Structural and functional characterization of IdiA/FutA (Tery_3377), an iron binding protein from the ocean diazotroph *Trichodesmium erythraeum*

Despo Polyviou[‡,1], Moritz M. Machelett[§,‡,1], Andrew Hitchcock[‡,1,2], Alison J. Baylay[‡], Fraser MacMillan[¶], C. Mark Moore[‡], Thomas S. Bibby[‡] and Ivo Tews[§,3]

From the [‡]Ocean and Earth Sciences, National Oceanography Centre Southampton, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK, the [§]Department of Biological Sciences, Faculty of Natural and Environmental Sciences Life Sciences, Building 85, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK, and the [¶]School of Chemistry, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK [1]The first authors should be regarded as Joint First Authors. [2]Present address: Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK.

Running title: *Structure and Function of Trichodesmium Tery_3377*

To whom correspondence should be addressed: [3] Ivo Tews, Tel.: +44 (0)23-8059-4415; E-mail: Ivo.Tews@soton.ac.uk.

Keywords: cyanobacteria, ABC transporter, crystal structure, microscopy, *Trichodesmium*, iron, IdiA, FutA, diazotroph, iron deficiency

ABSTRACT

Atmospheric nitrogen fixation by photosynthetic cyanobacteria (diazotrophs) strongly influences oceanic primary production and in turn affects global biogeochemical cycles. Species of the genus *Trichodesmium* are major contributors to marine diazotrophy, accounting for a significant proportion of the fixed nitrogen in tropical and subtropical oceans. However, *Trichodesmium* spp. are metabolically constrained by the availability of iron, an essential element for both the photosynthetic apparatus and the nitrogenase enzyme. Survival strategies in low-iron environments are typically poorly characterized at the molecular level, as these bacteria are recalcitrant to genetic manipulation. Here, we studied a homolog of the iron deficiency-induced A (IdiA)/ferric uptake transporter A (FutA) protein, Tery_3377, which has been used as an *in situ* iron-stress biomarker. IdiA/FutA has an ambiguous function in cyanobacteria, with its homologs hypothesized to be involved in distinct processes depending on their cellular localization. Using signal sequence fusions to GFP and heterologous expression in the model cyanobacterium *Synechocystis* sp. PCC 6803 we show that Tery_3377 is targeted to the periplasm by the twin-arginine translocase and can complement the deletion of the native *Synechocystis* ferric-iron ABC transporter periplasmic binding protein (FutA2). EPR spectroscopy revealed that purified recombinant Tery_3377 has specificity for iron in the Fe^{3+} state, and an X-ray crystallography determined structure uncovererd a functional iron substrate-binding domain, with Fe^{3+} penta-coordinated by protein and buffer ligands. Our results support assignment of Tery_3377 as a functional FutA subunit of an Fe^{3+} ABC-transporter, but do not rule out that it also has dual IdiA function.

Photosynthetic organisms have a higher requirement for iron than heterotrophs because of the dependency of multiple components of the photosynthetic electron transport chain on this essential element (1, 2). This requirement is amplified in nitrogen fixing cyanobacteria (diazotrophs) as Fe is also required by the nitrogenase enzyme (3, 4, 5). Therefore, marine diazotrophic cyanobacteria species such as the globally significant *Trichodesmium* have developed diverse strategies that permit survival at the fluctuating or chronically depleted iron concentrations typically

encountered in the surface waters of the tropical and subtropical oceans (5, 6, 7, 8, 9). *Trichodesmium* accounts for approximately half of the total marine nitrogen fixation (10), supplying both biologically available nitrogen and fixed carbon to oligotrophic ecosystems (11). Physiological experiments suggest *Trichodesmium* can acquire Fe from diverse sources (12, 7) and analysis of its genome reveals homologs to Fe-uptake systems, such as FeoAB for ferrous (Fe^{2+}) iron uptake and the periplasmic ferric hydroxamate binding protein FhuD (5). While 'omic' studies and similarity to homologous systems in other bacteria can be used to infer function, the inability to genetically engineer *Trichodesmium* means no Fe uptake mechanisms have been directly characterized *in vivo*, limiting our understanding of how this important species has adapted to acquire Fe from its environment.

Ubiquitous in cyanobacterial genomes and present in *Trichodesmium* are homologs of a protein variably referred to in the literature as IdiA (iron deficiency induced) or FutA (ferric uptake transporter). IdiA/FutA-like proteins are synthesized under Fe limitation in marine and freshwater cyanobacteria (13, 14), and as such are commonly exploited as *in situ* markers of Fe stress (15, 16, 17). Despite this, functional assignment of these proteins is presently ambiguous. homologs can have an intracellular localization (18) and are proposed to be associated with the cytoplasmic side of the thylakoid membrane to protect the acceptor side of photosystem II (PSII) against oxidative damage during Fe stress (18, 19). However, the same protein is also homologous to ferric binding proteins (FBPs) of ATP-binding cassette (ABC) transporters, found in some pathogenic bacteria and several cyanobacteria (20, 21, 22).

In the model freshwater cyanobacterium *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*), there are two characterized paralogues of IdiA/FutA, referred to as FutA1 (Slr1295) and FutA2 (Slr0513). FutA1 has a predominantly intracellular localization, associating with the thylakoid membrane, although it has also been found in the cytoplasmic membrane. In contrast, the FutA2 protein is a periplasmic substrate binding protein (SBP) associated with partner FutB and FutC subunits to form an active Fe^{3+} uptake ABC transporter (22, 23, 24, 25, 26).

Trichodesmium erythraeum IMS101 encodes only one IdiA/FutA homolog in its genome, Tery_3377 (17). Increased expression/production under Fe stress has been reported at both the transcriptional (3, 17), and protein levels (5, 13). Tery_3377 is therefore thought to play an important role in iron homeostasis.

In this study we investigated the role of Tery_3377 to understand whether it is performing a periplasmic function (FutA2-like) or an intracellular function (FutA1-like). We produced Tery_3377 in *futA1* and *futA2* mutants in the genetically tractable model cyanobacterium *Synechocystis* sp. PCC 6803 testing for complementation of mutant phenotypes. We also used signal peptide fusions to GFP to give insight into the cellular localization of Tery_3377. Further, we overproduce and purify recombinant Tery_3377 and confirm ferric iron (Fe^{3+}) coordination by EPR spectroscopy and structure determination, classifying Tery_3377 as a class IV Fe substrate binding domains (27). This functional and structural investigation of Tery_3377 helps to characterize its role in the iron-stress physiology of a globally significant microbe and suggests a potential need to reassess the roles of FutA/IdiA family proteins more generally.

RESULTS

Insights from sequence analysis

The protein sequence of the Tery_3377 protein is 58% and 57% identical to *Synechocystis* FutA1 and FutA2, respectively (Supplementary Information, Table S1). In *Synechocystis*, FutA1 and FutA2 are 52% identical to one another (21), yet appear to have distinct cellular function. Functional assignment based on primary sequence comparisons is therefore not possible. In addition, like the *Synechocystis futA1* and *futA2* genes, the Tery_3377 gene is located at an independent locus to genes encoding putative Fut ABC-transporter permease (FutB/Tery_3223) and ATPase subunits (FutC/Tery_3222), making functional inference from genomic context impossible.

Phylogenetic analysis was carried out to investigate grouping of Tery_3377 with IdiA/FutA homologs (Figure 1). A multiple sequence alignment was constructed from Tery_3377, FutA1 and FutA2, along with homologs from range of cyanobacterial

species, and this was used to generate a maximum likelihood phylogenetic tree. The *Synechocystis* FutA1 and FutA2 proteins are separated into distinct branches within the tree (branch support 0.97, approximate likelihood ratio test). Tery_3377 appears to be more closely related to the intracellular FutA1 protein than to FutA2. To confirm placement of the Tery_3377 protein within the phylogeny, we searched for homologs in other publicly available *Trichodesmium* genomes / metagenomes. No Tery_3377 homolog could be detected in the draft genome of *T. theibautii* H94 (Bioproject PRJNA278659); however this genome is known to be incomplete (28). A homolog was found in a metagenomic dataset from the Bermuda Atlantic Time Series (Bioproject PRJNA336846). This protein (encoded between nucleotides 342 and 1392 on contig 00877) clustered with Tery_3377 (Figure 1).

The Tery_3377 signal peptide targets GFP to the periplasm

To investigate the cellular localization of Tery_3377 a synthetic gene encoding the predicted signal peptide (aa 1-31) and first 4 residues of the mature protein (aa 32-35) fused to super-folder (sf)GFP was expressed in *Synechocystis* under the control of the *psbAII* promoter (Table S1 and Figure S1). For comparison, the predicted *Synechocystis* FutA1 and FutA2 signal peptides were also fused to sfGFP and expressed from the *psbAII* promoter. The *Escherichia coli* TMAO reductase (TorA) signal peptide that has previously been shown to transport GFP/YFP into the periplasm of *Synechocystis* was used as a positive control, and sfGFP alone, i.e. without signal sequence, acted as a marker of cytoplasmic expression (29 , 30).

The strains were visualized by confocal microscopy. Excitation of WT cells at 488 nm does not result in fluorescence emission in the GFP channel at 510-530 nm, but prominent auto-fluorescence from thylakoid membrane localized photosynthetic pigments was observed when emission was monitored at 600-700 nm (Figure 2A), as expected. Fusion of sfGFP to the TorA (Figure 2B) or FutA2 (Figure 2D) signal peptide results in bright fluorescent halos that surround the auto-fluorescence signal in merged images, indicating translocation of GFP across the cytoplasmic membrane to the periplasm. It is perhaps surprising that the FutA1 signal peptide also appears to localize some GFP to the periplasm or cytoplasmic membranes, albeit to a lesser extent than with the the TorA or FutA2 fusions (Figure 2C). Previous studies have suggested that FutA1 is mainly associated with thylakoid membranes, although it has been detected in the plasma membrane (26).

Fusion of the Tery_3377 signal peptide to GFP results in distinct halos (Figure 2E), although this is not as clear as for the TorA and FutA2 signal peptides (Figures 2B and 2D), and also showed similarity to the FutA1 fusion (Figure 2C). We changed the twin arginine residues in the Tery_3377 signal peptide to lysines, which prevented export of GFP and showed that the wild-type Tery_3377 signal peptide is recognized and exported by the twin-arginine translocase (TAT) system (Figures S2). Immunoblots with an anti-GFP antibody identified a protein approximately 3 kDa larger than sfGFP in the stain with the mutated signal peptide, consistent with the presence of the uncleaved signal peptide (Figure S3). In the strain producing just sfGFP without a signal peptide clear fluorescence signal mapping to the cytoplasm was observed (Figure 2F), as expected and as reported previously for YFP (30).

These results indicate that the Tery_3377 signal peptide can translocate GFP to the periplasm using the TAT system. Both of the native *Synechocystis* FutA signal peptides also result in some degree of cytoplasmic membrane or periplasmic localization of GFP. Based on previous work FutA1 is considered to be predominantly intracellular, hence this result is unexpected. It is likely that the mature protein sequences also play a role in cellular localization, but attempts to investigate this in strains in which GFP was fused to the full proteins rather than just the signal peptides were unsuccessful as the strains were either genetically unstable or GFP was cleaved from the protein (data not shown).

Tery_3377 is involved in Fe^{3+} transport across the cell membrane

The signal peptide fusions indicated that Tery_3377 is at least partly localized in the periplasm, presumably functioning as the periplamic binding protein component of a ferric uptake ABC-transporter. To gain further insight into the functional role of Tery_3377, the gene was codon optimized for expression in *Synechocystis* and integrated at the *psbAII* locus in Δ*futA1* and Δ*futA2 Synechocystis* mu-

tants. Production of Tery_3377 was confirmed by immunoblotting with the anti-IdiA antibody (raised against *Synechococcus elongatus* PCC 6302 IdiA), but was not detected using an antibody against *Synechocystis* FutA2 (Figure S4).

The high affinity iron ligand ferrozine (FZ) scavenges Fe^{2+} and cannot penetrate the cell membrane and thus is hypothesized to cause a Fe stress phenotype through reduction of ferrous iron uptake through the FeoAB system (31, 32, 33). This effect would be expected to be more severe in strains where ferric iron uptake is also reduced, such as the Δ*futA2* strain. Experiments were performed in biological triplicate and differences between FZ treated and non-treated cells for each strain were determined for growth and photosynthetic efficiency (Figure 3). A reduction in photosynthetic efficiency (F_v/F_m) is expected under iron stress because of the essential role of iron in the photosynthetic electron transport chain (4, 5).

In the presence of FZ, growth rates were significantly reduced by 13% in the WT ($P < 0.01$) and 15% in the Δ*futA1* strain ($P < 0.05$, Figure 3A) while the Δ*futA2* mutant lacking the Fe^{3+} transporter showed the greatest response, as expected (32% reduction, $P < 0.01$). The reduction in growth rate by FZ in both mutants was not as apparent when Tery_3377 was introduced to complement the knockout strains Δ*futA1* and Δ*futA2*, suggesting that Tery_3377 enables the cells to acquire ferric iron when ferrous iron is bound by FZ. Growth in presence of FZ also caused a reduction in photosynthetic efficiency (F_v/F_m) that was largest for Δ*futA2* (34%). The negative effect on the Δ*futA2* phenotype was again alleviated in the *tery_3377* complementation strain (Figure 3B). In contrast, the reduction in photosynthetic efficiency in the Δ*futA1* knockout is less severe than WT.

Tery_3377 binds ferric (Fe^{3+}) iron

Recombinant Tery_3377 with a C-terminal His-tag was over-produced in *E. coli* and purified in order to investigate its affinity and binding of Fe species. After incubation with Ammonium Fe(II) sulfate under aerobic conditions, the Tery_3377-Fe complex was isolated by size exclusion chromatography to remove unbound Fe. The purified protein-Fe complex was analyzed by Electron Paramagnetic Resonance (EPR) spectroscopy to determine the oxidation state of the bound Fe ligand. EPR spectra revealed an oxidized Fe^{3+} state present in the Tery_3377 sample (Figure 4). In addition to weaker signals between 800 – 1400 Gauss, a sharp and intense signal at about 1500 Gauss (equal to a g-value of 4.3) was clearly observed. This g-value (4.3) is indicative of a high-spin Fe^{3+} ion bound to Tery_3377. The typical fingerprint weaker signals are related to higher order spin transitions of this ferric iron.

Crystal structures of Tery_3377 reveal penta-coordination of Fe^{3+}

To study Fe^{3+} binding by Tery_3377, the structure of the iron-bound protein was determined using X-ray crystallography (Table 1). The overall fold of Tery_3377 assumes a classic small molecule ABC transporter substrate-binding domain fold, with two domains hinged over a substrate-binding cleft. The hinge-region is formed by two short beta-strands and connects the protein domains via a double crossover between the N- (residues 34-133 and 266-349) and C-terminus (134-265). This type of structure places Tery_3377 amongst Class IV Fe-binding proteins (in keeping with 34), and allows for structural classification within Cluster D of the substrate-binding domains (35, 36).

Iron substrate binding domains of different classes differ in Fe ligand coordination, determined by conserved interacting amino acid side chains of the substrate binding domain, often involving tyrosine residues. In Tery_3377 three tyrosine residues Tyr144, Tyr200 and Tyr201 from the C-terminal domain contribute to Fe binding. In the structure of Tery_3377-Ala (PDB: 6G7N), determined at 1.1 Å resolution, a buffer molecule from the crystallization mixture occupies two coordination sites, giving rise to a penta-coordinated Fe assigned as Fe^{3+} (Figure 5A). The geometry observed is trigonal bi-pyramidal, with the buffer molecule alanine carboxyl-group, Tyr174 and Try231 forming the trigonal plane with angles of 108°, 108° and 144°. Tyr230 and the amino-group from the buffer alanine coordinate the axial Fe positions.

Amino acid additives in the crystallization buffers lead to the most ordered and best diffracting crystals. However, Tery_3377 crystallized under a variety of conditions, including ones that lacked amino acid additives. We determined the structure of the protein crystallized with a monosaccha-

ride additive mixture and found it was devoid of an Fe-coordinating ligand in the binding site, with the iron instead coordinated by ordered water molecules (Figure 5B). Comparison of the Tery_3377-Water structure (PDB: 6G7P) to the Tery_3377-Ala complex shows that water molecules W1 and W2 take the positions of the alanine amino and carboxylate groups of the alanine ligand, respectively (Figure 5). Interestingly, in this complex an additional water molecule ligand (W3) is inserted between the coordinating oxygen atoms of W2 and Tyr231. In comparison to Tery_3377-Ala, where the observed angle between Tyr23-1O and Alanine-carboxyl oxygen was measured as 144°, this angle is split into two angles of roughly 73° measured between Tyr231-O, the inserted water molecule W3, and W2. Like in Tery_3377-Ala, Tyr230 and water W1 occupy the axial coordination sites in Tery_3377-Water.

To try and crystalline ligand-free Tery_3377 in order to examine structural changes upon iron binding, we citrate-treated Tery_3377 protein preparations prior to crystallization; citrate is known to chelate Fe and was added to remove all Fe from the protein preparation. The crystal structure (Tery_3377-Cit, PDB: 6G7Q) traps the citrate molecule with bound Fe in the substrate binding cleft, where the Fe is removed approximately 7 Å from the binding site (Figure 6). The Fe binding site (pink sphere in Figure 6) was free of metal, and the significance of this finding in the context of possible siderophore acquisition is discussed below .

DISCUSSION

We have investigated the function and structure of *Trichodesmium* Tery_3377, an Fe stress induced protein (iron deficiency induced A, IdiA) commonly used as an *in situ* marker of Fe stress (15, 16, 17). Phylogenetic analysis (Figure 1) groups Tery_3377 with the *Synechocystis* FutA1 protein, which has a suggested intracellular PSII protective function, rather than the FutA2 protein, a periplasmic protein with a role in Fe-transport. However, owing to the high level of sequence identity of Tery_3377 to both FutA1 and FutA2 proteins, and FutA1/2 proteins to one another, the function of these similar iron-binding proteins may be determined by their cellular localization. By fusion to sfGFP, Tery_3377 is shown to have a functional signal sequence for transport to the periplasm (Figure 2) by the TAT system (Figures S1 and S2). Consistent with this, Tery_3377 complements the ferrozine (FZ) induced Fe stress phenotype of the *Synechocystis* Δ*futA2* deletion strain (Figure 3). These data support functional assignment of _3377 as a subunit of a Fe-uptake ABC transporter.

Expression of Tery_3377 also complements the more subtle phenotype of the Δ*futA1 Synechocystis* mutant in the presence of FZ. This may indicate the possibility for an additional intracellular function in PSII protection, or alternatively result from increased ferric iron uptake by Tery_3377. Since localization data using FutA1 signal peptide GFP fusions resulted in some periplasmic localization (Figure 2)), FutA1 may have a role in iron uptake in addition to its reported intracellular function. Indeed, despite their apparently distinct cellular locations and functions, a *futA1/futA2* double mutant in *Synechocystis* has more severe iron-deficient growth and ferric iron uptake phenotypes than either single mutant, suggesting some degree of redundancy (20). In agreement with this, the phenotype of the double mutant is similar to that of mutants of the single copy Fut ABC-transporter permease (*futB*) and ATPase subunit (*futC*) genes (20). Hence the IdiA/FutA family of proteins may commonly display a dual function that requires a broader reassessment across cyanobacteria.

EPR spectroscopy and structural analysis confirms Tery_3377 binds Fe in the ferric oxidation state, Fe^{3+} (Figure 4). Tery_3377 crystallizes under various conditions, often requiring small-molecule additives. The best diffracting crystals were obtained with an additive mixture of amino acids (Figure 5A). An amino acid ligand, alanine, is identified in the Fe binding cleft; the amino acid is ordered and coordinates the Fe substrate. In contrast, other crystallization conditions, which did not show ligand coordination by organic molecules, gave less well-ordered crystals that diffracted to lower resolution and generally refined to higher R-values, indicating a lower quality of the resulting models (Table 1). We also determined the structure where organic ligands are absent from the Fe binding site, with three water molecules coordinating the Fe, in place of the two coordinating atoms seen with the alanine ligand structure (Figure 5B).

To determine the most likely coordination state of Fe in Tery_3377, we superposed the Tery_3377-Ala structure with a classic closed conformation FutA2 structure (*Synechocystis* PCC6803 FutA2, PDB: 2VP1) (Figure 7A). The coordinating oxygen of the carboxylic acid moiety of the alanine ligand ideally superposes with a tyrosine side chain oxygen, while the amino nitrogen of the alanine ligand superimposes with a nitrogen from an interacting histidine ligand (Figure 7B,C). The two coordinating amino-acids in FutA2 are contributed by a loop that is brought into the vicinity of the Fe binding site by domain closure around the hinge that brings the two lobes of the substrate binding domain closer together (37, 38) (Figure 7A). Interestingly, this motif is conserved in Tery_3377, and such a domain closure, while not observed under any of our crystallization conditions, can be proposed. We conclude that the alanine ligand mimics this Fe binding mode leading to a penta-coordinated state. This coordination state is corroborated by the low-temperature solution EPR spectroscopy, which revealed a strong signal at g=4.3 and less intense higher-order transitions at higher g-values, supporting generic Class IV Fe coordination by Tery_3377 in solution.

The observed binding of organic ligands together with Fe may just be a fortuitous observation, but it might also have biological significance. It could be argued that the high concentrations of alanine present in crystallization mixtures out-competes interactions with the N-terminal Fe-coordination residues His43 and Tyr44, thus leading to preferred crystallization of an open conformation. However, specific buffer molecules such as carbonate have previously been shown to coordinate Fe in substrate binding domains (39). Indeed, the binding site is large enough to fit a variety of amino acids in similar fashion to the shown alanine in Tery_3377-Ala. Hence Fe binding might concur with binding of different, larger molecules in the binding site. Nitrogen containing molecules might enrich in the periplasm, creating a concentration gradient over the outer membrane. Additional support for the plasticity in the binding of Tery_3377 comes from the observation of bound citrate in the binding cleft, clearly indicating the potential for binding of organic compounds (Figure 6). It is interesting to speculate that binding of organic ligands together with iron may represent a mechanism of siderophore recognition.

Summary

This study provides structural and mechanistic insight into a key protein important in iron acquisition and the iron stress response in the globally significant marine diazotroph *Trichodesmium*. These organisms are central to N_2 fixation, a highly significant process in the biogeochemical cycle, and whose biogeography is determined by the availability of iron. Using heterologous expression in the model cyanobacteria *Synechocystis* PCC6803 we provide evidence that Tery_3377, a commonly used iron-limitation biomarker, is localized to the periplasm with a suggested FutA2 function as an iron binding protein of an ABC transporter. However, Tery_3377 also complements deletion of FutA1, which could indicate it also functions in protection of PSII, suggesting a potential dual function of Tery_3377 and more generally the IdlA/Fut family of proteins in cyanobacteria. Structural and spectroscopic characterization of recombinantly produced Tery3̀377 confirm ferric iron binding under aerobic conditions. The structural data further show a remarkable plasticity of the substrate binding cleft, allowing organic molecules and potentially siderophores to bind in addition to Fe^{3+}.

EXPERIMENTAL PROCEDURES

Protein phylogeny analysis

Maximum likelihood phylogenetic analysis of protein sequences of Tery_3377 homologs from *Synechocystis* and selected bacteria was performed to identify possible functional clusters. homologs of the *Trichodesmium* Tery_3377 and two *Synechocystis* proteins FutA1 and FutA2 were identified by searching the NR protein database with BlastP (40). The top ten closest hits were retained for each query sequence, and clustered at 90% identity using CD-HIT (41) to reduce redundancy. For each cluster, one representative was retained for inclusion in the analysis. homologs from *Prochlorococcus marinus* MIT 9301, *Synechococcus elongatus* PCC 6301 and PCC 7942 and *Anabaena* sp. PCC 7120 were also included in the analysis as representatives of important primary producers, and sequences from *Haemophilus influenza* and *Neisseria meningitides* were chosen to represent pathogens.

The online platform Phylogeny.fr (42, 43) was utilized in "a la carte" mode to create an analysis pipeline. Sequences were firstly aligned with T-coffee (v11.00.8cbe486, 44) and the alignment was curated with G-blocks (v0.91b, 45). Subsequently, a maximum likelihood phylogenetic tree was constructed using PhyML v3.1/3.0, (46) and an approximate likelihood ratio test was carried out to compute branch supports (aLRT, 47). Finally, a graphical representation was produced with FigTree v1.4.3 (Institute of Evolutionary Biology, University of Edinburgh).

Growth conditions

Synechocystis strains (Table S2) were routinely grown in BG-11 media (48) supplemented with 5 mM glucose and buffered to pH 8.2 with 10 mM N-17 [tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid (TES)-KOH. Cultures were maintained at 30 °C, shaking at 150 rpm under constant illumination of ca. 50 µM photons m^{-2} s^{-1}. For growth on plates, 1.5% (w/v) agar and 0.3% (w/v) sodium thiosulphate were added. Antibiotics were included as appropriate (Table S2). Ferrozine (FZ) growth experiments were carried out in Fe (0.6 mM $FeCl_3$) YBG-11 media (recipe described by 49) at a FZ final concentration of 200 µM (FZ+) or no added FZ (FZ-). Growth was assessed under photo-autotrophic conditions (no glucose), in triplicate cultures illuminated by 30 µM photons m^{-2} s^{-1} of red light to minimize abiotic photoreduction of Fe in the presence of FZ (50).

Escherichia coli (Table S2) was grown shaking (250 rpm) at 37 °C in LB (Luria-Bertani) broth with 100 µg ml^{-1} ampicillin. Plates were set with 1.5% (w/v) agar.

Photosynthetic physiology

Photosynthetic efficiency of *Synechocystis* strains during FZ growth experiments was measured as F_v/F_m, an estimate of the apparent PSII photochemical quantum efficiency (51), using a FASTtracka™ MkII Fast Repetition Rate fluorometer (FRRf) integrated with a FastAct™ Laboratory system (Chelsea Technologies Group Ltd., Surrey, United Kingdom). Samples were dark adapted for 20 min prior to measurements. Presented results were calculated from the average of three technical replicates for each of three biological replicates.

Generation of *Synechocystis* strains

In this study, (1) pET21a was used for protein production in *E. coli*; (2) pACYC184 was used as a template to amplify the chloramphenicol resistance cassette; (3) pZEO was used as a template to amplify the zeocin resistance cassette; (4) pPD-FLAG was used as delivery vector for the introduction of genes into the *Synechocystis* genome in place of psbAII; and (5) pED151 was used for isolation of sfGFP.

For the generation of signal peptide fusions to sfGFP, DNA sequence corresponding to the predicted (in the case of *futA1*, *futA2*, *Tery_3377*) or known (*Escherichia coli torA*; 52) signal sequence and first 4 amino acid residues of the mature protein (as identified by SignalP v4.1/v3.0 sensitivity and TatP v1.0 servers 53, 54) (Table S1, Figure S1) was amplified from the respective organisms genomic (g)DNA (for details of all primers used in this study see Table S3). Approximately 400 bp immediately preceding the *psbAII* gene was joined to each signal sequence by overlap-extension (OLE)-PCR such that the ATG start codon of the signal sequence replaced the ATG start codon of *psbAII*. Using the same technique the chloramphenicol acetyl transferase (*cat*) gene from pACYC184 (NEB, UK) was inserted between the coding sequence for sfGFP (55) and $\tilde{4}00$ bp of sequence immediately downstream of *psbAII*. Finally, each *psbAII* upstream-signal sequence fragment was joined to the sfGFP-*cat*-*psbAII* downstream fragment, again using OLE-PCR. The psbAII locus is extensively utilized as a 'neutral site' that will not result in polar effects following gene insertion, and is commonly used for expression of recombinant genes (56).

To generate a variant of the Tery_3377 signal peptide in which the twin arginine residues were replaced by lysines (3377ss-KK-GFP) alternative primers were designed to anneal during OLE-PCR and both contained the first 18 bp of the Tery_3377 signal sequence with changed nucleotide sequence between positions 13-18 (AAA AAA in the place of AGA CGA) such that the translated sequence codes for two lysine residues in the place of the two arginine residues of the twin arginine transport (TAT) system.

Synechocystis was transformed as described by 57 with selection on chloramphenicol (8.5 µg ml^{-1}). Genome copies were segregated by repeated streaking on BG11 plates with increas-

ing antibiotic concentration up to 34 µg ml^{-1} and homozygous transformants were verified by PCR and sequencing (Eurofins MWG Operon, Ebersberg, Germany).

For the generation of slr1295 (Δ*futA1*) and slr0513 (Δ*futA2*) *Synechocystis* deletion mutants, the antibiotic resistance cassette for zeocin (ZeoR) from pZEO (58) or the *cat* gene were used. PCR products consisting of the sequence upstream of the *futA1*/*futA2* genes and the beginning of the open reading frame (ORF), or the end of the ORF and downstream flanking DNA, were amplified from *Synechocystis* gDNA. Using (OLE)-PCR, ZeoR or *cat* was inserted between the upstream and downstream PCR products to create a linear mutagenesis fragment. For *slr1295* (*futA1*) the entire gene apart from the first 22 bp was replaced with the zeo cassette. For *slr0513* (*futA2*) all but the first 66 bp and last 41 bp of the gene were replaced by the cat cassette. The deletion constructs were introduced into WT *Synechocystis* as described above, with selection and segregation on zeocin (2.5-20 µg ml^{-1}) or chloramphenicol (8.5-34 µg ml^{-1}). Full segregation of mutant strains was confirmed by PCR.

For expression of *tery_3377* in *Synechocystis*, the gene was codon optimized (Integrated DNA Technologies Inc, Leuven, Belgium) and cloned into the *Nde*I/*Bgl*II sites of pPD-FLAG (56). Transformation of *Synechocystis* with the constructed plasmid integrates *tery_3377* into the *Synechocystis* genome such that expression is driven from the *psbAII* promoter, and confers kanamycin resistance allowing selection of transformed cells using increasing concentrations of kanamycin (10-40 µg ml^{-1}). Segregation of transformants at the *psbAII* locus was confirmed by PCR and the sequence of the gene was verified by sequencing, as above.

Fluorescence microscopy

Synechocystis cells producing sfGFP fusion proteins were visualised using a Leica SP5 LSCM confocal microscope (Leica Microsystems, Wetzlar, Germany) or a axioscope Z plus epifluorescence microscope (Zeiss, Oberkochen, Germany). For confocal microscopy, samples were excited at 488 nm and emission was measured at 510-530 nm for sfGFP fluorescence or 600-700 nm for autofluorescence, as described by 59. For visualization of sfGFP by epifluorescence microscopy, excitation was set at 400-490 nm (max 441 nm) and emission was measured at 460-570 nm (max 485 nm).

Electron paramagnetic resonance

Oxidised Tery_3377 (550 µM) was placed in EPR quartz tubes (Wilmad) and shock-frozen in liquid nitrogen. X-band continuous wave (cw) EPR spectra were recorded on a Bruker eleXsys E500 spectrometer using a standard rectangular Bruker EPR cavity (ER4102T) equipped with an Oxford helium flow cryostat (ESR900). The spectrometer operates at X-band (about 9.4 GHz) frequency and uses a 1 mT modulation amplitude (peak to peak) and 2 mW microwave power. Spectra were recorded at cryogenic temperatures (5-6 K). A typical spectrum was recorded in about 15-30 mins.

Protein production, crystallisation and structure determination

To produce C-terminally His_6-tagged Tery_3377 in *E. coli*, the open reading frame (minus the putative sequence encoding the N-terminal signal peptide and stop codon) was amplified from the *Trichodesmium* genome. The resulting PCR product was digested and cloned into the *Nde*I/*Xho*I sites of pET21a(+) (Novagen). The sequence verified plasmid was introduced into *E. coli* BL21(DE3) and production of the recombinant protein was induced in cells grown to an OD_{600} of 0.6 by addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG), followed by growth for a furher 4 hours. Cells were harvested by centrifugation (6000 x g, 20 min, 4 °C), resuspended in binding buffer (50 mM Tris, 200 mM NaCl, 5% Glycerol, 20 mM Imodazole), broken (sonication for 10 sec on, 30 sec off; 2 min process time) and the cell-free extract was isolated as the supernatant following centrifugation (200000 x g, 45 min, 4 °C). His-tagged Tery_3377 was purified by affinity chromatography (1 ml HisTrap HP columns, GE Healthcare) and size-exclusion chromatography (HiLoad 16/60 Superdex 200, GE Healthcare).

Tery_3377 crystals were obtained in space group $P2_1$ with two molecules in the asymmetric unit with slightly different packing between Fe bound and Fe free forms of Tery_3377. Hanging drop vapor diffusion used conditions based on or optimized from the Morpheus screen (Molecular Dimensions). The reservoir composition leading to Tery_3377-Ala crystals was 0.1 M Amino acids Mix,

0.1 M Buffer System 2, pH 7.5, 50 % Precipitant Mix 1. Tery_3377-Ala data were collected at the Diamond Light Source beamline I02 to a resolution of 1.1 Å. Tery_3377-Water crystals were obtained with a reservoir composition of 0.12 M Monosaccharides Mix, 0.1 M Buffer System 2, pH 7.5, 50 % Precipitant Mix 1. Tery_3377-Water data were collected at the Diamond Light Source beamline I02 to a resolution of 1.5 Å. Tery_3377-Cit crystals were obtained with a reservoir composition of 0.1 M Carboxylic acids Mix, 0.1 M Buffer System 2, pH 7.5, 50 % Precipitant Mix 1. Tery_3377-Cit data were collected at ESRF beamline ID30B to resolution of 1.2 Å.

The XDS package (60) was used to index, integrate and scale diffraction data. Structures have been determined by molecular replacement with Phaser (61) using coordinates of apo-FutA2 from *Synechocystis* (PDB entry: 2VOZ, 21) and were further refined with REFMAC5 (62). During refinement, hydrogens were added in their riding positions, and refinement was iterated with water-building and model correction in Coot (63).

Immunoblotting

Synechocystis cells were resuspended in 100 µl of 10 mM Tris-HCl pH 7.3 and heated for 5 minutes at 100 °C followed by centrifugation at 10,000 rpm for 5 minutes at 4 °C. Samples were analyzed by SDS-PAGE on 4–12% NuPAGE acrylamide gradient gels (Life Technologies). Protein was blotted onto nitrocellulose membrane, which was blocked for 1 hr in Tris-buffered saline with Tween 20 (TSB-T) and 2% ECL Advance blocking reagent (GE Healthcare). Antibody incubations (1 h) and membrane washes were performed in TBS-T. Anti-IdiA and anti-Slr0513 (FutA2) antisera were a kind gift from E. Pistorius (64, 24), were used at concentrations of 1:5000 and 1:10000 respectively. Following application of ECL Advance detection reagent (GE Healthcare) chemiluminescence was visualized using a VersaDocTM CCD imager (Bio-Rad).

Acknowledgments: We thank Chris Holes of the macromolecular crystallization facility at the University of Southampton, staff at the Diamond Light Source and the European Synchrotron Radiation Facility for access and excellent user support through proposals MX11945 and MX1848, respectively. We thank staff at the Biomedical Imaging Unit (BIU), Southampton General Hospital for assistance with confocal microscopy.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: A.H. and T.S.B. conceived the study. D.P., M.M.M, A.H., T.S.B. and I.T. designed the experiments. D.P., M.M.M., A.H., F.M. and I.T. performed the experiments. All authors interpreted results and/or performed data analysis. D.P., M.M.M., A.H., T.S.B. and I.T. wrote the paper, which was edited and approved by all authors.

REFERENCES

1. Raven, J. A., Evans, M. C., and Korb, R. E. (1999) The role of trace metals in photosynthetic electron transport in o2-evolving organisms, *Photosynth. Res.* **60**(2-3), 111–150.
2. Blaby-Haas, C. E. and Merchant, S. S. (2012) The ins and outs of algal metal transport, *Biochimica et Biophys. Acta (BBA)-Molecular Cell Res.* **1823**(9), 1531–1552.
3. Shi, T., Sun, Y., and Falkowski, P. G. (2007) Effects of iron limitation on the expression of metabolic genes in the marine cyanobacterium trichodesmium erythraeum ims101, *Environ. Microbiol.* **9**(12), 2945–2956.
4. Richier, S., Macey, A. I., Pratt, N. J., Honey, D. J., Moore, C. M., and Bibby, T. S. (2012) Abundances of iron-binding photosynthetic and nitrogen-fixing proteins of trichodesmium both in culture and in situ from the north atlantic, *PLOS One* **7**(5), e35571.
5. Snow, J. T., Polyviou, D., Skipp, P., Chrismas, N. A., Hitchcock, A., Geider, R., Moore, C. M., and Bibby, T. S. (2015) Quantifying integrated proteomic responses to iron stress in the globally important marine diazotroph trichodesmium, *PLOS One* **10**(11), e0142626.
6. Moore, J. and Braucher, O. (2007) Observations of dissolved iron concentrations in the world ocean: implications and constraints for ocean biogeochemical models, *Biogeosciences Discuss.* **4**(2), 1241–1277.
7. Rubin, M., Berman-Frank, I., and Shaked, Y. (2011) Dust-and mineral-iron utilization by the marine dinitrogen-fixer trichodesmium, *Nat. Geosci.* **4**(8), 529–534.
8. Hansel, C. M., Buchwald, C., Ossolinski, J., Dyhrman, S., Van Mooy, B., Polyviou, D., et al. (2016) Dynamics of extracellular superoxide production by trichodesmium colonies from the sargasso sea, *Limnol. Oceanogr.* **61**(4), 1188–1200.
9. Falkowski, P. G. (1997) Evolution of the nitrogen cycle and its influence on the biological sequestration of co2 in the ocean, *Nat.* **387**(6630), 272.
10. Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B., and Carpenter, E. J. (1997) Trichodesmium, a globally significant marine cyanobacterium, *Sci.* **276**(5316), 1221–1229.
11. Westberry, T. K. and Siegel, D. A. (2006) Spatial and temporal distribution of trichodesmium blooms in the world's oceans, *Glob. Biogeochem. Cycles* **20**(4).
12. Achilles, K. M., Church, T. M., Wilhelm, S. W., Luther, G. W., and Hutchins, D. A. (2003) Bioavailability of iron to trichodesmium colonies in the western subtropical atlantic ocean, *Limnol. Oceanogr.* **48**(6), 2250–2255.
13. Webb, E. A., Moffett, J. W., and Waterbury, J. B. (2001) Iron stress in open-ocean cyanobacteria (synechococcus, trichodesmium, andcrocosphaera spp.): Identification of the idia protein, *Appl. Environ. Microbiol.* **67**(12), 5444–5452.
14. Singh, A. K., McIntyre, L. M., and Sherman, L. A. (2003) Microarray analysis of the genome-wide response to iron deficiency and iron reconstitution in the cyanobacterium synechocystis sp. pcc 6803, *Plant Physiol.* **132**(4), 1825–1839.
15. Webb, E. A., Jakuba, R. W., Moffett, J. W., and Dyhrman, S. T. (2007) Molecular assessment of phosphorus and iron physiology in trichodesmium populations from the western central and western south atlantic, *Limnol. Oceanogr.* **52**(5), 2221–2232.
16. Rivers, A. R., Jakuba, R. W., and Webb, E. A. (2009) Iron stress genes in marine synechococcus and the development of a flow cytometric iron stress assay, *Environ. Microbiol.* **11**(2), 382–396.
17. Chappell, P. D. and Webb, E. A. (2010) A molecular assessment of the iron stress response in the two phylogenetic clades of trichodesmium, *Environ. Microbiol.* **12**(1), 13–27.
18. Michel, K.-P. and Pistorius, E. K. (2004) Adaptation of the photosynthetic electron transport chain in cyanobacteria to iron deficiency: the function of idia and isia, *Physiol. Plantarum* **120**(1), 36–50.
19. Exss-Sonne, P., Tölle, J., Bader, K. P., Pistorius, E. K., and Michel, K.-P. (2000) The idia protein of synechococcus sp. pcc 7942 functions in protecting the acceptor side of photosystem ii under

oxidative stress, *Photosynth. Res.* **63**(2), 145–157.

20. Katoh, H., Hagino, N., and Ogawa, T. (2001) Iron-binding activity of futa1 subunit of an abc-type iron transporter in the cyanobacterium synechocystis sp. strain pcc 6803, *Plant Cell Physiol.* **42**(8), 823–827.
21. Badarau, A., Firbank, S. J., Waldron, K. J., Yanagisawa, S., Robinson, N. J., Banfield, M. J., and Dennison, C. (2008) Futa2 is a ferric binding protein from synechocystis pcc 6803, *J. Biol. Chem.* **283**(18), 12520–12527.
22. Kranzler, C., Lis, H., Finkel, O. M., Schmetterer, G., Shaked, Y., and Keren, N. (2014) Coordinated transporter activity shapes high-affinity iron acquisition in cyanobacteria, *The ISME J.* **8**(2), 409–417.
23. Fulda, S., Huang, F., Nilsson, F., Hagemann, M., and Norling, B. (2000) Proteomics of synechocystis sp. strain pcc 6803, *Eur. J. Biochem.* **267**(19), 5900–5907.
24. Tölle, J., Michel, K.-P., Kruip, J., Kahmann, U., Preisfeld, A., and Pistorius, E. K. (2002) Localization and function of the idia homologue slr1295 in the cyanobacterium synechocystis sp. strain pcc 6803, *Microbiol.* **148**(10), 3293–3305.
25. Waldron, K. J., Tottey, S., Yanagisawa, S., Dennison, C., and Robinson, N. J. (2007) A periplasmic iron-binding protein contributes toward inward copper supply, *J. Biol. Chem.* **282**(6), 3837–3846.
26. Huang, F., Parmryd, I., Nilsson, F., Persson, A. L., Pakrasi, H. B., Andersson, B., and Norling, B. (2002) Proteomics of synechocystis sp. strain pcc 6803 identification of plasma membrane proteins, *Mol. & Cell. Proteomics* **1**(12), 956–966.
27. Wang, S., Ogata, M., Horita, S., Ohtsuka, J., Nagata, K., and Tanokura, M. (2014) A novel mode of ferric ion coordination by the periplasmic ferric ion-binding subunit fbpa of an abc-type iron transporter from thermus thermophilus hb8, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **70**(1), 196–202.
28. Walworth, N., Pfreundt, U., Nelson, W. C., Mincer, T., Heidelberg, J. F., Fu, F., Waterbury, J. B., del Rio, T. G., Goodwin, L., Kyrpides, N. C., et al. (2015) Trichodesmium genome maintains abundant, widespread noncoding dna in situ, despite oligotrophic lifestyle, *Proc. Natl. Acad. Sci.* **112**(14), 4251–4256.
29. Spence, E., Sarcina, M., Ray, N., Møller, S. G., Mullineaux, C. W., and Robinson, C. (2003) Membrane-specific targeting of green fluorescent protein by the tat pathway in the cyanobacterium synechocystis pcc6803, *Mol. Microbiol.* **48**(6), 1481–1489.
30. MacGregor-Chatwin, C., Sener, M., Barnett, S. F., Hitchcock, A., Barnhart-Dailey, M. C., Maghlaoui, K., Barber, J., Timlin, J. A., Schulten, K., and Hunter, C. N. (2017) Lateral segregation of photosystem i in cyanobacterial thylakoids, *The Plant Cell* **29**(5), 1119–1136.
31. Kustka, A. B., Shaked, Y., Milligan, A. J., King, D. W., and Morel, F. M. (2005) Extracellular production of superoxide by marine diatoms: Contrasting effects on iron redox chemistry and bioavailability, *Limnol. Oceanogr.* **50**(4), 1172–1180.
32. Shaked, Y., Kustka, A. B., and Morel, F. M. (2005) A general kinetic model for iron acquisition by eukaryotic phytoplankton, *Limnol. Oceanogr.* **50**(3), 872–882.
33. Garg, S., Rose, A. L., Godrant, A., and Waite, T. D. (2007) Iron uptake by the ichthyotoxic chattonella marina (raphidophyceae): impact of superoxide generation1, *J. Phycol.* **43**(5), 978–991.
34. Wang, S., Ogata, M., Horita, S., Ohtsuka, J., Nagata, K., and Tanokura, M. (2014) A novel mode of ferric ion coordination by the periplasmic ferric ion-binding subunit fbpa of an abc-type iron transporter from *Thermus thermophilus* hb8, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **70**(1), 196–202.
35. Berntsson, R. P.-A., Smits, S. H., Schmitt, L., Slotboom, D.-J., and Poolman, B. (2010) A structural classification of substrate-binding proteins, *FEBS Lett.* **584**(12), 2606–2617.

36. Scheepers, G. H., Lycklama a Nijeholt, J. A., and Poolman, B. (2016) An updated structural classification of substrate-binding proteins, *FEBS Lett.* **590**(23), 4393–4401.
37. Tom-Yew, S. A., Cui, D. T., Bekker, E. G., and Murphy, M. E. (2005) Anion-independent iron coordination by the *Campylobacter jejuni* ferric binding protein, *J. Biol. Chem.* **280**(10), 9283–9290.
38. Koropatkin, N., Randich, A. M., Bhattacharyya-Pakrasi, M., Pakrasi, H. B., and Smith, T. J. (2007) The structure of the iron-binding protein, futa1, from *Synechocystis* 6803, *J. Biol. Chem.* **282**(37), 27468–27477.
39. Shouldice, S. R., Skene, R. J., Dougan, D. R., Snell, G., McRee, D. E., Schryvers, A. B., and Tari, L. W. (2004) Structural basis for iron binding and release by a novel class of periplasmic iron-binding proteins found in gram-negative pathogens, *J. Bacteriol.* **186**(12), 3903–3910.
40. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J. Mol. Biol.* **215**(3), 403–410.
41. Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, *Bioinforma.* **22**(13), 1658–1659.
42. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008) Phylogeny. fr: robust phylogenetic analysis for the non-specialist, *Nucleic acids research* **36**(suppl 2), W465–W469.
43. Dereeper, A., Audic, S., Claverie, J.-M., and Blanc, G. (2010) Blast-explorer helps you building datasets for phylogenetic analysis, *BMC Evol. Biol.* **10**(1), 8.
44. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-coffee: A novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol.* **302**(1), 205–217.
45. Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, *Mol. Biol. Evol.* **17**(4), 540–552.
46. Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, *Syst. Biol.* **52**(5), 696–704.
47. Anisimova, M. and Gascuel, O. (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative, *Syst. Biol.* **55**(4), 539–552.
48. Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., and Stanier, R. Y. (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria, *Microbiol.* **111**(1), 1–61.
49. Shcolnick, S., Shaked, Y., and Keren, N. (2007) A role for mrga, a dps family protein, in the internal transport of fe in the cyanobacterium synechocystis sp. pcc6803, *Biochimica et Biophys. Acta (BBA)-Bioenergetics* **1767**(6), 814–819.
50. Kranzler, C., Lis, H., Shaked, Y., and Keren, N. (2011) The role of reduction in iron uptake processes in a unicellular, planktonic cyanobacterium, *Environ. Microbiol.* **13**(11), 2990–2999.
51. Kolber, Z. S., Prášil, O., and Falkowski, P. G. (1998) Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols, *Biochimica et Biophys. Acta (BBA)-Bioenergetics* **1367**(1), 88–106.
52. Thomas, J. D., Daniel, R. A., Errington, J., and Robinson, C. (2001) Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (tat) pathway in escherichia coli, *Mol. Microbiol.* **39**(1), 47–53.
53. Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005) Prediction of twin-arginine signal peptides, *BMC bioinformatics* **6**(1), 167.
54. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) Signalp 4.0: discriminating signal peptides from transmembrane regions, *Nat. Methods* **8**(10), 785–786.
55. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein, *Nat. Biotechnol.* **24**(1), 79–88.

56. Hollingshead, S., Kopečná, J., Jackson, P. J., Canniffe, D. P., Davison, P. A., Dickman, M. J., Sobotka, R., and Hunter, C. N. (2012) Conserved chloroplast open-reading frame ycf54 is required for activity of the magnesium protoporphyrin monomethylester oxidative cyclase in synechocystis pcc 6803, *J. Biol. Chem.* **287**(33), 27823–27833.
57. Williams, J. G. (1988) [85] construction of specific mutations in photosystem ii photosynthetic reaction center by genetic engineering methods in synechocystis 6803, *Methods Enzymol.* **167**, 766–778.
58. Cereda, A., Hitchcock, A., Symes, M. D., Cronin, L., Bibby, T. S., and Jones, A. K. (2014) A bioelectrochemical approach to characterize extracellular electron transfer by synechocystis sp. pcc6803, *PLOS One* **9**(3), e91484.
59. Aldridge, C., Spence, E., Kirkilionis, M. A., Frigerio, L., and Robinson, C. (2008) Tat-dependent targeting of rieske iron-sulphur proteins to both the plasma and thylakoid membranes in the cyanobacterium synechocystis pcc6803, *Mol. Microbiol.* **70**(1), 140–150.
60. Kabsch, W. (2010) Xds, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **66**(2), 125–132.
61. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J. Appl. Crystallogr.* **40**(4), 658–674.
62. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **53**(3), 240–255.
63. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **60**(12), 2126–2132.
64. Michel, K.-P. and Pistorius, E. K. (1992) Isolation of a photosystem ii associated 36 kda polypeptide and an iron-stress 34 kda polypeptide from thylakoid membranes of the cyanobacterium synechococcus pcc 6301 grown under mild iron deficiency, *Zeitschrift für Naturforschung C* **47**(11-12), 867–874.

FOOTNOTES
Despo Polyviou was funded by a studentship from NERC, the Graduate School of the National Oceanography Centre Southampton (GSNOCS) and A.G. Leventis Foundation. Moritz M. Machelett was funded by a studentship from the National Oceanography Centre and Biological Sciences, University of Southampton.

Abbreviations: ABC, ATP-binding cassette; EPR, electron paramagnetic resonance; FBP, ferric binding protein; FZ, ferrozine; IdiA, iron deficiency induced protein A; PSII, photosystem II.

TABLES

Table 1: Crystallographic analysis: data collection and refinement statistics
Values in parentheses refer to the highest resolution shell.

	Tery_3377-Ala	**Tery_3377-Water**	**Tery_3377-Cit**
PDB accession code	6G7N	6G7P	6G7Q
Data collection			
Space group	$P2_1$	$P2_1$	$P2_1$
Cell dimensions			
a, b, c [Å]	56.2, 83.2, 65.2	56.0, 82.9, 64.9	56.3, 102.0, 57.7
α, β, γ [°]	90, 95.20, 90	90, 95.20, 90	90, 112.85, 90
Resolution [Å]	29.26-1.1 (1.130-1.1)	29.11-1.5 (1.540-1.5)	28.64-1.2 (1.230-1.2)
R_{merge} [%]	0.04759 (0.3528)	0.04115 (0.2744)	0.07207 (0.8711)
I/σ (I)	11.83 (3.00)	16.22 (4.13)	11.63 (1.55)
CC 1/2	0.998 (0.765)	0.999 (0.917)	0.999 (0.679)
Completeness [%]	99.42 (99.56)	95.48 (94.11)	98.30 (99.17)
Redundancy	3.2 (3.0)	3.5 (3.6)	5.8 (5.6)
Refinement			
Resolution [Å]	29.26-1.1 (1.129-1.1)	29.11-1.5 (1.539-1.5)	28.58-1.2 (1.231-1.2)
No. reflections	227890 (16772)	85685 (6274)	181790 (13467)
R_{work} [%]	0.1337 (0.2358)	0.1476 (0.1790)	0.1355 (0.2630)
R_{free} [%]	0.1512 (0.2377)	0.1739 (0.2050)	0.1648 (0.2840)
Monomers/AU	2	2	2
No. of atoms	6236	5656	6198
Protein	5250	5106	5193
Ligand/ion	24	11	49
Water	962	539	956
B-factors	16.42	18.12	17.88
Protein	15.24	17.90	16.46
Ligand/ion	27.24	34.35	26.75
Water	29.35	24.75	31.17
R.m.s. deviations			
Bond length [Å]	0.012	0.018	0.019
Bond angle [°]	1.57	1.89	2.01

FIGURES

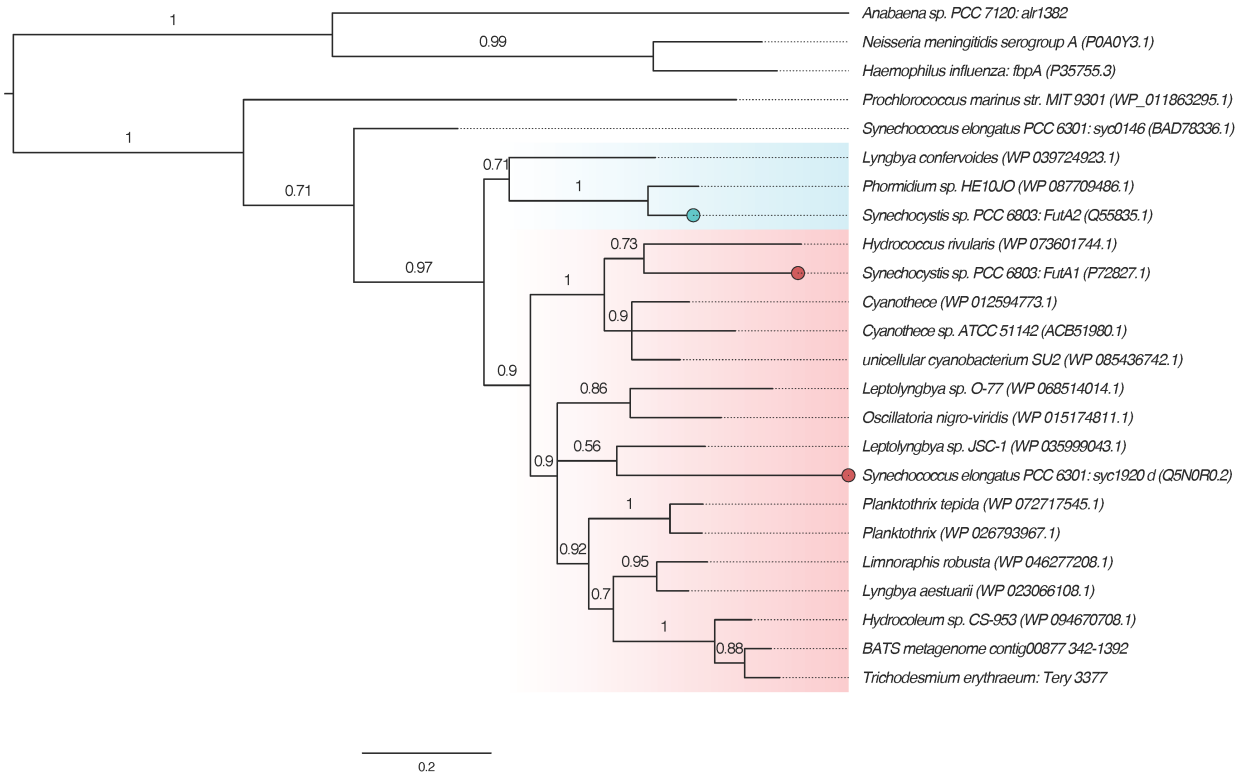

Figure 1: Maximum likelihood phylogenetic analysis of Tery_3377 homologs, based on amino acid sequence. homologs of Tery_3377 and *Synechocystis* FutA1 and FutA2 were identified using BlastP (40) and aligned along with sequences from the cyanobacteria *Synechococcus elongatus* PCC 6301, *Prochlorococcus marinus* MIT 9301, and *Anabaena* PCC 7120, and the pathogenic bacteria *Haemophilus influenza* and *Neisseria meningitides*. The graphical representation was produced using FigTree v1.4.3 and branch probabilities reported are calculated using an approximate likelihood ratio test (aLRT, 47). The *Synechocystis* FutA1 (red) and FutA2 (blue) proteins are separated into two groups, and Tery_3377 clusters within the FutA1-like group. Filled circles represent branches of the tree for which there is experimental evidence of intracellular (red) or Fe transport (blue) function. Database accessions for each protein (SwissProt where possible, otherwise RefSeq) are given in brackets. The scale bar represents the number of substitutions per site.

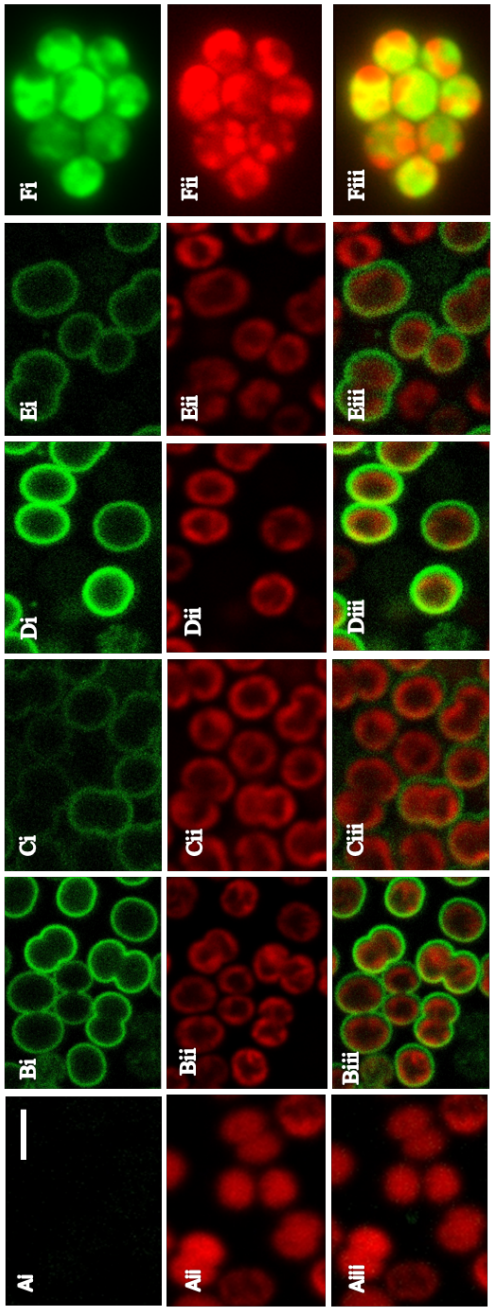

Figure 2: Localization of signal sequence-GFP constructs in *Synechocystis*. Using confocal microscopy the WT **(A)** was compared to strains expressing GFP fused to the signal sequences of the *E. coli* control TorA **(B)**, *Synechocystis* FutA1 **(C)**, *Synechocystis* FutA2 **(D)** and *Trichodesmium* Tery_3377 **(E)**; the GFP control, i.e. without signal sequence, is shown in **(F)**. Imaging was performed under identical settings with excitation at 488 nm and is shown as GFP emission at 510-530 nm **(i)** and as autofluorescence (represented by phycocyanin) at 600-700 nm **(ii)**; the merged images are shown in **(iii)**. Periplasmic localization of GFP is indicated by green fluorescent halos around the periphery of the cells in all strains **(B-E)** but not in WT **(A)** or the GFP control **(F)**, the latter showing cytosolic expression in absence of a signal peptide.

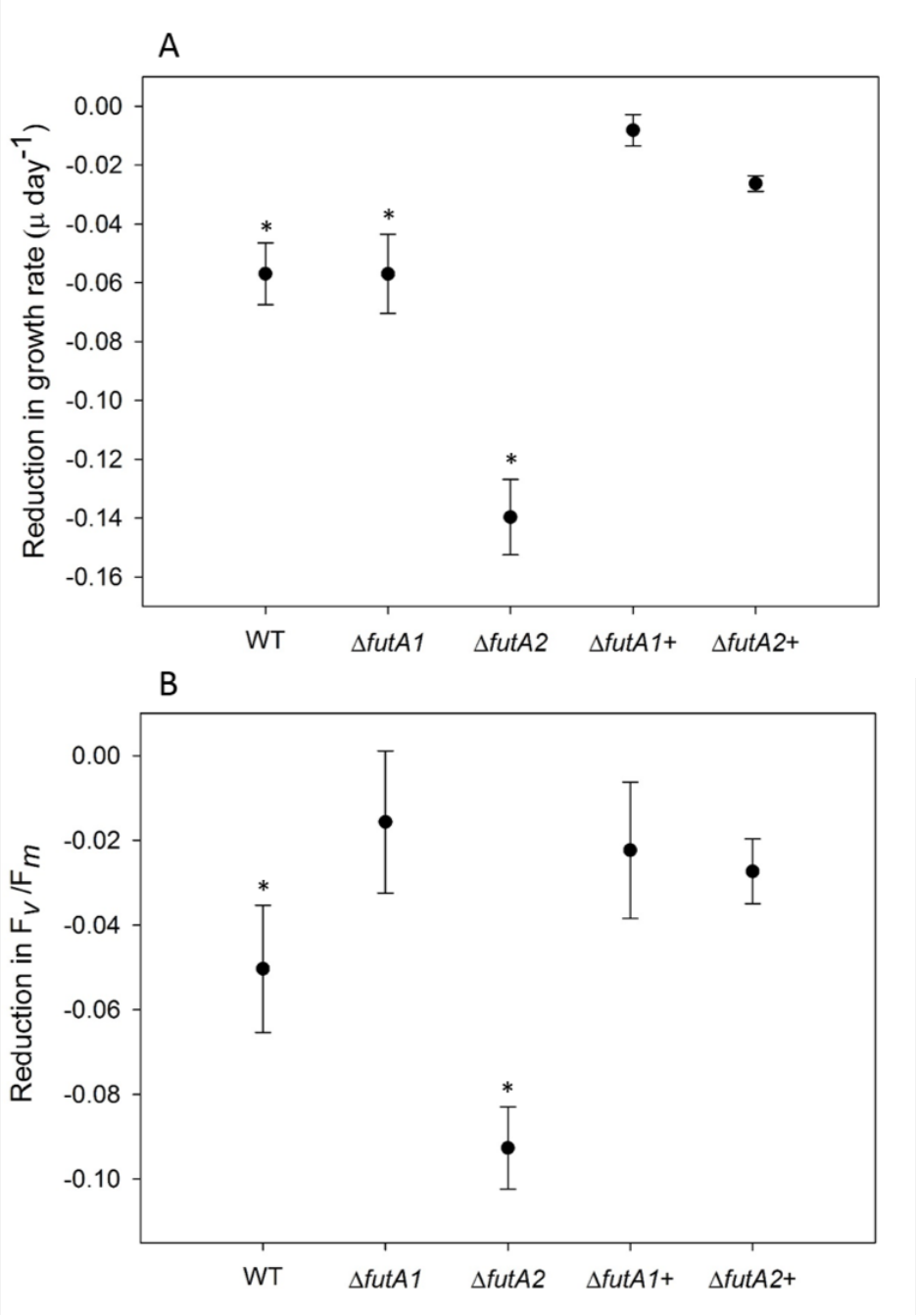

Figure 3: Reduction in growth and photosynthetic efficiency of strains grown in the presence of ferrozine. The decrease in growth rates measured as μ (day^{-1}) **(A)** and photosynthetic efficiency measured as F_v/F_m **(B)** during exponential growth of *Synechocystis* in YBG11 media when grown with and without the Fe^{2+} binding ligand ferrozine (FZ). Data are presented for WT, Δ*futA1* and Δ*futA2* mutants and the same strains expressing the *Trichodesmium tery_3377* gene (Δ*futA1*+ and Δ*futA2*+). Error bars show the standard deviation from the mean of three biological replicates. Statistical significance (*) was assessed using the Student's t-test ($p < 0.05$).

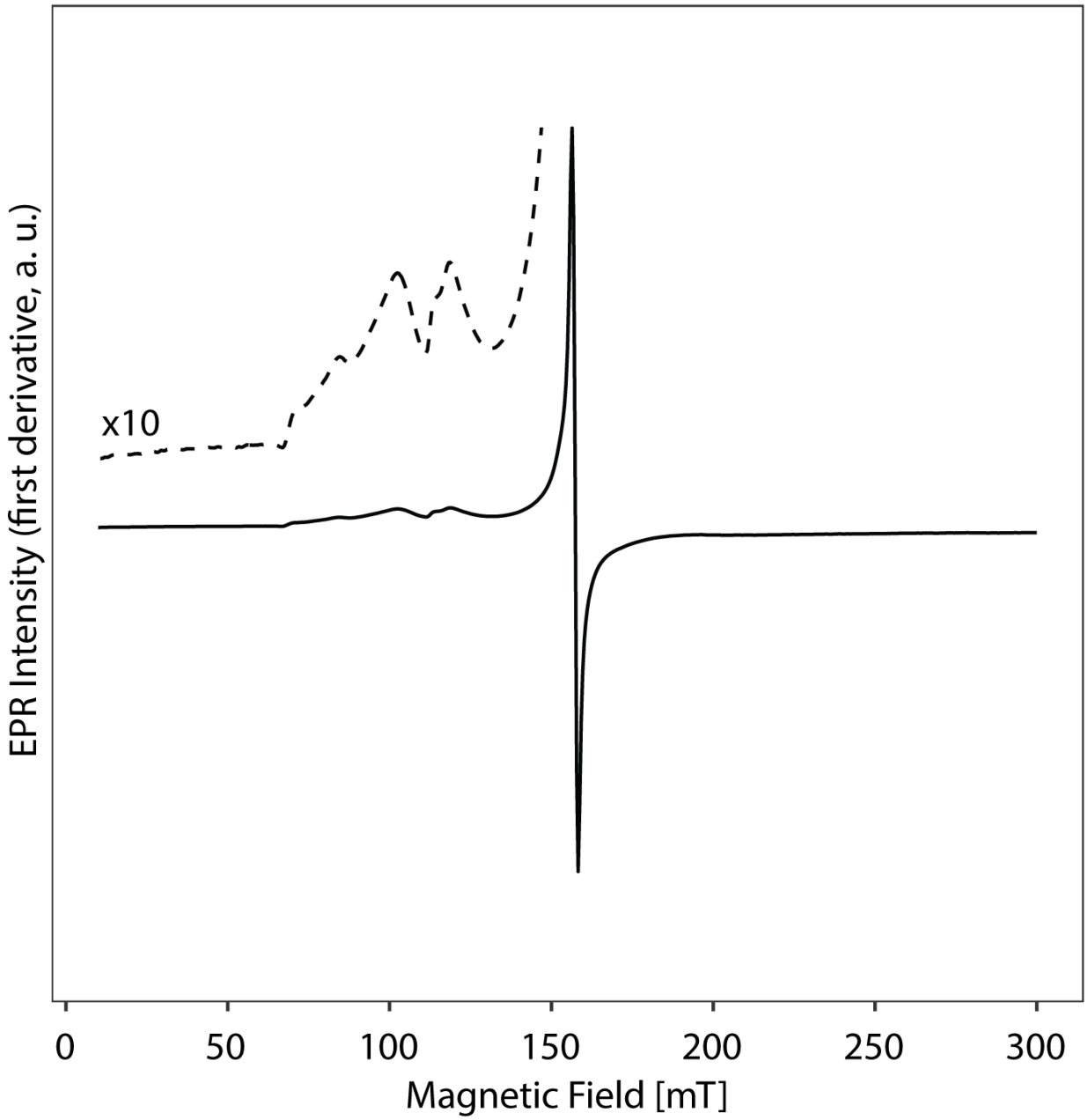

Figure 4: Electron paramagnetic resonance spectrum (EPR) of reconstituted Tery_3377 probing the iron centre. Recombinant purified Tery_3377 from *Trichodesmium erythraeum* ISM101 produced in *E. coli* was incubated with Ammonium Fe(II) sulfate under aerobic conditions; the complex was repurified by size exclusion chromatography to remove unbound Fe. The EPR spectrum was collected from a sample with 0.55 mM protein concentration.

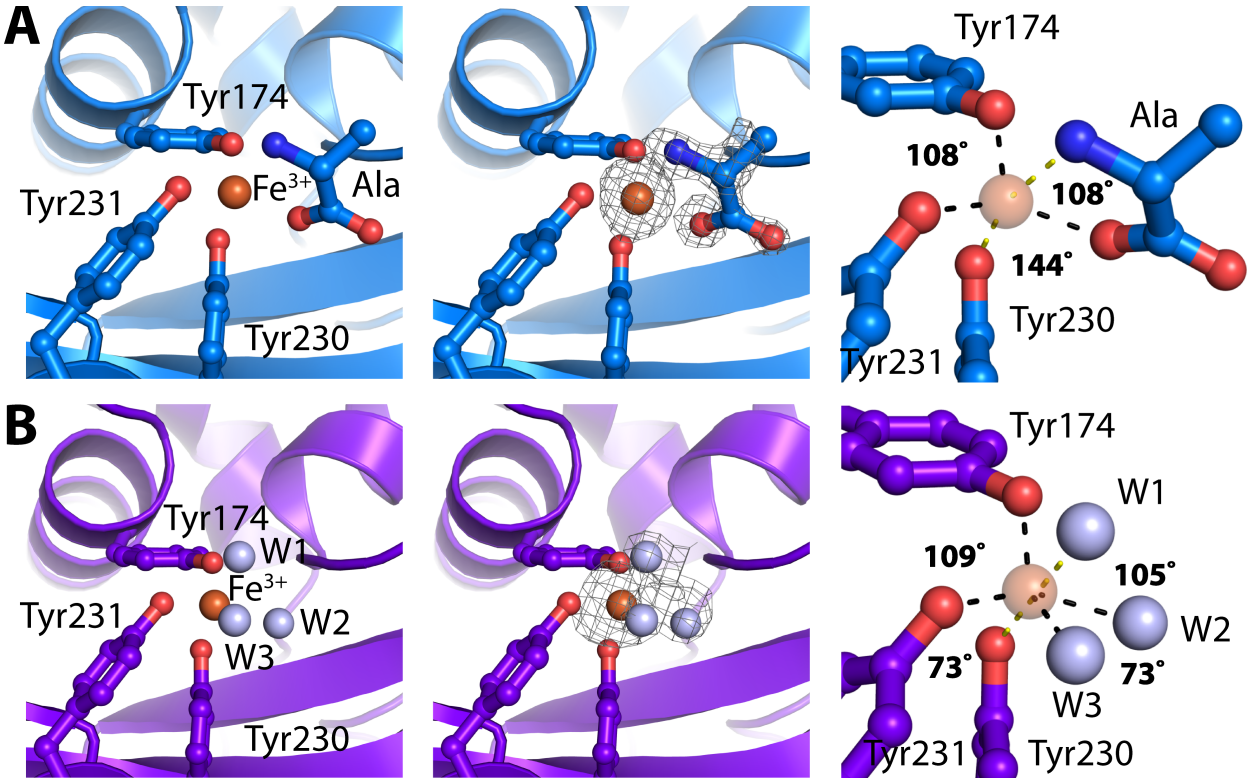

Figure 5: Structure analysis of Tery_3377. **(A)** Iron-binding site of Tery_3377-Ala with amino acid (ala) ligand (left); the electron-density map shown (middle) is obtained after molecular replacement and subsequent refinement without ligands (F_O - F_C, contoured at 3 sigma); ligand coordination (right) shows iron in a trigonal bi-pyramidal geometry; red: oxygen, blue: nitrogen, brown sphere: ferric-ion, yellow dashes: h-bonds to axial coordination sites. **(B)** Iron-binding site of Tery_3377-Water with water (W) ligands (left); the electron-density map shown (middle) is obtained after molecular replacement and subsequent refinement without ligands (Fo - Fc, contoured at 3 sigma); ligand coordination (right) shows iron in an octahedral geometry; red: oxygen, brown sphere: ferric-ion, gray sphere: water, yellow dashes: h-bonds to axial coordination sites.

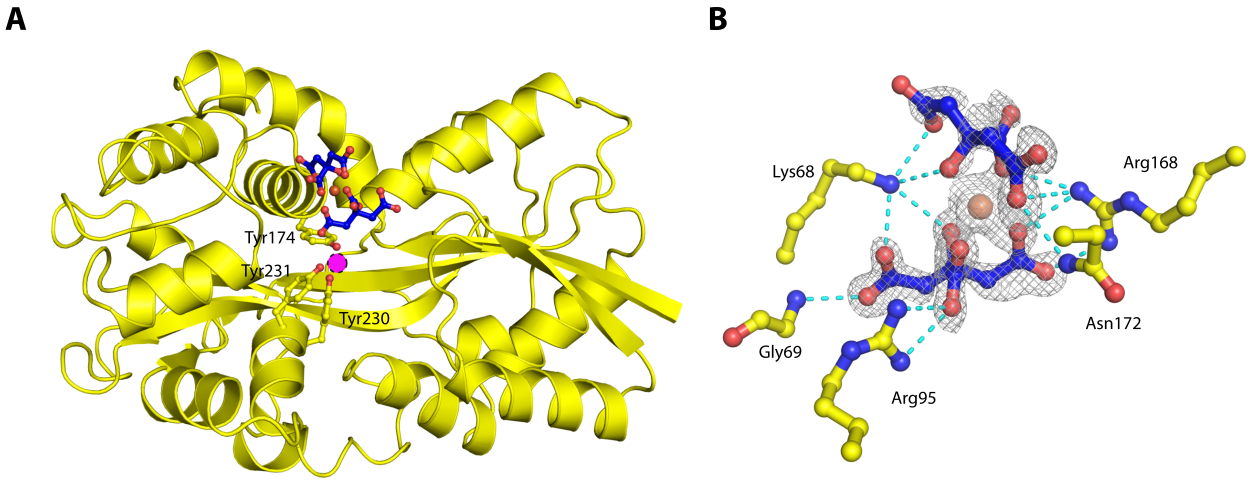

Figure 6: Iron binding site and coordination of citrate molecules. **(A)** The structure of Tery_3377-Cit. The fold is illustrated by the yellow ribbon, and the iron-binding site is shown by sticks representing the tyrosines shown in figure 5 (yellow) and two citrate molecules (blue). The iron coordinated by the two citrate molecules is shown as a brown sphere. The magenta circle indicates the position an iron ligand would occupy in iron-loaded Tery_3377. **(B)** Several protein amino acids, shown as yellow sticks, coordinate the citrate residues, shown in blue. The electron-density map is obtained after molecular replacement and subsequent refinement without ligands (F_O - F_C, contoured at 2 sigma). Red: oxygen, brown: ferric-ion, gray-mesh: electron density, blue dashes: coordinating interactions.

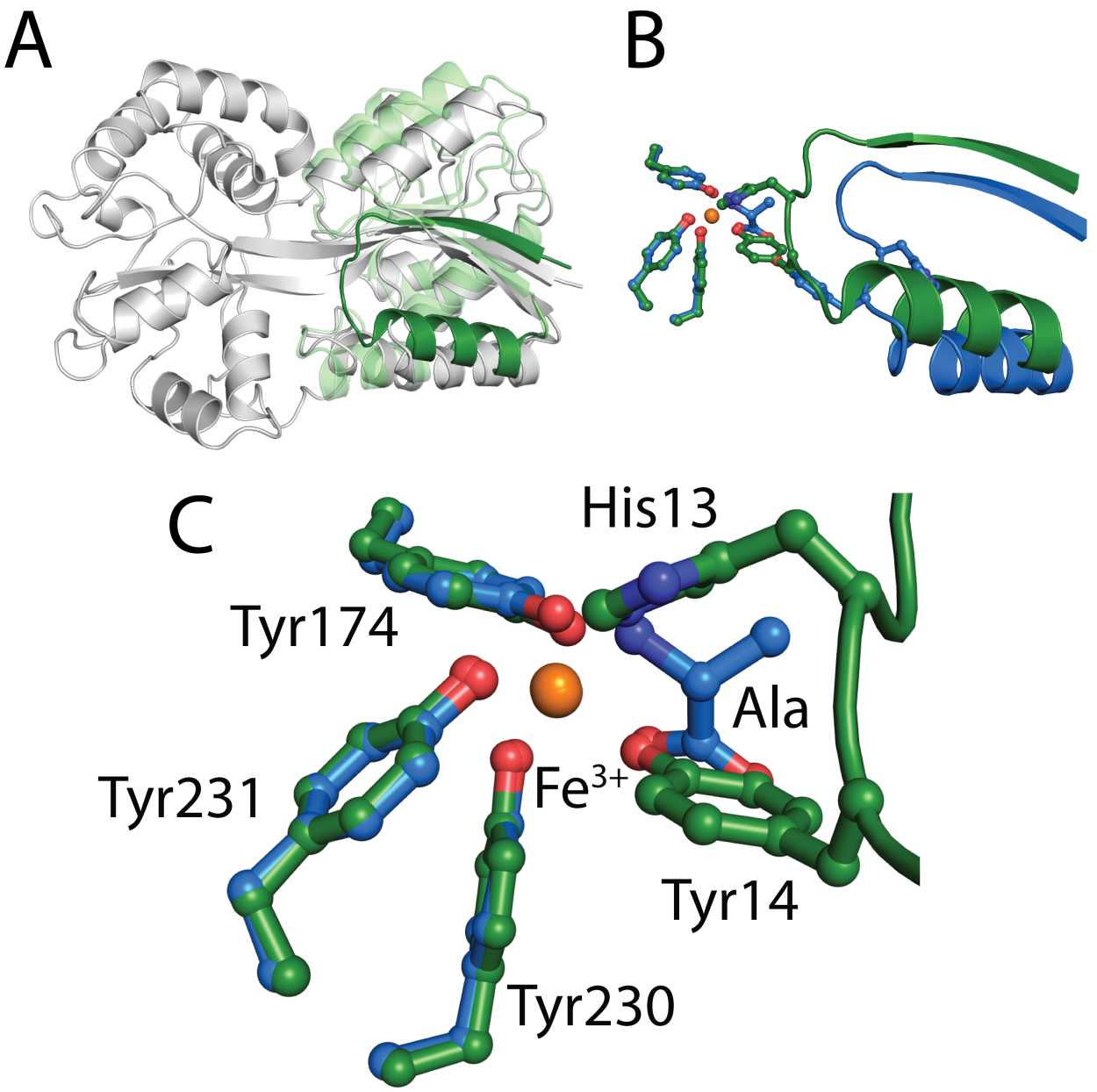

Figure 7: Analysis of iron coordination in Tery_3377. (A) The open state of FutA2 from *Synechocystis* sp. PCC6803 is Fe free (gray). When superposed with the closed Fe bound form of FutA2 from *Synechocystis* sp. PCC6803 (Fe-FutA2), the C-terminal domains can be brought into perfect overlap while the N-terminal domains reveal structural changes (green and light green) indicative of a closing of the substrate binding cleft. Notably, a loop region contributing binding residues to the Fe binding sites is observed in a different position (green) to the open state. **(B)** Superposition of Tery_3377-Ala (blue) with Fe-FutA2 (green). The loop from the N-terminal domain brings Fe co-ordinating tyrosine and histidine side chains into contact distance, allowing Fe coordination. **(C)** Coordination site of superposed Tery_3377-Ala (blue) with Fe-FutA2 (green). The equivalent amino acid residues to His13 and Tyr14, His43 and Tyr44 in Tery_3377, are not seen in contact with iron in either Tery_3377-Ala nor Tery_3377-Water. However, the polar atoms from the alanine ligand in Tery_3377-Ala superpose well with the coordinating atoms of His13 and Tyr14. The three co-ordinating tyrosine residues originating from the C-terminal domains occupy the same position in Tery_3377-Ala and Fe-FutA2. Coordinating residues (shown as sticks) and ferric iron (brown sphere).