HAL
open science

CyanoLyase: a database of phycobilin lyase sequences, motifs and functions

Anthony Bretaudeau, François Coste, Florian Humily, Laurence Garczarek, Gildas Le Corguillé, Christophe Six, Morgane Ratin, Olivier Collin, Wendy M Schluchter, Frédéric Partensky

To cite this version:

Anthony Bretaudeau, François Coste, Florian Humily, Laurence Garczarek, Gildas Le Corguillé, et al.. CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. Nucleic Acids Research, 2012, pp.6. 10.1093/nar/gks1091 . hal-01094087

HAL Id: hal-01094087

https://inria.hal.science/hal-01094087v1

Submitted on 11 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

D396–D401 Nucleic Acids Research, 2013, Vol. 41, Database issue Published online 21 November 2012
doi:10.1093/nar/gks1091

CyanoLyase: a database of phycobilin lyase sequences, motifs and functions

Anthony Bretaudeau[1], François Coste[2], Florian Humily[3,4], Laurence Garczarek[3,4], Gildas Le Corguillé[3,5], Christophe Six[3,4], Morgane Ratin[3,4], Olivier Collin[1], Wendy M. Schluchter[6] and Frédéric Partensky[3,4,*]

[1]GenOuest Platform, [2]Dyliss team, INRIA/Irisa, Campus de Beaulieu, 35042 Rennes Cedex, [3]UPMC-Université Paris VI, Station Biologique, Place Georges Teissier, 29680 Roscoff France, [4]CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment, Oceanic Plankton Group, 29680, [5]CNRS, FR2424 Analysis and Bioinformatics for Marine Science (ABiMS), 29680 Roscoff, France and [6]Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148, USA

Received August 29, 2012; Revised October 17, 2012; Accepted October 18, 2012

ABSTRACT

CyanoLyase (http://cyanolyase.genouest.org/) is a manually curated sequence and motif database of phycobilin lyases and related proteins. These enzymes catalyze the covalent ligation of chromophores (phycobilins) to specific binding sites of phycobiliproteins (PBPs). The latter constitute the building bricks of phycobilisomes, the major light-harvesting systems of cyanobacteria and red algae. Phycobilin lyases sequences are poorly annotated in public databases. Sequences included in CyanoLyase were retrieved from all available genomes of these organisms and a few others by similarity searches using biochemically characterized enzyme sequences and then classified into 3 clans and 32 families. Amino acid motifs were computed for each family using Protomata learner. CyanoLyase also includes BLAST and a novel pattern matching tool (Protomatch) that allow users to rapidly retrieve and annotate lyases from any new genome. In addition, it provides phylogenetic analyses of all phycobilin lyases families, describes their function, their presence/absence in all genomes of the database (phyletic profiles) and predicts the chromophorylation of PBPs in each strain. The site also includes a thorough bibliography about phycobilin lyases and genomes included in the database. This resource should be useful to scientists and companies interested in natural or artificial PBPs, which have a number of biotechnological applications, notably as fluorescent markers.

INTRODUCTION

Oxygenic phototrophic prokaryotes (i.e. cyanobacteria) share with the eukaryotic classes Rhodophyta (i.e. red algae) and Cryptophyta the presence of phycobiliproteins (PBPs), which are water-soluble proteins chromophorylated with brilliantly colored, linear-tetrapyrrolic pigments, called phycobilins (1). In red algae and most cyanobacterial species, different PBP types are assembled to form phycobilisomes (PBS), the major light-harvesting systems of these organisms, which are constituted of a central core surrounded by a number of radiating rods [usually six in cyanobacteria (2,3)]. Although the main antenna system of Cryptophyta is a membrane-intrinsic Lhc-type complex, like in all other photosynthetic eukaryotes (except red algae), cryptophytes possess a secondary one made of tightly packed aggregates of one PBP type, either phycocyanin (PC) or phycoerythrin (PE), located in the thylakoid lumen in the proximity of photosystems (4).

Although the PBP and phycobilin composition of the PBS core varies little, because it is always composed of allophycocyanin (APC) that binds phycocyanobilin (PCB) as its only chromophore, the structure of PBS rods is extremely variable between groups, and even within a given genus (3). In marine *Synechococcus*, for instance, six different pigment types have been described so far (5), based on the various PBP composition of their PBS rods. Indeed, the latter can comprise one to three of four possible PBP types: PC, PE-I or PE-II and phycoerythrocyanin [PEC, so far only found in freshwater species; see (3)]. Furthermore, the phycobilin composition of each individual PBP itself varies, because it may bind one to three types of the four different possible phycobilin types: PCB, phycoerythrobilin (PEB), phycourobilin (PUB) and phycoviolobilin (PVB), which are isomers with distinct

*To whom correspondence should be addressed. Tel: +33 2 9829 2564; Fax: +33 2 9829 2324; Email: partensky@sb-roscoff.fr

© The Author(s) 2012. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 11, 2014

spectral properties. PBPs generally consist of two subunits, α and β, organized into hexamers, and each subunit has either one (α- or β-APC, α-PC and α-PEC), two (α-PEI and β-PC) or three chromophore binding cysteinyl sites (β-PEI, α-PEII and β-PEII). Given this complexity, the phycobilin lyases, that is the enzymes that catalyze the ligation of chromophores to PBPs, constitute a particularly wide and diversified group of proteins (6,7). Among those which have been biochemically characterized, most are highly specific *in vivo* as they can ligate only one phycobilin type at one particular PBP binding site. However, CpcS (a member of the S/U clan) can bind either PCB or PEB to one specific site, Cys-82 (consensus numbering), of a variety of PBPs (α- and β-APC and PE, β-PC and PEC) and is therefore more universal (8). Furthermore, some enzymes of the E/F clan are bifunctional, because they can both bind a chromophore (either PCB for PecE/F or PEB for RpcG) to α-PC and change its chemical configuration into another isomer [i.e. PVB or PUB, respectively; see (9–13)].

Here, we describe CyanoLyase, a sequence and motif database dedicated to the annotation of phycobilin lyases and related proteins. Indeed, these enzymes are often poorly annotated in public databases, especially sequences coming from genome projects. Given the fact that PBPs have a growing number of biotechnological and biomedical applications [see, e.g. (14) and references therein], this resource should be very useful to all scientists and companies interested in natural or artificial PBPs. Furthermore, the knowledge of the phycobilin lyase content of any given cyanobacterial strain can be used to predict the pigmentation of its PBS, even if the latter was previously unknown, and therefore, Cyanolyase provides a list of the known and predicted chromophores at all binding positions of PBPs for most strains of the database. CyanoLyase also contains bioinformatic tools, BLAST (15) and a new pattern analysis suite Protomata [see (16) and http://tools.genouest.org/tools/protomata], that allow users to rapidly retrieve and annotate all lyases present in any new genome using a whole proteome file in Fasta format. In addition, it provides tables specifying the function of lyases and phyletic profiles [i.e. patterns of presence of orthologs in a set of genomes; see, e.g. (17)] allowing the user to determine the co-occurrence of lyase genes in the different strains of the database.

DATA COLLECTION AND CURATION

The CyanoLyase database is mainly composed of sequences of characterized or presumed phycobilin lyases retrieved from genomes of cyanobacteria, red algae or cryptophytes. However, in view of forthcoming evolutionary studies of this interesting enzyme group, the database also comprises sequences of a number of phylogenetically related proteins, such as NblB that is involved in PBS degradation during nitrogen starvation (18) or IaiH involved in iron–sulfur cluster biosynthesis (19), as well as other proteins with no characterized function to date. At the time of writing, CyanoLyase accounted 954 sequences of phycobilin lyases and related proteins, coming from 84 genomes (mainly cyanobacteria). These sequences have been classified into three main clans [i.e. proteins sharing a common 3D structure; see, e.g. (20)] and 30 different families [i.e. groups of orthologous sequences; see, e.g. (21)], a modification and extension of the previous classification proposed by Schluchter *et al.* (7). For members of the S/U and E/F clans, the 3D structure was predicted using the Protein Fold Recognition Server Phyre 2 (22), while there is so far no structure that fits members of the T clan in public databases. The E/F clan was further subdivided into two subclans, based on both the phylogeny and the fact that enzymes belonging to E/F subclan 1 form either heteroduplexes or fusion proteins, whereas members of E/F subclan 2 apparently do not. Each family in our classification gathers proteins that we assume to have the same biological function. Some families were further divided into subfamilies, based on phylogenetic analyses, which for instance often split apart marine picocyanobacteria sequences from other cyanobacteria (7).

To build the dataset included in the CyanoLyase database, an initial set of biochemically characterized phycobilin lyases was compiled from an extensive literature survey and similarity searches were conducted using BLAST to retrieve highly homologous sequences in all cyanobacterial genomes available in RefSeq. After attributing these sequences to a given family and/or subfamily, conserved amino acid motifs were computed using Protomata learner. Then, using both BLAST and Protomatch (see Tools section below for details about Protomata learner and Protomatch), more distantly related sequences were identified in public databanks (e.g. RefSeq protein) and added to existing or newly created families.

Sequences stored in the database are tagged as 'sure' or 'unsure'. 'Unsure' sequences are sequences that, based on similarity or the presence of a conserved motif, have been affiliated to a given family, but for which matching scores are too low to ascertain that they have the same function as other members of this family (e.g. CMR092C from *Cyanidioschyzon merolae* or the two paralogous sequences AM1_4215 and AM1_C0217 from *Acaryochloris marina* have been attributed to the CpcS family, but tagged 'unsure' because the identity to other members of this well-conserved family is lower than 53%). 'Unsure' sequences are indicated in italics in the family description pages and are not used for motif and phylogeny inferences.

All the sequences stored in CyanoLyase were manually curated, and modifications were made for some open reading frames (ORFs) that were not accurately defined in public databanks (GenBank). For instance, some protein sequences were missing a few residues at the N-terminus and were therefore extended, whereas others seemingly had too long N-termini due to a misplaced start codon and were then shortened. Also, two sequences (CpcSIII in *Cylindrospermopsis raciborskii* CS-505 and in *Raphidiopsis brookii* D9) were found to have a bacterial group II intron insertion, which led to the prediction of two independent ORFs by the annotation algorithms. In

Downloaded from http://nar.oxfordjournals.org/ by guest on December 11, 2014

this case, the intronic region was suppressed and the two ORFs fused. Whenever such an alteration was made to a sequence, this was reported in the corresponding remark field.

Because most of the sequences of CyanoLyase come from GenBank, NCBI record IDs are included in the database, when available, to keep track of the origin of the data.

DATA ACCESS

A free, public access to the CyanoLyase database, tools and other features are available at http://cyanolyase.genouest.org. A brief description of this group of enzymes, some database statistics and direct access to bioinformatic tools are available from the 'Home Page'. Several browsing methods are available to access the information contained in the database: curated data, applications and additional features are arranged under independent pull-down menus, the main ones called 'Genomes', 'Families', 'Functions', 'Pigmentation', 'Blast', 'Protomatch', 'Phyletic profile' and 'References'.

The 'Genomes' page lists all the genomes where at least one 'true' phycobilin lyase has been identified (i.e. not only a related sequence) and provides some information about strain taxonomy, classification, environment and the sequencing center and status. This list, like all others in CyanoLyase, is sortable and filterable to ease the navigation. It can also be exported in various formats (csv and pdf). Clicking on a genome displays information about this genome, links to its RefSeq record (if available), some bibliographic references and the list of phycobilin lyase or related sequences that were found in this genome.

The 'Families' page displays the classification of sequences included in the database that are divided in clans, subclans, families and subfamilies. For each level, a hyperlink leads to a brief description, some bibliographic references and the list of sequences belonging to this group and, for most of the two lower levels, additional links give access to amino acid motifs. Each sequence stored in CyanoLyase has a dedicated page with details about the genome where the sequence comes from and the family in which it is classified.

Because CyanoLyase keeps track of NCBI record IDs (GenBank, RefSeq) of sequences and genomes when available, it offers users the possibility to display the genomic context of each phycobilin lyase gene (Figure 1) using the NCBI Sequence Viewer (http://www.ncbi.nlm.nih.gov/projects/sviewer/). Using this tool, the user gets access to the genomic organization around lyase genes. This option is of particular interest because the latter genes are frequently organized in clusters with other genes involved in PBS biosynthesis and regulation (5).

TOOLS

Some bioinformatic tools are directly available on the CyanoLyase website to perform analyzes of new sequences to find novel members of the phycobilin lyase family. A BLAST (version 2.2.26+) form allows users to search for sequence similarity of query sequences in CyanoLyase databanks. These databanks (that can be selected using a scrolldown menu) comprise not only all individual phycobilin lyase and related protein families but also the whole proteomes of all cyanobacteria and red algae included in the database. BLAST results can be downloaded in various formats. Sequences already recorded in CyanoLyase and having an associated GenBank ID are highlighted in the result page by the presence of the abbreviation 'CL' (for CyanoLyase), just before the ORF ID in the BLAST result.

CyanoLyase also gives access to an original motif discovery and matching tool suite: Protomata (version 2.0; http://tools.genouest.org/tools/protomata). This software allows the user to discover motifs in sets of related sequences, focusing only on most conserved regions, represented as blocks with a letter size proportional to the amino acid frequency at any position (Figure 2). Most often, multiple blocks are detected for each dataset, and each block can be found in all the sequences or only a subset of them. With this tool, it is possible to detect the regions common to all the sequences of a protein family, but also regions shared only by some sequences that may constitute a subfamily. Using Protomata learner, a motif has been generated for most lyase family and subfamily. These motifs are available online and can be used directly from the web interface to search for motif matching on new protein sequences using Protomatch.

Using both BLAST and Protomatch allows users to rapidly retrieve all the putative phycobilin lyases or related sequences present in any new genome within a few minutes, directly in a web browser. The same tools can also be used to search for such sequences in public databanks such as RefSeq protein or NR. This type of search will be done on a regular basis by CyanoLyase authors to keep the database updated.

In addition, a phylogeny program has been integrated into CyanoLyase. It performs a succession of three tasks: (i) using Muscle (23), it automatically generates multiple alignments of protein sequences for each level from subfamilies to clans, (ii) using trimAl (24), it eliminates gaps and low-quality regions and (iii) using PhyML (25), it generates Maximum Likelihood trees. The multiple alignments and phylogenetic trees (under both Newick format and radial visualization) are available in the 'Phylogeny' boxes that appear in the description of each group of sequences at all levels of the classification.

ADDITIONAL FEATURES

The 'Function' menu summarizes in a table the characterized or predicted functions of phycobilin lyases in terms of the bound chromophore (PCB or PEB), enzymatic activity (lyase or lyase-isomerase), substrate (apoprotein targeted by the enzyme) and binding sites (cysteine positions at which phycobilins are bound). These data refer to literature data listed in the 'References' menu (see below). It is important to note that phycobilin lyases often have a lower specificity *in vitro* than *in vivo* (9,26), but we chose to display their *in vivo* properties.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 11, 2014

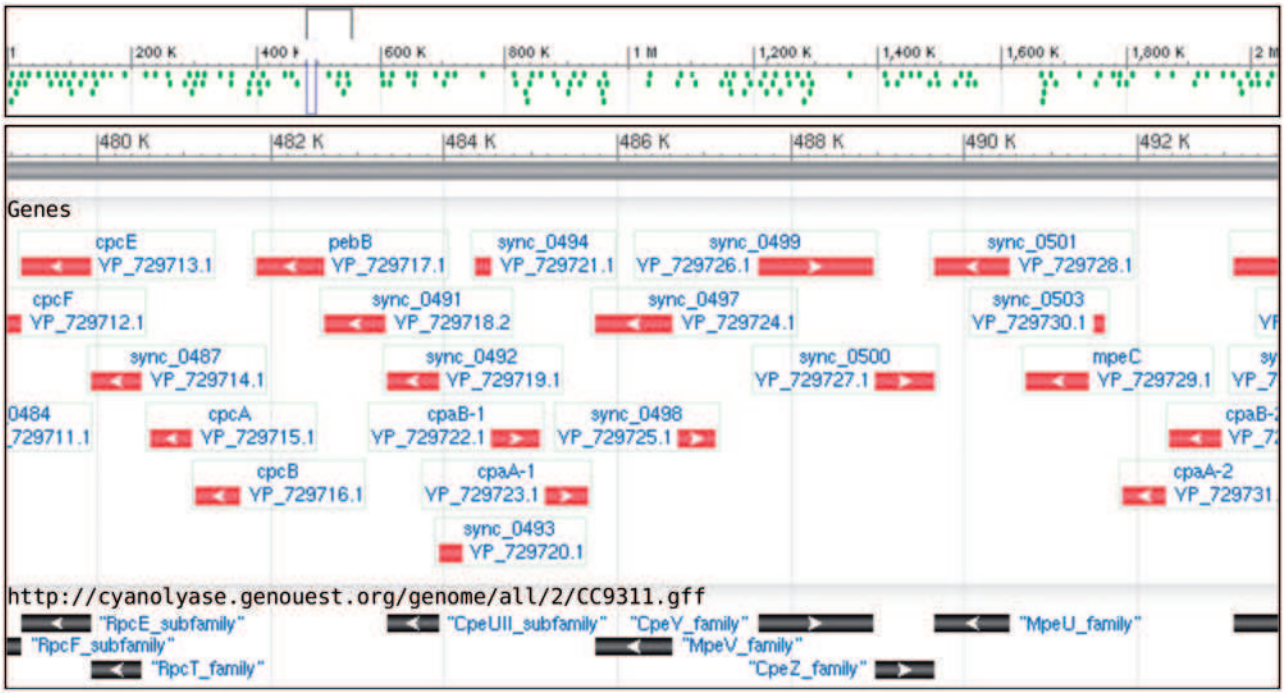

Figure 1. Use of the Browser NCBI to visualize the genome context of lyase genes in the genome of *Synechococcus* sp. CC9311. The upper track shows the region with a low level of zoom and the middle track with a higher level of zoom and show (in red) all ORFs with their IDs (or gene names) in the RefSeq record. The bottom track displays (in black) genes present in the CyanoLyase database and their corresponding name.

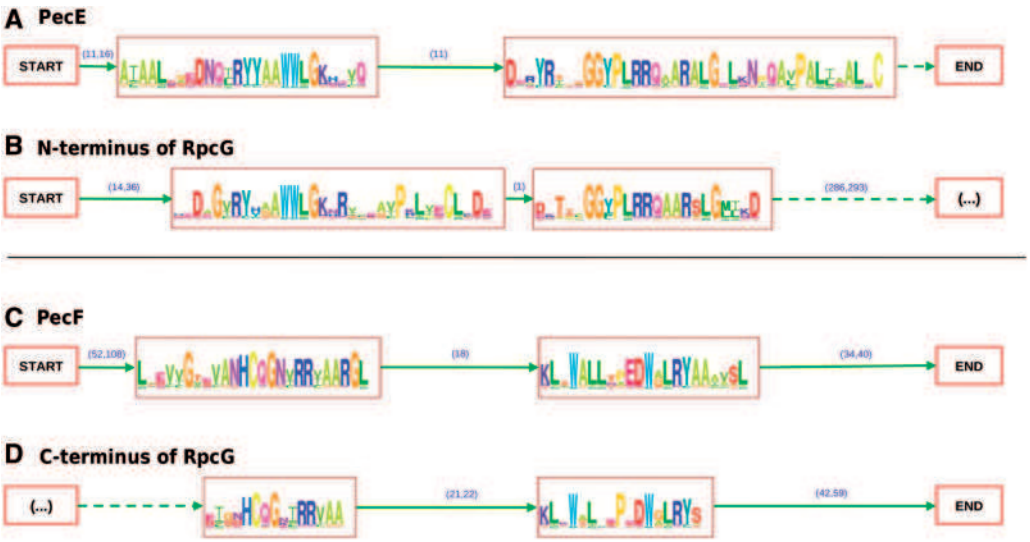

Figure 2. Example of motifs created by Protomata learner for members of the E/F subclan 1. Motifs for the PecE (**A**) and PecF (**C**) subunits of the PCB:Cys-84 α-PEC lyase-isomerase PecE/F(12) and for the corresponding N-terminus (**B**) and C-terminus (**D**) of the closely related fusion protein RpcG, a PEB:Cys-84 α-PC lyase-isomerase (9). Note the similarity of motifs independently found by Protomata learner for these two sets of protein sequences, and in particular the presence of the well-conserved motifs YyaAWWL, biochemically characterized as essential for the lyase activity of PecE (13) and nHCQGn, conferring its isomerase activity to PecF (13), in the N- and C-terminus of RpcG, respectively.

The 'Pigmentation' page displays the PBP content for each strain listed in the genome page and the chromophores found at the different binding sites of PBP subunits. The latter information was either determined biochemically (bold letters) or predicted by CyanoLyase authors (normal letters), based on the presence of specific phycobilin lyases in the genome.

The 'Phyletic profiles' page shows the distribution of phycobilin lyases or related proteins in the different genomes of the CyanoLyase database. It allows the user to easily spot sequences that are present in a comparable set of genomes, and this is particularly useful to predict potential heterodimers. For instance, this tool allowed us to find that marine *Synechococcus* spp. likely possess a

Downloaded from http://nar.oxfordjournals.org/ by guest on December 11, 2014

CpcU-like lyase (that we called CpcU-II), which likely makes an heterodimer with CpcS-II, by analogy with CpcU-I, which forms a heterodimer with CpcS-I in a number of freshwater cyanobacteria (7,27). In contrast, CpcS-III does not seem to form heterodimers (8).

The 'References' page comprises an exhaustive bibliography about the function of the different lyases and the description of genomes included in the database. These references are referred to, when appropriate, in the different sections of the CyanoLyase website.

IMPLEMENTATION AND DATA EXPORT

CyanoLyase was developed using the Symfony 2 PHP framework (http://symfony.com/). Some modules that were used to build the application are available under an open source license on github (https://github.com/genouest/). The web interface was designed with XHTML, CSS and JQuery, and data were integrated into a MySQL database. These data can be exported in Fasta format from the web interface either as individual sequences or as multiple sequence files, for example if one wishes to retrieve all members of a particular family or of a genome.

CURRENT SCOPE AND FUTURE PERSPECTIVE

The first release of CyanoLyase provides an extensive collection of phycobilin lyases and related proteins, classified in clans, subclans, families and subfamilies. The website also gives access to bioinformatic tools to ease the annotation of these sequences in forthcoming genomes of PBP-containing organisms. As such, the website will be updated regularly as new data become available and will therefore be a long-term resource. Users can monitor directly from the web interface the latest changes that have occurred in the database using the corresponding scrolldown menu.

CyanoLyase aims to be a reference resource about the classification of phycobilin lyases and their respective predicted or characterized functions. Future versions could include sequences from metagenomic samples or viruses, and the same kind of resources could be built for other complex and poorly annotated groups of protein sequences, such as the polypeptide linkers that maintain the PBS assembly (3).

ACKNOWLEDGEMENTS

The Authors thank Dr M. Czjzek for her useful hints concerning protein structures that helped them designing the new classification of phycobilin lyases used in the database. Cyanolyase would not be possible without the genome data publicly available at the National Center for Biotechnology Information.

FUNDING

Agence Nationale de la Recherche Scientifique (ANR), Microbial Genomics Programme PELICAN, [ANR-09-GENM-030]; European Union's Seventh Framework Programmes (FP7) MicroB3 and MaCumBa (287589 and 311975, respectively); National Science Foundation grant [MCB-0843664]. Funding for open access charge: ANR contract [ANR-09-GENM-030].

Conflict of interest statement. None declared.

REFERENCES

1. Apt,K.E., Collier,J.L. and Grossman,A.R. (1995) Evolution of the phycobiliproteins. *J. Mol. Biol.*, **248**, 79–96.
2. Glazer,A.N. (1982) Phycobilisomes: structure and dynamics. *Annu. Rev. Microbiol.*, **36**, 173–198.
3. Sidler,W.A. (1994) Phycobilisome and phycobiliprotein structure. In: Bryant,D.A. (ed.), *The Molecular Biology of Cyanobacteria*. Kluwer Academic Publishers, The Netherlands, pp. 139–216.
4. Glazer,A.N. and Wedemayer,G.J. (1995) Cryptomonad biliproteins—an evolutionary perspective. *Photosynth. Res.*, **46**, 93–105.
5. Six,C., Thomas,J.C., Garczarek,L., Ostrowski,M., Dufresne,A., Blot,N., Scanlan,D.J. and Partensky,F. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol.*, **8**, R259.
6. Scheer,H. and Zhao,K.H. (2008) Biliprotein maturation: the chromophore attachment. *Mol. Microbiol.*, **68**, 263–276.
7. Schluchter,W.M., Shen,G., Alvey,R.M., Biswas,A., Saunée,N.A., Williams,S.R., Miller,C.A. and Bryant,D.A. (2010) Phycobiliprotein biosynthesis in cyanobacteria: structure and function of enzymes involved in post-translational modification. *Adv. Exp. Med. Biol.*, **675**, 211–228.
8. Zhao,K., Su,P., Tu,J., Wang,X., Liu,H., Plöscher,M., Eichacker,L., Yang,B., Zhou,M. and Scheer,H. (2007) Phycobilin:cystein-84 biliprotein lyase, a near-universal lyase for cysteine-84-binding sites in cyanobacterial phycobiliproteins. *Proc. Natl Acad. Sci. USA*, **104**, 14300–14305.
9. Blot,N., Wu,X.-J., Thomas,J.-C., Zhang,J., Garczarek,L., Böhm,S., Tu,J.M., Zhou,M., Plöscher,M., Eichacker,L. *et al.* (2009) Phycourobilin in a unique trichromatic phycobiliprotein is formed post-translationally by a novel phycoerythrobilin lyase-isomerase. *J. Biol. Chem.*, **284**, 9290–9298.
10. Jung,L.J., Chan,C.F. and Glazer,A.N. (1995) Candidate genes for the phycoerythrocyanin alpha subunit lyase. Biochemical analysis of *pecE* and *pecF* interposon mutants. *J. Biol. Chem.*, **270**, 12877–12884.
11. Storf,M., Parbel,A., Meyer,M., Strohmann,B., Scheer,H., Deng,M.G., Zheng,M., Zhou,M. and Zhao,K.H. (2001) Chromophore attachment to biliproteins: specificity of PecE/PecF, a lyase-isomerase for the photoactive 3(1)-cys-alpha 84-phycoviolobilin chromophore of phycoerythrocyanin. *Biochemistry*, **40**, 12444–12456.
12. Zhao,K.H., Deng,M.G., Zheng,M., Zhou,M., Parbel,A., Storf,M., Meyer,M., Strohmann,B. and Scheer,H. (2000) Novel activity of a phycobiliprotein lyase: both the attachment of phycocyanobilin and the isomerization to phycoviolobilin are catalyzed by the proteins PecE and PecF encoded by the phycoerythrocyanin operon. *FEBS Lett.*, **469**, 9–13.
13. Zhao,K.H., Wu,D., Zhou,M., Zhang,L., Böhm,S., Bubenzer,C. and Scheer,H. (2005) Amino acid residues associated with enzymatic activities of the isomerizing phycoviolobilin-lyase PecE/F. *Biochemistry*, **44**, 8126–8137.
14. Sekar,S. and Chandramohan,M. (2008) Phycobiliproteins as a commodity: trends in applied research, patents and commercialization. *J. Appl. Phycol.*, **20**, 113–136.
15. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinfo.*, **10**, 421.
16. Coste,F. and Kerbellec,G. (2005) A similar fragments merging approach to learn Automata on proteins. In: Gama,J., Camacho,R., Brazdil,P., Jorge,A. and Torgo,L. (eds), *Proceeding 16th European Conference on Machine Learning (ECML 2005)*. Springer-Verlag, Berlin, Heidelberg, Porto, Portugal, pp. 522–529.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 11, 2014

17. Aravind,L. (2000) Guilt by association: contextual information in genome analysis. *Genome Res.*, **10**, 1074–1077.
18. Dolganov,N. and Grossman,A.R. (1999) A polypeptide with similarity to phycocyanin alpha-subunit phycocyanobilin lyase involved in degradation of phycobilisomes. *J. Bacteriol.*, **181**, 610–617.
19. Morimoto,K., Sato,S., Tabata,S. and Nakai,M. (2003) A HEAT-repeats containing protein, IaiH, stabilizes the iron-sulfur cluster bound to the cyanobacterial IscA homologue, IscA2. *J. Biochem.*, **134**, 211–217.
20. Rawlings,N.D., Morton,F.R., Kok,C.Y., Kong,J. and Barrett,A.J. (2008) MEROPS: the peptidase database. *Nucleic Acids Res.*, **36**, D320–D325.
21. Mulkidjanian,A.Y., Koonin,E.V., Makarova,K.S., Mekhedov,S.L., Sorokin,A., Wolf,Y.I., Dufresne,A., Partensky,F., Burd,H., Kaznadzey,D. *et al.* (2006) The cyanobacterial genome core and the origin of photosynthesis. *Proc. Natl Acad. Sci. USA*, **103**, 13126–13131.
22. Kelley,L.A. and Sternberg,M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.*, **4**, 363–371.
23. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
24. Capella-Gutierrez,S., Silla-Martinez,J.M. and Gabaldon,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
25. Guindon,S., Dufayard,J.F., Lefort,V., Anisimova,M., Hordijk,W. and Gascuel,O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *System. Biol.*, **59**, 307–321.
26. Zhao,K.H., Zhang,J., Tu,J.M., Bohm,S., Ploscher,M., Eichacker,L., Bubenzer,C., Scheer,H., Wang,X. and Zhou,M. (2007) Lyase activities of CpcS- and CpcT-like proteins from *Nostoc* PCC7120 and sequential reconstitution of binding sites of phycoerythrocyanin and phycocyanin beta-subunits. *J. Biol. Chem.*, **282**, 34093–34103.
27. Saunée,N.A., Williams,S.R., Bryant,D.A. and Schluchter,W.M. (2008) Biogenesis of phycobiliproteins: II. CpcS-I and CpcU comprise the heterodimeric bilin lyase that attaches phycocyanobilin to Cys-82 of beta-phycocyanin and Cys-81 of allophycocyanin subunits in *Synechococcus* sp. PCC 7002. *J. Biol. Chem.*, **283**, 7513–7522.

Downloaded from http://nar.oxfordjournals.org/ by guest on December 11, 2014