TET:FATGEENNS-180       TP:TET:FATGEENNS-180       SE:TET:FATGEENNS-180       FS:TET:FATGEENNS-180       FS:TET:FATGEENNS-206       FS:TET:FATGEENNS-219       MTA TEN FATGAKENNS-181	Av C Cp Cr MI Mp	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1335 At	ENKEK CPCHGE QVDET KVV ENKEK CPCHGE QVDAT KVVV EDKEI CPCHGE QVNAQEKVV EDKEI CPCHGE QVNAQEKVV EDKEI CPCHGE QVDAT KVV ENKEK CPCHGE QVDAT KVV	IOPA ELS   ALV ATVTEDEK   V     IOPA ELS   ALC   ATVQD-DNIV     IOPA ELS   ALC   ATVQD-DNIV     IOPA ELS   ALS   AKTEN-DRIV     IOPA ELS   ALA   VNVSEDDK   V     IOPA ELS   ALA   VNVSEDDK   V     IOPA ELS   ALA   VNVTEDDK   V     IOPA ELS   ALA   VNVTEDDK   V     IOPA ELS   ALA   AVTTEDDN   V     IOPA ELS   ALA   ATVATDDK   V     IOPA ELS   ALA   ATVTEDDN   V     IOPA ELS   ALA   ATVATDDK   V     IOPA ELS   ALA   ATVTEDDN   V     IOPA ELS   ALA   ATVTEDDN   V     IOPA ELS   ALA   ATVTEDDN   V     IOPA ELS   ALA   ATVENDESCUT   V     IOPA ELS   ALA   ALVENDESCUT   V     IOPA ELS   ALA   ALA   ALVENDESCUT     IOPA ELS   ALA   ALA   ALVENDESCUT     IOPA ELS   ALA   ALA   ALA     IO	ITP     TFATGEKENV-179       ITP     TFATGEE     179       SE     TFATGEE     180       FATEQDE     180       FTP     TFATEQDE     180       TTP     TFT     FATEQDE     180       TTP     TFT     FATEQDE     180       TTP     TFT     FATEQDE     180       TTP     TFT     FATEQE     180       TTP     TFT     FATEQE     180       TTP     TFT     FATGEE     180       TTP     TFT     FATGEE     180       TTP     TFT     FATGEE     180       TTP     TFT     FATGEE     180       TTP     TFT     FATGEL     180       TT     TFT     FATGEL     180       TT     TFT     TFT     TGL       TT     TT     TT     179       TT     TT     TT     180       TT     TT     TT     180       TT     TT     TT <t< th=""><th>Av C Cp Cr MI Mp N O Ps Sco</th><th>Anabaona variabilis Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus</th></t<>	Av C Cp Cr MI Mp N O Ps Sco	Anabaona variabilis Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1335 At Ps	ENKEK CPCHGE QVDET CKVV ENKEK CPCHGE QVDATCKVVV EDKEI CPCHGE QVDATCKVVV EDKEI CPCHGE QVDACCKVVV EDKEI CPCHGE QVDACCKVVV ENKEK CPCHGE QVDATCKVVV ENKEK CPCHGE QVDATCKVVV	ISP ALV ATVTEDEK     VICP ELS ALC ATVQD-DN V     VICP ELS ALC ATVQD-DN V     VICP ELS ALA VNVEEDKV     VICP ELS ALA VNVTEGDKVL     VICP ELS ALA VNVTEGDKVL     VICP ELS ALA VNVTEGDKVL     VICP ELS ALA ATVAD-DK     VICP ELS ALA ATVAD-VICPDN     VICP ELS ALA ATVAD-VICPDN     VICP ELS ALA ATVAD-VICPON     VICP ELS ALA AVVICPONSKIV     VICP ELS ALA ALANDSDTVI     VICP ELS ALA ALANDNDSKIV     VICP ELS ALA ALANDRAKINGSKVI     VICP ELS ALA ALANDRAKING	ITP     TFATGEKENW-179       ITP     TFATGEENWS-179       ISE     TFATGEENWS-180       ISE     TFATEODENWA-180       IST     TFATEODENWA-238	Av C Cp Cr MI N N O Ps Sco Scy	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechocystis
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1545 Tp CCMP1335 At Ps So	ENKEK CPCHGE QVDET CKVV ENKEK CPCHGE QVDATCKVVV EDKEI CPCHGE QVDATCKVVV EDKEI CPCHGE QVDACCKVVV EDKEI CPCHGE QVDATCKVV ENKEK CPCHGE QVDATCKVV ACKEM CPCHGE QVDATCKVV ENKEK CPCHGE QVDATCKVV ENKEL CPCHGE QVDATCKVV ENKEL CPCHGE QVDATCKVV ENKEL CPCHGE QVDATCKVV	ISP ALV ATVTEDRK V     ISP ALS ALV ATVTEDRK V     ISP ALS ALS AKTEN-DRIV     ISP ALS ALS AKTEN-DRIV     ISP ALS ALS AKTEN-DRIV     ISP ALS ALA VNVSEDDTV     ISP ALS ALA VNVSEDDTV     ISP ALS ALV ADVTEDTVV     ISP ALS ALV ADVTEDTVV     ISP ALS ALV ADVTEDTVV     ISP ALS ALA AVNO-DKT     ISP ALS ALA ADVTEDDTVV     ISP ALS ALA ADVTEDDTVV     ISP ALS ALA ADVTEDDTVV     ISP ALS ALA ADVTEDDTVV     ISP ELS ALA ATVTDDKTV     ISP ELS ALA ATVTDDKVV     ISP ELS ALA ADVTEDDTVV     ISP ELS ALAADIEACOVAESGLUT	ITP TET FATGER ONV-     179       ITP TET FATGEE ONS-     179       SE TET FATGEE ONS-     180       SE TET FATGEE ONS-     180       TTP TET FATGEE ONS-     179       ST TET FATGEE ONS-     206       EST TET FATGLO ON A-     238       ESP TET FATAL (A 2000 N - 219     219       MTA TEN FATGEA ON A- 234     24       FVP TET FATGEA ON A- 234     24	Av C Cp Cr MI Mp N O Ps Sco Scy So	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechocystis Spinacia oleracea
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1335 At Ps So Zm	ENKER CPCHGE QVDET EKV ENKER CPCHGE QVDAT EKV EDKET CPCHGE QVDAT EKV EDKET CPCHGE QVDAT EKV EDKET CPCHGE QVDAT EKV ENKER CPCHGE QVDAT EKV	ISP ALV ATVTEDRK V     ISP ALX ALVATVTEDRK V     ISP ALX ALVATVTEDRK V     ISP ALX ALXANDALX     ISP ALXANDALX	TP: TBT FATGER     179       TP: TBT FATGER     179       SE TDT FATGER     179       SE TDT FATGER     180       SE TDT FATGER     180       TP TET FATGER     180       TP TET FATGER     180       TS TDT FATGER     180       ST TT TT FATGER     180       FST TDT FATGER     180       FST TT TT FATGER     180       FST TDT FATGER     206       FST TDT FATGER     238       FST TDT FATGER     238       FST TDT FATGER     219       MTA TEN FATGER     229       VP TST FATGEN     247       VP TST FATGEN     247       VP TST FATGEN     247       VP TST FATGEN     247	Av C Cp Cr MI Mp N O Ps Sco Scy So Te	Anabaona variabilis Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococsus Spinacia oleracea Thermosynechoccoccus elongatus
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1545 Tp CCMP1335 At Ps So Zm Av ATCC 29413	ENKIK CPCHGS QMDET KKVT ENKIK CPCHGS QMDAT KKVT EDKIT CPCHGS QMAQCKVT EDKIT CPCHGS QMAQCKVT EDKIT CPCHGS QMACKVT ENKIK CPCHGS QMAE KVT ENKIK CPCHGS QMAE ENKIK CPCHGS QMAC ENKIK CPCHGS QMAC	GS2 ELS   ALVEATVTEDEK   V     GS2 ELS   ALC   ATVQD-DN   V     GS2 ELS   ALC   ATVQD-DN   V     GS2 ELS   ALC   ATVQD-DN   V     GS2 ELS   ALA   VNVSEDDK   V     GS2 ELS   ALA   VNVSEDDK   V     GS2 ELS   ALA   VNVTEDDK   V     GS2 ELS   ALA   ALVTEDDK   V     GS2 ELS   ALA   ALVELDKSCL   T     GS2 ELS   ALA   ALVELDKSCL   T <th>TP: TET: FATGER ON V-179     TP: TET: FATGER ON V-179     SE: TET: FATGER ON V-180     SE: TET: FATGER ON V-180     TP: TET: FATGER ON V-180     TP: TET: FATGER ON V-180     TTP: TET: FATGER ON V-180     FATER FATGER ON V-238     FST TET: FATGLE ON V-238     FSP TET: FATGLE ON V-234     FVP WET: FATGER NA V-234     FVP TET: FATGER NA V-234     FVP NET: FATGER NA V-178</th> <th>Av C Cp Cr MI Mp N O Ps Sco Scy So</th> <th>Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococcus Synechocostis Spinacia oleracea Thermosynechococcus elongatus Thalassiosira pseudonana</th>	TP: TET: FATGER ON V-179     TP: TET: FATGER ON V-179     SE: TET: FATGER ON V-180     SE: TET: FATGER ON V-180     TP: TET: FATGER ON V-180     TP: TET: FATGER ON V-180     TTP: TET: FATGER ON V-180     FATER FATGER ON V-238     FST TET: FATGLE ON V-238     FSP TET: FATGLE ON V-234     FVP WET: FATGER NA V-234     FVP TET: FATGER NA V-234     FVP NET: FATGER NA V-178	Av C Cp Cr MI Mp N O Ps Sco Scy So	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococcus Synechocostis Spinacia oleracea Thermosynechococcus elongatus Thalassiosira pseudonana
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1545 Tp CCMP1335 At Ps So Zm Av ATCC 29413 C ATCC 51142	ENKEK CPCHGE QVDET KVV ENKEK CPCHGE QVDAT KVVV EDKEI CPCHGE QVDAT KVVV EDKEI CPCHGE QVDAT KVVV EDKEI CPCHGE QVDAT KVVV ENKEK CPCHGE QVDAT KVVV AGKEN CPCHGE QVDAT KVVV ENKEK CPCHGE QVDAT KVVV	GP2 ELS   ALV ATVTEDEK   V     GP2 ELS   ALC   ATVQD-DNIV     GP2 ELS   ALC   ATVQD-DNIV     GP2 ELS   ALC   ATVQD-DNIV     GP2 ELS   ALA   VNVSEDDK     GP2 ELS   ALA   VNVSEDDK     GP2 ELS   ALA   VNVSEDDK     GP2 ELS   ALA   VNVTEDDT     GP2 ELS   ALA   VNVTEDDT     GP2 ELS   ALA   ALVTEDDNV     GP2 ELS   ALA   ATVATDDK     GP2 ELS   ALA   ALVTEDDNV     GP2 ELS   ALA   ALVTEDDNV     GP3 ELS   ALA   ALVATDDK     GP4 ELS   ALA   ALVTEDDNV     GP5 ELS   ALA   ALVTEDDNV     GP6 ELS   ALA   ALVENDKSGK     GP7 ELS   ALAADVDDNSGK   V     GP6 ELS   ALAADVDDOKGK   V     GP7 ELS   ALAADVDDOKGK   V     GP8 ELS   ALAADVDDOKK   V     GP9 ELS   ALAADVDDOKK   V     GP9 ELS   ALACDVDOKGK   V     GP9 ELS   ALACDVDOK <th>ITP: ToT: FATGER ENDY-179     ITP: ToT: FATGER ENDS-179     SE. TOT: FATEQUE ONA-180     FTP: ToT: FATEQUE ONA-180     ITP: TOT: FATEQUE ONA-180     ITTP: TOT: FATEQUE ONA-180     ITT: FATELEONA-180     ITT: FATELEONA-181     ITT: FATELEONA-238     IST: TOT: FATELEONA-238     IST: TOT: FATELEONA-238     IST: TOT: FATELEONA-238     IST: TOT: FATELEONA-234     IST: TOT: PATELEONA-234     IST: TOT: PATELEONA-234</th> <th>Av C Cp Cr MI Mp N O Ps Sco Scy So Te Tp</th> <th>Anabaona variabilis Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococsus Spinacia oleracea Thermosynechoccoccus elongatus</th>	ITP: ToT: FATGER ENDY-179     ITP: ToT: FATGER ENDS-179     SE. TOT: FATEQUE ONA-180     FTP: ToT: FATEQUE ONA-180     ITP: TOT: FATEQUE ONA-180     ITTP: TOT: FATEQUE ONA-180     ITT: FATELEONA-180     ITT: FATELEONA-181     ITT: FATELEONA-238     IST: TOT: FATELEONA-238     IST: TOT: FATELEONA-238     IST: TOT: FATELEONA-238     IST: TOT: FATELEONA-234     IST: TOT: PATELEONA-234	Av C Cp Cr MI Mp N O Ps Sco Scy So Te Tp	Anabaona variabilis Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococsus Spinacia oleracea Thermosynechoccoccus elongatus
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1545 Tp CCMP1335 At Ps So Zm Av ATCC 29413 C ATCC 51142 C CCY0110	ENKER CPCHGE QYDET CKVV ENKER CPCHGE QYDATCKVVV EDKEI CPCHGE QYDACCKVVV EDKEI CPCHGE QYDACCKVVV ENKER CPCHGE QYDACCKVVV ENKER CPCHGE QYDATCKVVV ENKER CPCHGE QYDACCKVVV ENKER CPCHGE QYDACCKVVV	ISPA	HTP   TFAT   GEK   179     HTP   TFAT   GEK   179     HTP   TFAT   GEE   170     HSE   TGT   FATEQD   180     HSE   FATEGE   179   180     HST   FATEGE   170   170     ST   FATEGE   180   180     HST   FATEGE   180   179     HST   FATEGE   180   179     HST   FATEDE   180   180     HST   FATEDE   180   180     HST   FATEDE   180   238     HST   FATEDE   TGE   234     HVP   FATE   FATERE   234     HVP   FATERE   HST   2	Av C Cp Cr MI Mp N O Ps Sco Scy So Te Tp	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococcus Synechocostis Spinacia oleracea Thermosynechococcus elongatus Thalassiosira pseudonana
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1545 Tp CCMP1545 At Ps So Zm Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424	ENKIRK CYCHGE QYDAT CKWY ENKIRK CYCHGE QYDAT CKWY EDKIR I CYCHGE QYDAT CKWY EDKIR I CYCHGE QYDAT CKWY EDKIR I CYCHGE QYNAC KWY ENKIRK I CYCHGE QYDAT CKWY ENKIRK CYCHGE QYDAT CKWY ENKIRK CYCHGE QYNAE KWY ENKIRK CYCHGE QYNAE KWY ENKIR CYCHGE QYNAC KWY	ISPA   ALVATVTED   V     ISPA   ALVATVTED   V     ISPA   ALSAKTEN-DKIV   D     ISPA   ALSAKTEN-DKIV   V     ISPA   ALAVNVSEDDKV   V     ISPA   ALAVNVTEDDKV   V     ISPA   ALAATVAD-DKV   V     ISPA   ALAATVTDDDKV   V     ISPA   ALAAATVTDDDKV   V     ISPA   ALAAATVSVLEDGVVA   V     ISPA   ALAAATVSVLEDGVA   V     ISPA   ALAAATVSVLEDGVA   V     ISPA   ALAAADIDEAGKVU   V     ISPA   ALAACDVDD-GKVV   V     ISPA   ALAACDVDD-GKVV   ISPA     ISPA   ALAACDVDD-GKVV   V     ISPA   ALAACVDD-GKVV   ISPA     ISPA   ALAACVDD-GKVV   ISPA     ISPA   ALVAVDD-AVVD-GKVV   ISPA	TP: TB T. FATGER WAY-179     TP: TB T. FATGER WAY-180     TB T. FATGER WAY-180     SE TD T. FATGER WAY-180     TP: TB T. FATGER WAY-180     TP: TB T. FATGER WAY-180     TF: TB T. FATGER WAY-180     TF: TB T. FATGER WAY-180     TS TD T. FATGER WAY-180     TS TD T. FATGER WAY-180     TT TB T. FATGER WAY-180     TT TB T. FATGER WAY-180     FAT TD T FATGER WAY-180     FST TD T. FATGER WAY-181     FY TB T. FATGER WAY-177     SS TT TP TG K WAY 177     SS TT TP TG NG R WY-177	Av C Cp Cr MI Mp N O Ps Sco Scy So Te Tp	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococcus Synechocostis Spinacia oleracea Thermosynechococcus elongatus Thalassiosira pseudonana
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1545 Tp CCMP1335 At Ps So Zm Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120	ENKIK CPCHGS QMDET KKVT ENKIK CPCHGS QMDET KKVT EDKIFI CPCHGS QMAQSKVT EDKIFI CPCHGS QMAQSKVT EDKIFI CPCHGS QMACKVT ENKIFI CPCHGS QMAESKVT AGSIMOPCHGS QMAESKVT ENKIFI CPCHGS QMACSKVT ENKIFI CPCHGS QMACSKVT DQC QCCHGF R DAESSVT DQC QCCHGF R APOSSVT	ISP ALVATVTEDK     ISP ALS ALVATVTEDK     ISP ALS ALS AKTEN-DRIV     ISP ALS ALS AKTEN-DRIV     ISP ALS ALS AKTEN-DRIV     ISP ALS ALA VNVEDDK     ISP ALS ALA VNVEDDK     ISP ALS ALA VNVEDDK     ISP ALS ALS AKTEN-DRIV     ISP ALS ALS AKTEN-DRIV     ISP ALS ALS AKTEN-DRIV     ISP ALS ALS AKTEN-DRIV     ISP ALS ALA AVVEDDVV     ISP ALS ALA ALVDEDK     ISP ALS ALA ALVDEDK     ISP ALS ALA ALVDEDK     ISP ALS ALA ALVDUEDK     ISP ALA ALADICEASCLUT     ISP ALA ALADICEASCLUT     ISP ALA ALADIDEAGKU     ISP ALA ALOUDD-GKV     ISP ALA ALOUDD-GKV     ISP ALA ALOUDD-GKV     ISP ALACUUD-GKV     ISP KV VAVID-LA W     ISP KV VAVID-LA W     ISP KV VAVID-LA W     ISP KV VAVID-LA W     <	ITP: TET: FATGER ON V-179     ITP: TET: FATGER ON V-170     SE: TET: FATGER ON V-180     ISE: TET: FATGER ON V-180     ISE: TET: FATGER ON V-180     ISE: TET: FATGER ON V-180     ITP: TET: FATGER ON V-180     IST: TET: FATGER ON V-200     ST: TET: FATGER ON V-200     FATGER ON V-181     VP: VET: FATGER ON V-200     FT: TET: FATGER ON V-200     FT: TET: FATGER ON V-200     FY: TET: TET: FATGER ON V-200     FY: TET: FATGER ON V-177     SE: TT: TET: TET: FATGER V-177     SF: TN: TET: TET: TER V-177     SF: TN: TET: TET: TER V-177     SF: TN: TET: TET: TER V-177     SF: TN: TET: TET: TEN: V-177	Av C Cp Cr MI Mp N O Ps Sco Scy So Te Tp	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococcus Synechocostis Spinacia oleracea Thermosynechococcus elongatus Thalassiosira pseudonana
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1545 Tp CCMP1545 At Ps So Zm Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424	ENKIK CPCHGS QMDET KKVV ENKIK CPCHGS QMDAT KKVV EDKIS I CPCHGS QMAQ KKVV EDKIS I CPCHGS QMAQ KKVV ENKIS I CPCHGS QMAC KVV ENKIS I CPCHGS QMAE KVV AGKIM CPCHGS QMAE KVV ENKIS I CPCHGS QMAC KVV I CPCHS VV I CPCH	ISP ALVATVTEDEK     VICE ELS ALC ATVQD-DN V     VICE ELS ALC ATVQD-DN V     VICE ELS ALA VAVEDUC     VICE ELS ALA VAVEDUC     VICE ELS ALA VAVEDUC     VICE ELS ALA VAVEDUC     VICE ELS ALA AVADUTEDT     VICE ELS ALA AVADUTEDT     VICE ELS ALA ATVADUC     VICE ELS ALA AVVEDUC     VICE ELS ALA AVVELOUC	TP: TB T. FATGER WAY-179     TP: TB T. FATGER WAY-180     TB T. FATGER WAY-180     SE TD T. FATGER WAY-180     TP: TB T. FATGER WAY-180     TP: TB T. FATGER WAY-180     TF: TB T. FATGER WAY-180     TF: TB T. FATGER WAY-180     TS TD T. FATGER WAY-180     TS TD T. FATGER WAY-180     TT TB T. FATGER WAY-180     TT TB T. FATGER WAY-180     FAT TD T FATGER WAY-180     FST TD T. FATGER WAY-181     FY TB T. FATGER WAY-177     SS TT TP TG K WAY 177     SS TT TP TG NG R WY-177	Av C Cp Cr MI Mp N O Ps Sco Scy So Te Tp	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococcus Synechocostis Spinacia oleracea Thermosynechococcus elongatus Thalassiosira pseudonana
MI Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506 Sco PCC 7335 Scy PCC 6803 Cr Cp Mp CCMP1545 Tp CCMP1545 Tp CCMP1335 At Ps So Zm Av ATCC 29413 C ATCC 51142 C CCY0110 C PCC 7424 N PCC 7120 O PCC 6506	ENKIK CPCHGE QVDET KVV ENKIK CPCHGE QVDAT KVVV EDKIFI CPCHGE QVDAT KVVV EDKIFI CPCHGE QVDAT KVVV EDKIFI CPCHGE QVDAT KVVV ENKIK CPCHGE QVDAT KVVVV ENKIK CPCHGE QVDAT KVVV ENKIK CPCHGE QVDAT KVVVV ENKIK CPCHGE QVDAT KVVVV ENKIK CPCHGE QVDAT KVVVV ENKIK CPCHGE QVDAT KVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV	ISP ALVATVTEDEK     VICE ELS ALC     ATVALL     ATVALL     ATS     AKTEN-DKIV     CIP ELS     ATA     CIP ELS     ATA     VIVESEDKV     CIP ELS     ATA     ATA     CIP ELS     ATA     ATA     CIP ELS     ATA     CIP ELS     ATA     ATA     CIP ELS     ATA	ITP:   T.F.A.T.GEK.E.N.V   179     ITP:   T.F.T.GEE   ITP:     ISE   T.G.T.F.T.GEE   ITP:     ISE   T.G.T.F.T.GEE   ITP:     ISE   T.G.T.F.T.GEE   ITP:     ISE   T.F.T.GEE   ITP:     ISE   T.F.T.T.GEE   ITP:     ISE   T.F.T.T.GEE   ITP:     IST   T.F.T.GEE   ITP:     IST   T.F.T.GEL   ITP:     IST   T.F.T.GEL   ITP:     IST   T.F.T.T.GLE   ITP:     IST   T.F.T.GLE   ITP:     IST   T.T.F.T.GLE   ITP:     IST   T.T.F.T.GLE   ITP:	Av C Cp Cr MI Mp N O Ps Sco Scy So Te Tp	Anabaona variabilis Cyanothece Cyanophora paradoxa Chlamydomonas reinhardtii Mastigocladus laminosus Micromonas pusilla Nostoc Oscillatoria Picea sitchensis Synechococcus Synechococcus Synechocostis Spinacia oleracea Thermosynechococcus elongatus Thalassiosira pseudonana

#### Figure 4

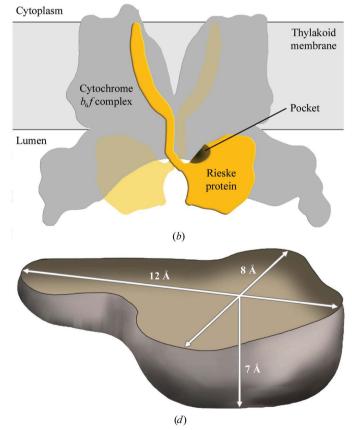
Sequence alignment of PetCs from various species. Apart from the two thermophilic cyanobacteria *T. elongatus* BP-1 and *M. laminosus*, all species are mesophilic. Fully, highly and relatively conserved residues among the cyanobacterial species are shaded dark green, green and light green, respectively. Alignments of PetC from eukaryotes (shaded pink) and PetC2 from cyanobacteria (shaded blue) are also shown for comparison [the alternative Rieske protein of *Anabaena variabilis* ATCC 29413 is part of the PetC4 group (Schneider & Schmidt, 2005; Arnold, 2001), which shows a high sequence similarity to the PetC2 group]. The following symbols are used in the alignments: X for residues of the pocket wall, an asterisk for residues in the protruding loop next to the pocket and  $\Delta$  for residues involved in binding of Wat202. Sequence alignment was performed using *ClustalW2* (Larkin *et al.*, 2007).

#### 3.4. Structurally important water molecules

The crystal structure of TePetC contained 85 water molecules (Fig. 2c), of which three have been assigned as internal water molecules (Wat202, Wat204 and Wat208). These water molecules are completely surrounded by the protein matrix and might be important for both protein structure and function. Notably, the  $F_o - F_c$  map revealed an unusual strong electron-density peak at the position corresponding to Wat202, which could indicate the presence of a metal ion at this site (Fig. 3c). Calcium and potassium could be excluded by anomalous difference Fourier maps at long wavelengths ( $\lambda = 1.73$  and 1.75 Å) and atomic emission spectroscopy of the purified protein (data not shown). Also, owing to the coordination geometry of Wat202 (Harding, 2002, 2006), a sodium ion is very unlikely at this position. In order to find differences in the internal water molecules in structures from higher



plants and prokaryotes, our high-resolution prokaryotic structure was compared with the corresponding eukaryotic structure from spinach (PDB entry 1rfs; Carrell et al., 1997). Owing to their low resolution, water molecules could not be resolved in other PetC crystal structures (Baniulis et al., 2009; Yamashita et al., 2007; Stroebel et al., 2003) and hence they did not help in this comparison. Our analysis showed that Wat204 and Wat208 are (with small deviations) coordinated by mainchain atoms. Wat204 (for simplicity, TePetC numbering is used) does not exist in the SoPetC crystal structure. In the TePetC structure Wat204 is coordinated by the backbone carbonyl O atom of Ile64 and the amine group of Leu161, whereas the corresponding atoms in the SoPetC structure are directly connected by a hydrogen bond. Although the presence or absence of this internal water molecule is not correlated with strong structural changes, it may have an impact on protein stability, as has been shown for many other proteins (see, for example, Luntz et al., 1989; Helms, 2007; Takano et al., 1997). The third internal water molecule in the TePetC structure, Wat202, is coordinated by the side chains of Tyr69 and Asp77 and the main-chain carbonyl of Val91. Surprisingly, in other cyanobacterial PetC1s Tyr69 is usually replaced by Phe (Fig. 4), which is unable to form a hydrogen bond to the water molecule (Fig. 3c). Assuming that at least three hydrogen bonds are necessary to bury a water molecule



#### Figure 5

View of the putative ligand-binding site. (a) Hydrogen-bonding network of the water cluster filling the pocket. (b) Model of the dimeric cytochrome  $b_{6f}$  complex in the thylakoid membrane with the Rieske protein (yellow) and its pocket (highlighted as a dark spot). (c) Surface representation of the pocket, with aromatic (Phe, Trp and Tyr; green), hydrophobic (Ala, Cys, Ile, Leu, Pro and Val; light green) and hydrophilic residues (at the edge, *i.e.* Asp87, Thr89 and Asn105; pale red) indicated. (d) Schematic model with the approximate dimensions of the pocket.

(Rashin *et al.*, 1986), this replacement will result in the loss of this internal water molecule. In contrast, in SoPetC Tyr69 is replaced by Trp, which also enables the binding of a water molecule (Fig. 3*c*). Since eukaryotic PetCs generally contain Trp69 or Tyr69 and Asp77 (Fig. 4), the internal water molecule Wat202 should be fully conserved in eukaryotes. The superposition of prokaryotic and eukaryotic PetCs (Fig. 3*a*) indicates a significant role of Wat202 in the conformation of residues Tyr69–Asp77. Also, the complete conservation of Wat202 in eukaryotes suggests a function of this internal water molecule within the eukaryotic cytochrome  $b_6 f$  complex.

#### 3.5. The surface pocket as a putative ligand-binding site

The protein surface of TePetC revealed a deep pocket which is located at the junction of the cluster-binding subdomain and the large subdomain, and is filled by a cluster of water molecules (Fig. 5a). This pocket was also detected by the metaPocket software for the identification of ligand-binding sites (Huang, 2009). In the structure of the whole cytochrome  $b_6 f$  complex the pocket is oriented towards the membrane surface on the lumen side (Fig. 5b). It most probably forms a ligand-binding site which is highly hydrophobic inside (owing to residues Ile84, Ala149, Val151, Phe163, Pro165, Trp178 and Trp179) and hydrophilic at the edge (owing to residues Asp87, Thr89 and Asn105) (Figs. 5a, 5c and 5d). These residues are completely or highly conserved in PetC1 (Fig. 4). A buried cluster of water molecules as seen in Fig. 5(a) may be fixed by the hydrophilic interior of this pocket. It is possible that this water easily dissociates upon complex formation with a ligand. The fact that mainly C-terminal residues (Trp166–Thr180) contribute to the formation of this pocket may account for its absence and presence in respiratory-type (Iwata et al., 1996; Hunsicker-Wang et al., 2003) and b<sub>6</sub>f-type Rieske proteins, respectively (Baniulis et al., 2009; Yamashita et al., 2007; Stroebel et al., 2003; Carrell et al., 1997). Also, the presence of an empty pocket in all available crystal structures of cytochrome  $b_6 f$  complexes indicates a transient and/or weak interaction with the potential binding partner. The discovery of a membrane-facing hydrophobic pocket suggests the binding of a lipophilic ligand, which may be a plastoquinone, chlorophyll or carotenoid.

Asp87 (Fig. 5*a*), which is located at the edge of the pocket as part of a protruding loop (Fig. 2*c*), is also a putative interaction site with the hinge region or another transmembrane element of the cytochrome  $b_6f$  complex. Interestingly, this residue and two other well conserved residues of this loop, Asn82 and Lys85 (Fig. 2*c*), are replaced in PetC2 and PetC4 (Fig. 4). This may have a serious impact on the ligand-binding properties of the pocket.

#### 4. Conclusions

The Rieske protein TePetC is a key subunit of the cytochrome  $b_6f$  complex from the thermophilic cyanobacterium *T. elongatus*. In this study, we overexpressed the soluble domain of TePetC in its redox-active holo form followed by

detailed spectroscopic and crystallographic characterizations. Owing to the high-resolution (2.0 Å) crystal structure, three internal water molecules with potential impact on the stability and function of the protein could be resolved for the first time for cyanobacterial PetC. While one of them is completely conserved in all known eukaryotic PetCs, it is missing in all other known cyanobacterial PetCs. Also, the conformation of loop Tyr69–Asp77 and the presence of a *cis*-peptide bond at Gly142-Pro143 confirm the similarities between TePetC and eukaryotic PetCs. The fact that protein structures are usually more conserved than amino-acid sequences (Holm & Sander, 1996) indicates that higher plant chloroplasts may originate from a cyanobacterium closely related to *T. elongatus*.

We also discovered a putative hydrophobic ligand-binding pocket on the TePetC protein surface which may regulate electron transport from plastoquinone to cytochrome *f*. While previous studies have indicated the importance of conformational changes ('tethered diffusion') in this step (Yan & Cramer, 2003; Yu *et al.*, 2008), it is still unclear how this sensing mechanism could be triggered.

In conclusion, the well resolved structure of the thermophilic cyanobacterial Rieske protein reported in this study may stimulate new structure-based approaches to clarify both the evolution and the regulation of electron-transport processes within the cytochrome  $b_6 f$  complex.

We gratefully acknowledge the critical input of and careful reading of the manuscript by Dr Gabor Bernat and Dr Sascha Rexroth (RUB). We also greatly appreciate the help of the beamline staff at SPring-8 and the financial support of the Industrialized Countries Instrument Education Cooperation Programme (ICI-ECP), which enabled a stay of SV at Kyoto University. This work was also supported by Grants in Aid for Scientific Research from MEXT of Japan (to KT and KM).

#### References

- Arnold, M. (2001). Doctoral dissertation. University of Regensburg, Germany.
- Baniulis, D., Yamashita, E., Whitelegge, J. P., Zatsman, A. I., Hendrich, M. P., Hasan, S. S., Ryan, C. M. & Cramer, W. A. (2009). J. Biol. Chem. 284, 9861–9869.
- Baniulis, D., Yamashita, E., Zhang, H., Hasan, S. S. & Cramer, W. A. (2008). Photochem. Photobiol. 84, 1349–1358.
- Bernát, G. & Rögner, M. (2011). Bioenergetic Processes of Cyanobacteria, edited by G. A. Peschek, C. Obinger & G. Renger, pp. 573–606. Dordrecht: Springer.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Acta Cryst. D54, 905–921.
- Carrell, C. J., Schlarb, B. G., Bendall, D. S., Howe, C. J., Cramer, W. A. & Smith, J. L. (1999). *Biochemistry*, **38**, 9590–9599.
- Carrell, C. J., Zhang, H., Cramer, W. A. & Smith, J. L. (1997). Structure, 5, 1613–1625.
- Chi, Y.-I., Huang, L.-S., Zhang, Z., Fernández-Velasco, J. G. & Berry, E. A. (2000). *Biochemistry*, **39**, 7689–7701.
- Crofts, A. R., Guergova-Kuras, M., Kuras, R., Ugulava, N., Li, J. & Hong, S. (2000). *Biochim. Biophys. Acta*, **1459**, 456–466.

Veit et al. • Thermophilic cyanobacterial b<sub>6</sub>f-type Rieske protein **1407** 

- De la Mora, E., Lovett, J. E., Blanford, C. F., Garman, E. F., Valderrama, B. & Rudino-Pinera, E. (2012). *Acta Cryst.* D68, 564–577.
- DeLano, W. L. (2002). PyMOL. http://www.pymol.org.
- Freskgård, P. O., Mårtensson, L. G., Jonasson, P., Jonsson, B. H. & Carlsson, U. (1994). *Biochemistry*, 33, 14281–14288.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.
- Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999). *Protein Sci.* 8, 1843–1849.
- Gubernator, B., Seidler, A., Rögner, M. & Szczepaniak, A. (2003). Protein Expr. Purif. 29, 8–14.
- Harding, M. M. (2002). Acta Cryst. D58, 872-874.
- Harding, M. M. (2006). Acta Cryst. D62, 678-682.
- Helms, V. (2007). ChemPhysChem, 8, 23-33.
- Henderson, R. (1990). Proc. R. Soc. London Ser. B, 241, 6-8.
- Holm, L. & Sander, C. (1996). Science, 273, 595-603.
- Holton, B., Wu, X., Tsapin, A. I., Kramer, D. M., Malkin, R. & Kallas, T. (1996). *Biochemistry*, **35**, 15485–15493.
- Huang, B. (2009). OMICS, 13, 325-330.
- Hunsicker-Wang, L. M., Heine, A., Chen, Y., Luna, E. P., Todaro, T., Zhang, Y. M., Williams, P. A., McRee, D. E., Hirst, J., Stout, C. D. & Fee, J. A. (2003). *Biochemistry*, 42, 7303–7317.
- Iwata, S., Saynovits, M., Link, T. A. & Michel, H. (1996). Structure, 4, 567–579.
- Kallas, T. (2011). Photosynthesis: Plastid Biology, Energy Conversion and Carbon Assimilation, edited by J. J. Eaton-Rye, B. C. Tripathy & T. D. Sharkey, pp. 501–560. Dordrecht: Springer.
- Knapp, S., Mattson, P. T., Christova, P., Berndt, K. D., Karshikoff, A., Vihinen, M., Smith, C. I. & Ladenstein, R. (1998). *Proteins*, **31**, 309–319.
- Kumar, S., Tsai, C.-J., Ma, B. & Nussinov, R. (2000). J. Biomol. Struct. Dyn. 17, 79–85.
- Kurisu, G., Zhang, H., Smith, J. L. & Cramer, W. A. (2003). *Science*, **302**, 1009–1014.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, 23, 2947–2948.
- Link, T. A. (2001). Handbook of Metalloproteins, edited by A. Messerschmidt, R. Huber, T. L. Poulos & K. Weighardt, pp. 518– 531. New York: Wiley.
- Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.

- Luntz, T. L., Schejter, A., Garber, E. A. & Margoliash, E. (1989). Proc. Natl Acad. Sci. USA, 86, 3524–3528.
- Martinez, S. E., Huang, D., Ponomarev, M., Cramer, W. A. & Smith, J. L. (1996). *Protein Sci.* 5, 1081–1092.
- McRee, D. E. (1999). J. Struct. Biol. 125, 156-165.
- Murray, J. W., Garman, E. F. & Ravelli, R. B. G. (2004). J. Appl. Cryst. 37, 513–522.
- Nakamura, Y. et al. (2002). DNA Res. 9, 123-130.
- Otwinowski, Z. & Minor, W. (1997). Methods Enzymol. 276, 307–326.
- Pace, C. N. (2000). Nature Struct. Biol. 7, 345-346.
- Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000). Nature Struct. Biol. 7, 380–383.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Nature Struct. Biol. 6, 458–463.
- Rashin, A. A., Iofin, M. & Honig, B. (1986). *Biochemistry*, **25**, 3619–3625.
- Schneider, D., Berry, S., Volkmer, T., Seidler, A. & Rögner, M. (2004). J. Biol. Chem. 279, 39383–39388.
- Schneider, D., Jaschkowitz, K., Seidler, A. & Rögner, M. (2000). Indian J. Biochem. Biophys. 37, 441–446.
- Schneider, D. & Schmidt, C. L. (2005). *Biochim. Biophys. Acta*, **1710**, 1–12.
- Schneider, D., Skrzypczak, S., Anemüller, S., Schmidt, C. L., Seidler, A. & Rögner, M. (2002). J. Biol. Chem. 277, 10949–10954.
- Stroebel, D., Choquet, Y., Popot, J. L. & Picot, D. (2003). *Nature* (*London*), **426**, 413–418.
- Summerfield, T. C., Toepel, J. & Sherman, L. A. (2008). *Biochemistry*, **47**, 12939–12941.
- Takano, K., Funahashi, J., Yamagata, Y., Fujii, S. & Yutani, K. (1997). J. Mol. Biol. 274, 132–142.
- Taylor, T. J. & Vaisman, I. I. (2010). BMC Struct. Biol. 10, S5.
- Tsunoyama, Y., Bernát, G., Dyczmons, N. G., Schneider, D. & Rögner, M. (2009). J. Biol. Chem. 284, 27875–27883.
- Vagin, A. & Teplyakov, A. (2010). Acta Cryst. D66, 22-25.
- Winn, M. D. et al. (2011). Acta Cryst. D67, 235-242.
- Yamaoka, T., Satoh, K. & Katoh, S. (1978). Plant Cell Physiol. 19, 943–954.
- Yamashita, E., Zhang, H. & Cramer, W. A. (2007). J. Mol. Biol. 370, 39–52.
- Yan, J. & Cramer, W. A. (2003). J. Biol. Chem. 278, 20925–20933.
- Yano, J., Kern, J., Irrgang, K. D., Latimer, M. J., Bergmann, U., Glatzel, P., Pushkar, Y., Biesiadka, J., Loll, B., Sauer, K., Messinger, J., Zouni, A. & Yachandra, V. K. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 12047–12052.
- Yu, C.-A., Cen, X., Ma, H.-W., Yin, Y., Yu, L., Esser, L. & Xia, D. (2008). Biochim. Biophys. Acta, 1777, 1038–1043.