A Colorful Pallet of B-Phycoerythrin Proteoforms Exposed by a Multimodal Mass Spectrometry Approach

Tamara, Sem; Hoek, Max; Scheltema, Richard A.; Leney, Aneika C.; Heck, Albert J.R.

DOI:
10.1016/j.chempr.2019.03.006

Document Version
Peer reviewed version

Citation for published version (Harvard):
Tamara, S, Hoek, M, Scheltema, RA, Leney, AC & Heck, AJR 2019, 'A Colorful Pallet of B-Phycoerythrin Proteoforms Exposed by a Multimodal Mass Spectrometry Approach', *Chem*, vol. 5, no. 5, pp. 1302-1317. https://doi.org/10.1016/j.chempr.2019.03.006

Chem

CellPress

A Colorful Pallet of B-phycoerythrin Proteoforms Exposed by a Multimodal Mass Spectrometry Approach

Sem Tamara[a,b], Max Hoek[a,b], Richard A. Scheltema[a,b], Aneika C. Leney[c,*] and Albert J. R. Heck[a,b,d,*]

[a]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands
[b]Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands
[c]School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
[d]Lead Contact
*Correspondence: A.Leney@bham.ac.uk (A.C.L.) and A.J.R.Heck@uu.nl (A.J.R.H.)

SUMMARY (150/150)

Cyanobacteria and red algae represent some of the oldest lifeforms on the planet. During billions of years of evolution they have fine-tuned the structural details of their light-harvesting antenna, called phycobilisomes, that represent one of the most efficient systems for light-harvesting and energy transfer. Yet, the exact details of phycobilisome assembly and energy transfer are still under investigation. Here, we employed a multi-modal mass spectrometric approach to unravel the molecular heterogeneity within B-phycoerythrin, the major phycobiliprotein in the red algae *P. cruentum*. B-phycoerythrin consists of 12 subunits $(\alpha\beta)_{12}$ arranged in ring with the central cavity housing a linker (γ) subunit which is crucial for stabilizing B-phycoerythrin within the phycobilisome. Using top-down MS we unravel the heterogeneity in the γ proteoforms, characterizing the distinct γ chains and multiple isobaric chromophores they harbor. Our data highlight the key role γ plays in phycobilisome organization that enables optimal light transmission.

B-phycoerythrin, phycobilisome, -phycobilins, chromophorylation, proteoform, complex heterogeneity, top-down MS, native MS, multi-modal mass spectrometry, photosynthetic complexes

INTRODUCTION

Phycobilisomes are large light-harvesting antennas that facilitate the conversion of light into chemical energy in different species of cyanobacteria and red algae[1,2]. These MDa protein assemblies are formed by a morphologically distinct core complex and rod-like assemblies that are attached to the core[3]. Both structural units consist of stacked disc-shaped phycobiliproteins (PBP), which are themselves multi-chain protein complexes with distinct photochemical properties. The core of the phycobilisome is primarily comprised of allophycocyanin (APC; λ_{max} 651 nm)[4] while the rods incorporate phycocyanin (PC; λ_{max} 620 nm)[5] and phycoerythrin (PE; λ_{max} 565 nm)[6] that are situated proximal (PC) or distal (PE) to the core. Specific topologies of PBPs within phycobilisomes facilitate spontaneous excitation energy flow as energy transitions decrease from the rods to the core[7]. The distinct photochemical properties of these PBP types are largely defined by tetrapyrrole prosthetic groups (called bilins) that are covalently attached to the cysteine residues of the polypeptide chains[8].

The phycoerythrin family of PBPs are unique to red algae and cyanobacteria and have the most pronounced fluorescent and colorant properties of all PBPs with fluorescence quantum yield (Q) in the range of 0.82-0.98[9]. As such, phycoerythrins (PEs) have numerous biotechnological applications as dyes and fluorescent tags [10,11]. One of the most studied PEs is B-phycoerythrin (B-PE; Q = 0.98)[12–15], which is the most abundant PBP (~42% of all colorant proteins) in the red algae *Porphyridium cruentum*[16]. B-PE is known to be a hetero-

13-mer that contains six α, six β, and one linker protein subunit termed γ[6]. The primary architecture of the B-PE assembly consists of two overlaid disc-shaped $(\alpha\beta)_3$ hexamers which form a central cavity that is filled by a single γ subunit[17]. The B-PE complex has two types of bilin prosthetic groups that are covalently bound to cysteine residues: phycoerythrobilin (PEB; λ_{max} 550 nm) and phycourobilin (PUB; λ_{max} 498 nm)[18]. The maximum absorbance of B-PE is at 565 nm, which originates from B-PE assemblies harboring a high content of PEB molecules. It has been well-documented that each 17.8 kDa α chain carries 2 PEB prosthetic groups, while the 18.5 kDa β chain harbors 3 PEB molecules. Typically, bilins are connected *via* a single thioester bond to the cysteine residue, however one of the bilin prosthetic groups of the β chain is connected through two thioester linkages[19]. Compared to the available knowledge about the α and β chains, the nature of the γ chain and the chromophores it carries has so far remained much more elusive.

The γ chain is important as it stabilizes the tertiary structure of phycoerythrins by holding the discs of $(\alpha\beta)_3$ hexamers together[15,17]. Moreover, the bilin prosthetic groups the γ chain harbors enhance the light-absorbance properties without increasing the spacing of phycobiliproteins[2]. Additionally, it has been proposed that the γ chain provides energetic decoupling protecting the photosynthetic reaction center from damage induced by excessive photoexcitation[20]. Initially, in biochemical studies of PEs, the γ subunit was identified as a single band on sodium dodecyl sulfate (SDS) gels and was assumed to be a single protein[6]. Later, reversed phase liquid chromatography (RP-LC) revealed that the γ subunit is represented by at least three distinct polypeptide chains in B-PE assemblies[16,21], however, the exact amino acid sequences and positions of attached bilins were not determined. Overall, the γ subunits are expected to have molecular weights in the range of 27-35 kDa based on the sequences of predicted genes with likely up to four bilins attached to them, supposedly two PEBs and two PUBs[6]. In the related R-PE complex the γ subunit was long proposed to harbor 4 chromophores as well, however five distinct chromphorylated peptides were detected, which was rationalized by presence of several distinct γ subunits[8,22]. Recently, a structural model was reported for the entire phycobilisome from the red alga *Griffithsia pacifica* based on cryo-EM data revealing more details on the structure and conformation of the γ linker subunit[2]. Predicted structures outlined the presence of a chromophore-binding domain on γ that could carry up to 5 bilin molecules. However, due to the variations of the γ subunit sequences within single and, in particular, different algae strains, it is likely that not all γ subunits are identically modified. Moreover, since extensive "class" averaging was performed to obtain the cryo-EM images, heterogeneity in these γ protein sequences present within the PE core and their bilin modifications are difficult to resolve. Thus, alternative methods are indispensable to distinguish and analyze B-PE variants and γ proteoforms separately, allowing the extent of post-translational processing events to be individually characterized and quantified.

Mass spectrometry (MS) is a rapidly emerging tool to monitor protein isoforms present within protein complexes[23–25]. Due to their difference in mass, proteoforms and modifications contained within can be rapidly distinguished and quantified. Indeed, preliminary MS work has already been utilized to reveal details on the bilin architecture and amino acid sequence of the γ subunit in R-phycoerythrin[26,27]. The combination of multiple MS approaches provides complimentary information that is not obtainable from a single MS method, which proved to be advantages for analysis of highly heterogeneous proteins and protein complexes[25,28–30]. Here, we use a combination of bottom-up, top-down, and native MS to explore the structural heterogeneity present within the protein assembly B-PE and its constituent subunits in the red algae *P. cruentum*. We unequivocally determine the co-occurrence of multiple variants of the B-PE assembly and link each of these variants to distinct proteoforms of the α, β, and γ subunits. In our work four distinct polypeptide chains of the γ subunit are identified (one more than previously reported for B-PE from *P. cruentum*[13]), quantified, and fully characterized. These chains harbor different number of bilins ranging from 3 to 5 including both PEB and PUB molecules (more heterogeneous than previously reported for B-PE from *P. cruentum*[21]). Thus, the complete B-PE assembly can carry up to 35 bilin chromophores. In our work, by linking fragment signatures to the structures of isobaric prosthetic groups we unambiguously characterize and position each

PEB or PUB bilin on the γ chains. Such information is relevant as, ultimately, the photochemical properties of the B-PE assembly are a result of the interplay between all co-assembled proteins and the chromophore groups they harbor. Moreover, because topologies of phycobiliproteins in a phycobilisome are influenced by a linker protein and number/type of carried bilin molecules[2] different γ subunits define ordering of B-PE within phycobilisome rods.

RESULTS

Heterogeneity of B-phycoerythrin probed by native MS

As a first step to unravel the structural heterogeneity within B-PE, we measured high-resolution native mass spectra of the fluorescent assembly. These spectra indicated that the B-PE complex is heterogeneous, with several co-occurring charge states present, all originating from assemblies with molecular weights between 260 and 270 kDa (Figure S1A), in line with literature and as expected for $\alpha_6\beta_6\gamma$ assemblies[15]. To further investigate the factors influencing the heterogeneity observed in the native mass spectrum of B-PE, we then performed native tandem MS (native top-down MS/MS) experiments on this assembly. The z = 37+ ions of different B-PE variants were isolated and subsequently subjected to collisional activation. These tandem mass spectra revealed the ejection of α and β subunits and residual complexes of B-PE wherein α, β or combinations of the two subunits had been eliminated (Figure 1). Upon collisional dissociation of different precursor ions originating from different assembly variants, the released α and β subunits always had identical masses while the residual high molecular weight fragment complexes, formed by the loss of an α and/or β subunit, exhibited clear mass differences that, thus, can be attributed to the presence of different forms of the γ subunit.

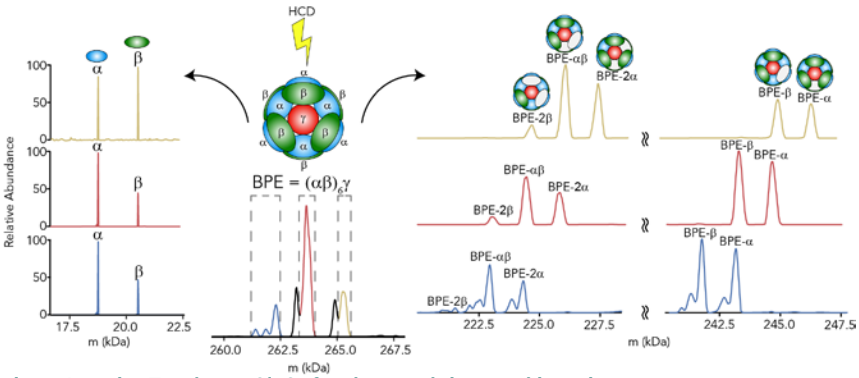

Figure 1. Native Top-down MS/MS of B-phycoerythrin Assembly Variants

Deconvoluted mass spectra observed following HCD fragmentation of B-phycoerythrin (B-PE) precursor ions sprayed under native conditions. Precursor ions (z = + 37) corresponding to B-PE species enclosed with dashed boxes were subjected individually to HCD, which resulted in the ejection of α and β monomeric subunits (left) and the concomitant formation of residual fragmented complexes (right) missing a single α or β subunit or combinations of subunits (2α, 2β, or αβ).

Dissociation of B-PE complex was limited to maximum of two ejected subunits. We did not observe the γ subunit to be ejected, likely due to its topological inaccessibility as it is buried inside the cavity formed by toroids of $(\alpha\beta)_3$ hetero-hexamers. Alternatively, in case the γ is present in a large number of proteoforms within B-PE assembly, even if it gets detached from the complex its intensity would be spread over numerous ion peak distributions, thus avoiding detection. Finally, we performed pseudo-MS3 experiment (Supplemental Experimental Procedures) to test whether it is possible to detect γ subunit detaching from residual complexes formed upon collisional activation. For this, following in-source activation residual complexes corresponding to B-PE assembly lacking 2α, αβ, or 2β were isolated and fragmented with HCD (Figure S1B). Alongside the monomeric products of dissociation in this experiment we observed dimers and trimers of α/β subunits, however neither the leftover parts of precursors containing γ subunit nor intact γ subunit itself were detected (Figure S1C-D).

Characteristics of proteins comprising the B-phycoerythrin assemblies

To identify the heterogeneity behind the γ subunit as revealed by native MS (here and earlier by Leney et al.[15]), we next denatured B-PE, digested it into peptides and the resulting peptides were analyzed by bottom-up LC-MS/MS. The data revealed that the most abundant proteins in the sample were as expected the α and β chains of B-PE. Identifying any γ subunits present, however, is more challenging since their mature sequences as well as positions, type, and number of bilins have not been explicitly reported. Thus, a database was set up incorporating all of the possible sequences obtained from the *P. cruentum* genome[31]. By comparing the peptides identified with this dedicated protein database, the presence of 4 different γ subunits could be revealed in the list of the most abundant identified proteins (Figure 2A and Table S1). Next to these we also detected peptides originating from other linker protein(s) that did not carry any bilin molecules, albeit typically all at a much lower abundance. Such linker proteins are more prominent for PBPs in the proximal parts of the rods in relation to the core of the phycobilisome[32]. Along with B-PE-related proteins bottom-up LC-MS/MS revealed subunits of R-PE detected albeit with significantly lower abundances, likely because of their similar biochemical properties that resulted in their co-purification.

To verify the presence of multiple γ subunits within B-PE, we next denatured B-PE and separated the intact proteins using reversed-phase HPLC. Consistent with the bottom-up results, the data showed two abundant signals corresponding to the α (~9 min retention time) and β subunits (12.5 min) as well as several lower abundant peaks with shorter retention times (2.5-5 min) (Figure 2B). Peak splitting observed for β subunit was attributed to shifted retention time of oxidized proteoforms. Additionally, we observed a peak at 5.5 min retention time that was assigned to the colorless linker protein (Figure S2; fraction A06). The shorter retention times of the γ subunits can be rationalized by these proteins harboring more hydrophilic residues than the α and β subunits. To verify whether the eluting proteins are chromophorylated subunits of B-PE we measured the absorbance spectra of the fractions (Figure 2C). Clearly, three distinct absorption profiles were observed that resembled previously reported absorbance spectra of the B-PE subunits[13]. Additionally, we measured the absorbance for fractions corresponding to different γ chains (Figure S2; fractions A02-A04), whereby we observed that all these fractions resulted in alike absorbance profiles. However, such absorption data cannot directly reveal the number and positions of the bilin prosthetic groups these subunits harbor. Therefore, we next set out to further characterize all proteoforms of the γ subunits.

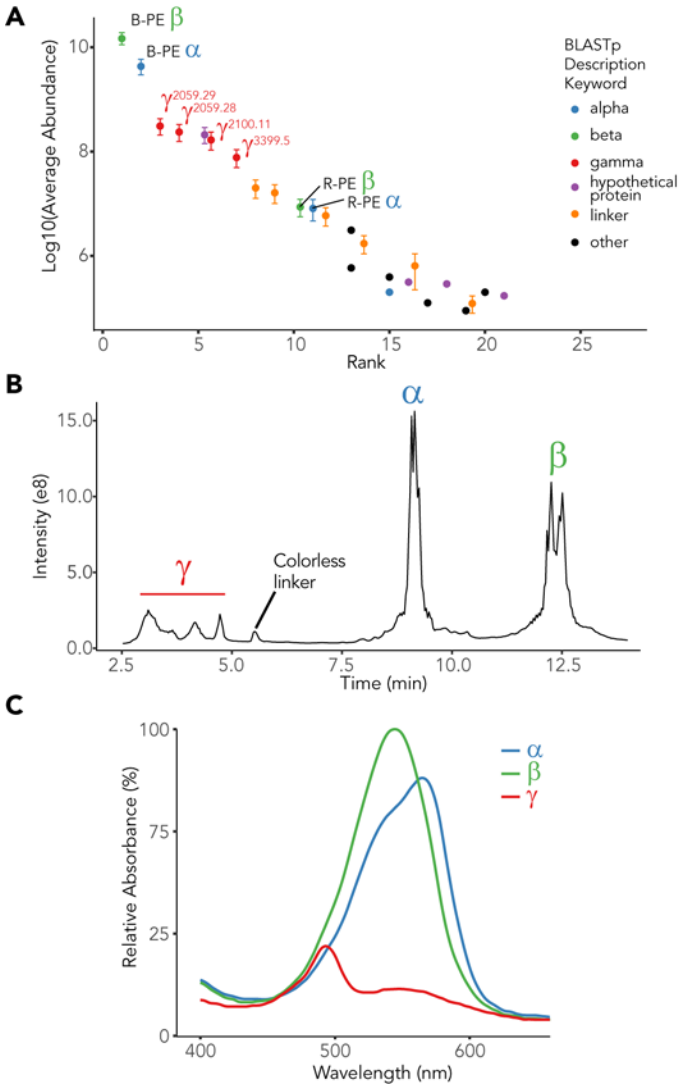

Figure 2. Identification of Distinctive B-phycoerythrin Subunits

Overview of all proteins, with distinct physico- and photochemical properties, identified in the B-phycoerythrin sample. (A) Proteins identified in the bottom-up LC-MS/MS analysis and ranked based on the combined abundance of respective peptides in LC-MS. The errorbars represent standard error of the mean abundance. (B) Reversed-phase (RP) LC separation of the intact subunit proteins in the B-phycoerythrin sample represented by base peak intensities against the retention time. (C) Absorbance spectra of fractions collected after RP-LC, corresponding to the α (blue), β (green), and four γ (red) subunits.

Characterizing γ Subunit with "Sequence Tags" and Mass Matching

Although traditional shotgun LC-MS/MS methods provide means for fast and sensitive identification of proteins, the information on mature proteoforms and their chromophorylation stoichiometries is lost because of protein digestion. Top-down MS circumvents this problem by analyzing the proteins intact and, therefore, potentially allows for the identification of all proteoforms present.

In top-down MS experiments of non-modified proteins the backbone fragments typically provide direct sequence information. However, the γ proteoforms studied here harbor various bilin modifications. These chromophores heavily influence and complicate the observed dissociation patterns (Figure S3A). Moreover co-isolation of co-eluting proteoforms limit further the straightforward retrieval of sequence information. Therefore, we here characterized the γ proteins using an alternative approach, wherein the

proteoforms were identified by mass matching to the theoretical proteoform masses and further verified with bottom-up LC-MS/MS.

Using fast low resolution (7,500 at 200 m/z) recording of mass spectra, several γ proteoforms were successfully resolved by top-down LC-MS (Figure S3B; Supporting Data 1), consistent with the broad elution peaks observed in the total ion chromatogram corresponding to the γ subunits (Figure 2B). The intact masses alone provide significant insight into the heterogeneity present within the γ subunit with multiple molecular weights being identified for each proteoform (Figure S3C). Elucidating their sequences from intact masses only, however, is challenging. This is due to two reasons. Firstly, the protein sequences of B-PE γ subunits from *P. cruentum* are missing from the conventionally used protein databases (*e.g.* UniProt). Secondly, information on the bilin content and localization in the γ protein sequences is incomplete. For these reasons, we designed an unbiased screening approach by building a custom database that incorporated all of the sequences from the *P. cruentum* genome with varying number of chromophorylations. However, in our initial attempts no matches were found between the experimental and theoretical masses. Thus, we hypothesized that considering the number of γ proteoforms identified, post-translational sequence processing events could have occurred.

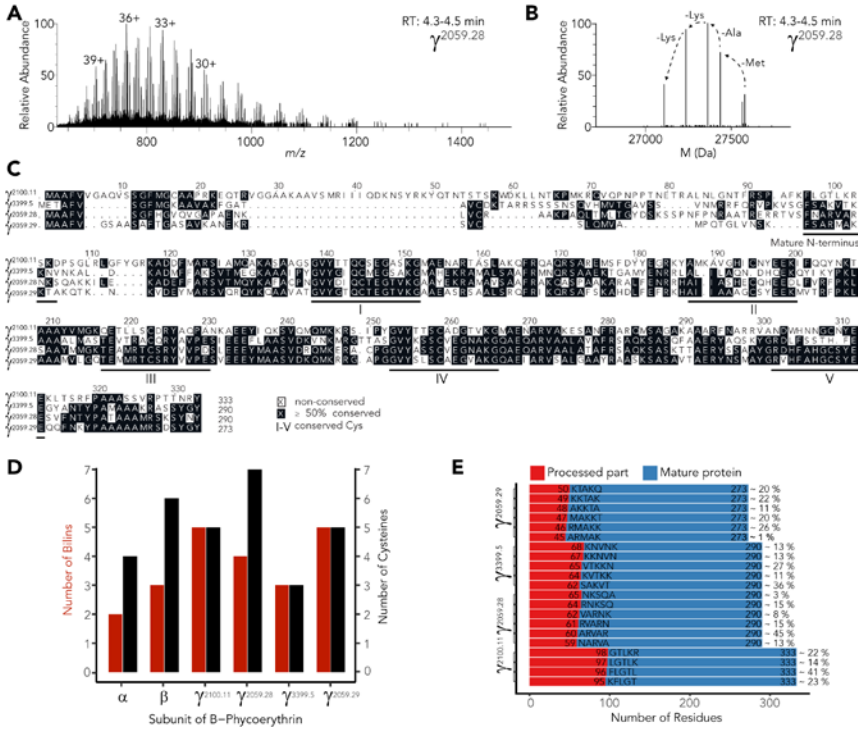

Figure 3. Determination of the Pallet of γ Subunit Proteoforms Identified by Top-down LC-MS

(A) A full LC-MS scan displaying a mixture of charge envelopes for different co-eluting proteoforms of $\gamma^{2059.29}$. (B) Deconvolved mass spectrum of (A) reveals several proteoforms with mass differences that originate from the sequential deletion of specific amino acid residues due to protein processing. (C) Predicted sequences of the γ subunits aligned by using the MUSCLE algorithm. Conserved regions corresponding to potential chromophore-binding sites are annotated with roman numbers. (D) The number of chromophorylations observed (red bars) on the most abundant proteoform and total number of cysteines on the detected B-PE subunits (black bars). (E) Relative abundances of each of the processed sequence variants for each γ subunit.

While MS scans of the α and β subunits were dominated by a single most abundant proteoform (Figure S4), for the γ subunit we observed the co-elution of various proteoforms differing in masses likely originating from the addition or deletion in the sequence of a few amino acid residues (Figure 3A, B). Excitingly, by searching these "sequence tags" against

the custom database of the γ subunit proteoforms we identified one of the γ subunits, which we named $\gamma^{2059.29}$ wherein the superscript refers to the contig number annotated in the DNA sequencing[31]. The observed position of the "sequence tag" indicated a N-terminal processing of this chain, which previously was proposed for γ subunits of R-PE, based on the fact that they require a transit peptide for transfer into the chloroplasts[33]. Interestingly, upon further analysis, top-down LC-MS revealed truncated sequence variants for all the γ subunits detected showing that all of them in the final phycobilisome complex require cleavage of the transit peptide prior to complex assembly. Indeed, consistent with these results, no peptides in the N-terminal transit peptides of the γ subunits were detected by bottom-up (data not shown).

To investigate γ protein processing and chromophorylation sites, the sequences of the γ subunits were aligned making use of the MUSCLE algorithm[34]. This alignment revealed that all the γ subunits contain a conserved region, which has been recently pointed out as the chromophore binding domain of the γ subunits[2]. In three out of four γ chains this domain contained 5 conserved cysteine residues, which can be regarded as potential sites for bilin attachment ($\gamma^{3399.5}$ is missing 2 out of total 5 conserved cysteines) (Figure 3C). We extended our custom database of theoretical proteoforms with truncated forms of γ based on the experimentally detected "sequence tags", which indicated at mature N-termini.
Using the extended custom database, we identified multiple sequence variants of the B-PE subunits with varying number of occupied chromophorylation sites (Figure 3D). This is in contrast to the recent cryo-EM study, wherein only complete occupancy of cysteine residues was reported, highlighting the advantages mass spectrometry can provide in revealing the structural heterogeneity within the B-PE assembly. Using top-down LC-MS, we were able to assess the relative abundance of each of the γ proteoforms, the data revealing both the extent of protein processing and the variable bilin occupancies for each γ subunit (Figure 3D, E). Two out of four γ subunits were primarily represented by proteoforms having all their cysteine residues fully occupied by bilins. The polypeptide chain of $\gamma^{2059.28}$, which has the highest number of cysteine residues, was not detected in a form fully saturated with chromophores. Finally, $\gamma^{3399.5}$, which lacks two conserved cysteines, was observed harboring only 1, 2, or 3 bilin molecules. In agreement with previously reported data, the α and β subunits predominantly carry 2 and 3 bilins per subunit, respectively, resulting in half of the total number of cysteines being occupied (Figure 3D). Heterogeneity of γ subunit revealed in our work for B-PE from *P. cruentum* is in agreement with cryo-EM data recently reported for phycobilisome from *G. pacifica*, which has R-PE as a primary type of phycoerythrin. For *G. pacifica* 5 distinct isoforms of γ subunit were detected, of which four harbored 5 chromophores and one had 4 bilin molecules attached[2]. The complete list of matched proteoforms of α, β, and the various γ subunits of B-PE from *P. cruentum* can be found in the Supporting Data 1. Additionally, to verify the most prominent proteoform within B-PE assembly we collected RP-LC fraction corresponding to the $\gamma^{2059.29}$ and performed direct characterization of the most abundant proteoform with EThcD MS/MS. All 5 predicted bilins with isobaric masses of 586.279 Da were detected by corresponding fragments and positioned at the conserved chromophorylation sites (Figure S5 and Supplemental Experimental Procedures).

Reconstruction of the Native Mass Spectrum of B-PE from the Qualitative and Quantitative Data on All α, β, and γ proteoforms

Mass matching of features extracted from the top-down LC-MS runs resulted in the identification of a wide variety of proteoforms of all B-PE subunits. To validate this, we used a recently developed computational approach[30] to recreate a native mass spectrum based on the intensities of mass features detected in the top-down LC-MS data. Thus, the mass of the γ proteoforms based on their average abundances from the intact LC-MS data were plotted alongside the mass of the B-PE complex as determined by native MS whereby the mass of $(\alpha\beta)_6$ had been subtracted (Figure 4). Direct comparison of these two profiles showed a high correlation of 0.9 (Figure 4) indicating that the γ subunits that participate in formation of different B-PE variants have been explicitly and correctly identified by our top-down mass spectrometry approach. Additionally, it confirms that the γ subunits are the dominant factor contributing to the mass heterogeneity within the full B-PE assembly.

Based on this analysis we conclude that the most abundant B-PE assembly is formed by $(\alpha\beta)_6\gamma^{2059.29}$ with 35 attached chromophores in total (Figure 4 and Figure S6).

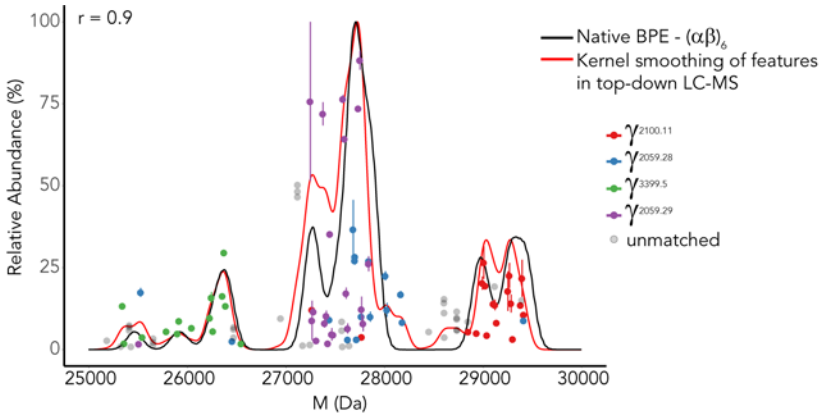

Figure 4. Reconstruction of the Features Observed in the Native Mass Spectrum of the Intact B-phycoerythrin Assembly Based on the Proteoforms Detected in the Top-down LC-MS Analysis

Summing over all mass features detected in the top-down LC-MS runs provides a mass profile that correlates well with the profile observed in the native mass spectra of B-phycoerythrin (here displayed by subtraction of the $(\alpha\beta)_6$ mass). Mass features are color-coded in accordance with the matching γ proteoform masses from the custom database. Error bars represent standard error of the mean calculated for three technical replicates.

Further molecular diversity introduced by the isobaric PEB and PUB chromophores

The absorbance maxima observed for the B-PE complex and its subunits (Figure 2C, Figure S7) indicate the presence of two types of chromophores, namely, phycoerythrobilin (PEB, absorbance maximum at 550 nm) and phycourobilin (PUB, absorbance maximum at 498 nm). In the PubChem database, the molar mass of PEB (CID 5289229) is 586.689 g/mol while the molar mass of PUB (CID 5289229) is 590.721 g/mol. Because the chromophore moiety binds loosely to the cysteine residue – being readily detached upon HCD in bottom-up or top-down LC-MS/MS experiments – the mass of the bilin molecule could be determined. Our data showed that masses of the majority of chromophorylated peptides from B-PE subunits indicate at mass shift of 586 Da corresponding to PEB (Figure S8). For some positions we also observed several PEB derivatives that displayed the addition or deletion of 1-2 hydrogens, however not a single peptide was observed with mass shift of 590 Da, which is the theoretical mass of PUB. Based on this evidence we conclude that contrary to the theoretically expected masses both PUB and PEB bilin moieties when attached to B-PE subunits are isobaric and have a monoisotopic mass of 586.279 Da. Having identical masses, the distinct difference in absorbance of PEB and PUB can be reasoned as the chromophores have different π conjugation systems. Thus, taking into account that the double carbon-carbon bond is nearly twice as strong as the single bond (*D* = 602 kJ/mol and 346 kJ/mol, respectively)[35], we hypothesized that upon MS/MS the extended π conjugation system of PEB should prevent formation of fragments containing two inner pyrroles (annotated as BC in Figure 5A, B), thus producing distinctive fragmentation signatures different than for PUB. Indeed, MS/MS spectra of the chromophorylated peptides of the B-PE subunits were dominated by either three-pyrrole (*m/z* 466.23^{+1} and 464.22^{+1}) or two-pyrrole (*m/z* 343.17^{+1}) fragment ions (Figure 5B) in the low mass region indicative for either PEB or PUB chromophore, respectively. Moreover, peptides chromophorylated with PEB displayed fragment ions consistent with three pyrroles closest to the attachment site (*m/z* 466.23^{+1} annotated as ABC in Figure 5A, B). This fragmentation pattern agrees with the proposed conjugation system in the PEB molecule. Notably, slightly higher abundance of tri-pyrrole fragment BCD versus ABC is supported by the bond energetics with a dissociation energy of 618 kJ/mol for ABC and 602 kJ/mol for BCD calculated by summing dissociation energies of respective bonds (C-S and C-C single bonds for ABC; C=C double bond for BCD)[35].

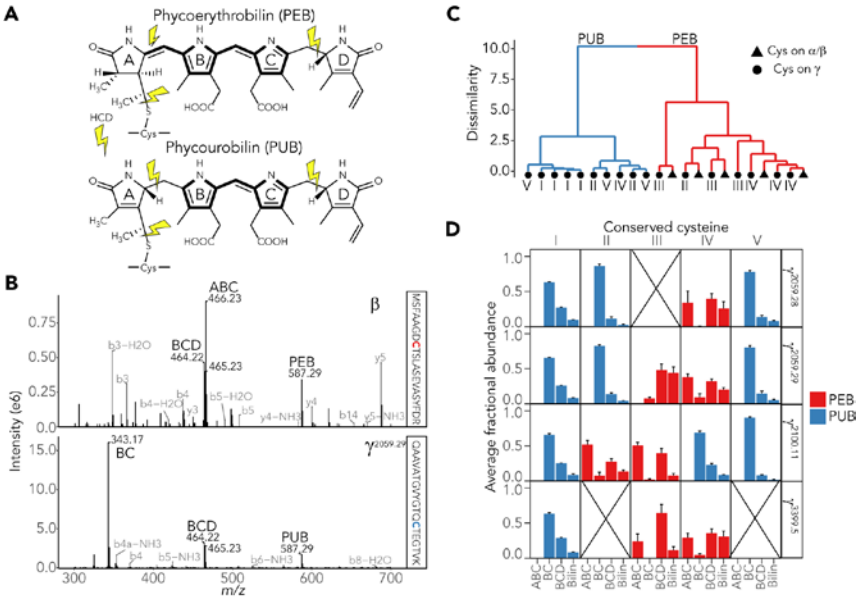

Figure 5. Differentiating between isobaric phycoerythrobilin (PEB) and phycourobilin (PUB) moieties attached to the B-PE subunits

(A) Chemical structures of the isobaric PEB and PUB moieties attached to cysteine residues with the most prominent fragmentation channels observed upon HCD indicated. (B) Examples of tandem mass spectra of peptides chromophorylated with PEB (β subunit) or PUB ($\gamma^{2059.29}$ subunit). (C) Dendrogram representing hierarchical clustering of chromophorylation sites based on abundances of four characteristic fragments (ABC, BC, BCD, and Bilin). Each dot or triangle represents chromophorylation site on α/β or γ subunit, respectively. (D) Abundances of bilin fragments calculated for each of the conserved cysteines of the γ chains, color-coded based on the hierarchical clustering in (C). Error bars represent standard error of the means.

Fragmentation signature peaks allowed us to unambiguously characterize all the high-stoichiometric chromophorylations in each of the B-PE subunits by quantifying the abundance of characteristic bilin fragments. First, chromophorylated peptides from α/β subunits displayed similar bilin fragmentation and were clustered together based on four characteristic fragments supporting presence of PEB on all sites (Figure 5C). Hierarchical clustering directly provided two groups of chromophorylation sites in agreement with expected presence of two bilin types suggested by absorbance profiles of B-PE subunits (Figure 2B). Distinctively, doubly-linked PEB on β subunit (Cys50-Cys61) produced less ABC fragment upon dissociation compared to the singly-linked PEB molecules (Figure S9), ultimately allowing us to distinguish Cys-PEB, Cys-PUB and Cys-PEB-Cys linked bilins.
In the γ subunits several conserved cysteine residues – annotated by roman numbers I-V in Figure 3C – displayed a distinct preference for either PEB or PUB moieties (Figure 5D). Recently, it was proposed that for the red algae *G. pacifica* the bilins of the γ subunits connect with those of the β subunit to allow efficient energy transfer within rods of the phycobilisome[2]. Using the recently published structural model of the phycobilisome from *G. pacifica* we color-coded in Figure 6 bilins of the γ and β subunits of phycoerythrins in accordance with the bilin types revealed for B-PE subunits from *P. cruentum* in our study. Consistent with *G. pacifica*, our identified PUB sites on the γ subunit within B-PE in *P. cruentum* also connect with the PEB molecules on the β subunits allowing them to transfer energy efficiently when the initial chromophore is excited. The data acquired within this work, therefore, offer an explanation as to how we determined that PUB molecules of B-PE carried by the most abundant γ subunits are participating in the shortest energy transfer pathways indicated by the black arrow lines in the Figure 6A. In a phycobilisome, such a layout of chromophores in the phycoerythrin, away from the core parts of the rods, can extend the spectral range for more efficient light-harvesting and allow for more energetic excitation-relaxation transitions of PUB molecules to happen prior to less energetic primarily PEB-mediated energy transfer within phycoerythrins in the proximal parts of the rods of a phycobilisome. To a lesser investigated extent PUB groups of the γ subunits might act as light quenchers by absorbing less energetic emission from excited PEB molecules.

The overall structure of stacked phycoerythrin complexes reveals that the γ subunits from one phycoerythrin complex link to another *via* its N-terminus, being in close proximity to bilin groups on conserved cysteines III and IV of the γ subunit from subsequent phycoerythrin (Figure 6B). Diverse bilin combinations at these positions (Figure 5D) may regulate the ordering of B-PE complexes in the rods of the phycobilisome.

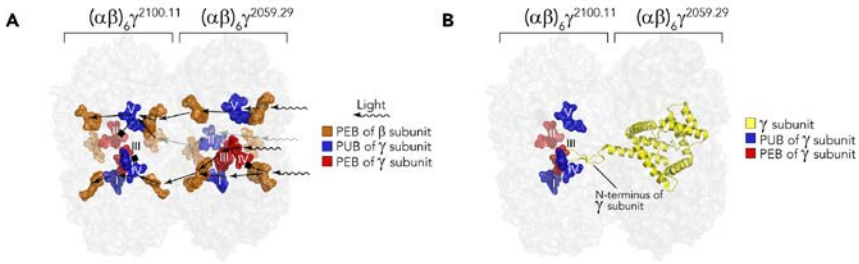

Figure 6. γ Subunit Facilitates Energy Transfer and Linking of Phycoerythrins in Phycobilisome
(A) Energy transfer pathways within phycoerythrin complexes as stacked in the structural model of the intact phycobilisome (PDB entry: 5Y6P, phycobilisome from *G. pacifica*). The bilin chromophores on the γ subunits are colored to resemble bilin types determined in the current study for $\gamma^{2059.29}$ and $\gamma^{2100.11}$. Rings of $(\alpha\beta)_6$ are represented as the grey transparent surface. The black lines schematically represent the shortest energy transfer pathways through the phycobilisome complex. (B) N-terminus of the γ subunit from one phycoerythrin lies within the complex in close proximity to bilin groups on conserved cysteines III and IV (Figure 3C) of the γ subunit from a neighboring phycoerythrin complex.

DISCUSSION

The distinct photochemical properties of each phycobiliprotein drive an increasing demand for their industrial utilization. However, progress in discovering the molecular details of how these protein complexes function and – thus – opening new biotechnological capabilities has been hindered due to their high complexity and structural heterogeneity. Moreover, it is only when we can identify how the phycobilisome components function individually, that we can attempt to unravel the mechanistic details behind how the intact phycobilisome operates. One of the ways to extend understanding of the light harvesting machineries and improve their exploitation is through explicit characterization of phycobiliprotein variants and constituent proteoforms. Here, by using different tiers of mass spectrometric analysis we were able to determine the heterogeneity of B-PE in unprecedented detail. Native MS allowed us to detect multiple variants of the intact B-PE assembly and gain insights into its stoichiometry. Top-down LC-MS on the intact subunits revealed the heterogeneity within the B-phycoerythrin subunits and provided means for characterization and quantification of the prominent proteoforms. Lastly, bottom-up LC-MS/MS facilitated identification and localization of prosthetic groups on each of the B-PE subunits. Taken together, B-phycoerythrin was detected as mixture composed of six α, six β and one of four distinct γ subunits: $\gamma^{2059.29}$, $\gamma^{2059.28}$, $\gamma^{3399.5}$, and $\gamma^{2100.11}$ whereby 35 bilin molecules decorate on average each of the B-PE protein complexes. Finally, we demonstrated that by using bottom-up LC-MS/MS it was possible to unambiguously distinguish between isobaric tetra-pyrrole chromophores attached to each of the modified cysteine residues.

Together, our work reveals high levels of structural heterogeneity present within B-PE. Interestingly, this heterogeneity is confined to the γ subunit; the subunit that links B-PE's together and is essential for its stability. Our results indicate that 4 distinct γ subunits are present which is in agreement with the cryo-EM structure whereby γ subunits are required to link the individual PE complexes in the rods of the phycobilisome[2]. Furthermore, different N-terminal regions of distinct γ subunits that participate in linker-linker contacts might influence absorbance and emission properties of involved prosthetic groups, as protein microenvironment significantly influences fluorophore-mediated light transmission[36]. Additionally, we show the most abundant γ subunits ($\gamma^{2059.29}$ and $\gamma^{2100.11}$) have all 5 cysteines saturated with bilins. However, it is important to note that this is not the case for all γ subunits as $\gamma^{2059.28}$ and $\gamma^{3399.5}$ carry 4 and 3 bilins, respectively. The γ subunit is the only phycobiliprotein within phycoerythrin to contain phycourobilin (PUB). The

nature of PUB is crucial for the absorption of 495 nm light through the phycobilisome. Here, we locate different combinations of PUB and PEB groups on γ subunits wherein two chromophore positions that participate in linking of phycoerythrins demonstrate unique chromophorylation patterns. Thus, we speculate that varying chromophorylation patterns and distinct primary structures of linker subunits drive joining and ordering of phycoerythrins for efficient light transmission throughout the rods of phycobilisomes. Overall, we expect that the detailed molecular knowledge gathered here will provide strong foundation for further investigations into how these large macromolecular machines function and add important detail about how energy may be most efficiently transferred through these light-harvesting complexes. Moreover, unravelling the complexity of the phycobilisome will prove essential for the further applications of such systems in science and industry.

EXPERIMENTAL PROCEDURES

Sample Preparation

B-phycoerythrin (B-PE) was purchased from Thermo Fischer Scientific. For bottom-up LC-MS/MS analysis B-phycoerythrin was reduced with 20 mM TCEP at room temperature for 30 minutes and alkylated with 20 mM chloroacetamide for 30 minutes in the dark. Digestion of proteins was done overnight at 37 °C with trypsin (Promega Benelux, Leiden, The Netherlands) at a protein-to-enzyme ratio of 50:1 (weight/weight). Samples were kept at pH > 7 prior to LC-MS/MS in order to prevent interconversion between phycourobilin and phycoerythrobilin.

For top-down LC-MS/MS protein samples were buffer exchanged into 0.1% formic acid by using 3 kDa molecular weight cutoff centrifuge filters (Amicon Ultra, Merck KGaA, Darmstadt, Germany) and then diluted to 1 µg/µl final concentration.
B-PE sample was prepared for native MS experiments by several cycles of buffer exchange into aqueous ammonium acetate. Centrifugal filters (Amicon, Ultra Merck KGaA, Darmstadt, Germany), which were used in the buffer exchange procedure, had a molecular weight cutoff at 10 kDa. The final concentration of the ammonium acetate was 300 mM and the pH was adjusted to 7.5.

Bottom-up LC-MS/MS Analysis

Separation of the peptides from the digested B-PE was performed on an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany). Samples were loaded on a 100 µm x 20 mm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 µm) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50 µm x 500 mm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 µm) (Agilent Technologies, Amstelveen, The Netherlands). A 2-5 µL injection of peptides was used, corresponding to ~0.05 µg of material. The LC-MS/MS run time was set to 40 min with flow rate of 300 nL/min. Mobile phases A (water/0.1% formic acid) and B (80% ACN/0.1% formic acid) were used for gradient time of 35 minutes: 13 to 44% B for 20 minutes, and 44 to 100% B over 3 minutes. Samples were analyzed on a Thermo Scientific Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer. Nano-electrospray ionization was achieved using a coated fused silica emitter (New Objective, Cambridge, MA, USA) biased to 2 kV. The mass spectrometer was operated in positive ion mode and the spectra were acquired in the data-dependent acquisition mode. Full MS scans were acquired with resolution setting set to 60,000 (200 m/z) and at a scan mass range of 375 to 2,000 m/z. Automatic Gain Control (AGC) target was set to 4e5 with maximum injection time of 50 ms. Data dependent-MS/MS (dd-MS/MS) scans were acquired at 30,000 resolution (at 200 m/z) and with mass range of 200 to 2,000 m/z. AGC target was set to 5e4 with maximum of injection time defined at 54 ms. 1 µscan was acquired both for full MS and dd-MS/MS scans. Data dependent method was set to isolation and fragmentation for the cycle time set to 5 seconds. Parameters for isolation/fragmentation of selected ion peaks were set as follows: isolation width – 1.6 Th; HCD normalized collision energy (NCE) – 28%; mass analyzer – Orbitrap.

Top-down LC-MS/MS Analysis

Chromatographic separation of intact protein samples was conducted on a Thermo Scientific Vanquish Flex UHPLC system equipped with MAbPac RP 2.1 mm x 50 mm

column. 2 µg of material was loaded on the column heated to 80 °C. LC-MS runtime was set to 22 minutes with flow rate of 250 µL/min. Gradient elution was performed using mobile phases A (water/0.1% formic acid) and B (ACN/0.1% formic acid): 25 to 46% B for 14 minutes.

All top-down MS experiments were performed on a Thermo Scientific Q Exactive HF-X instrument (Thermo Fisher Scientific, Bremen, Germany)[37]. LC-MS data were collected with instrument set to the Intact Protein Mode. For analysis of intact proteins, a resolution of 7,500 at 200 *m/z* was used. Full MS scans were acquired for the range of 150 to 2,000 m/z with AGC target set to 3e6. Maximum of injection time was defined at 16 ms with 1 µscan recorded.

Absorbance Measurements

B-PE subunits separated with RP-LC were collected at the time of elution following loading of 25 µg of material on the Thermo Scientific MAbPac RP LC column (Thermo Fisher Scientific). Absorbance spectra were measured for the range 400-750 nm on the Thermo Scientific Multiscan GO spectrophotometer (Thermo Fisher Scientific, Ratastie, Finland). Fractions of 250 µl corresponding to each subunit were loaded into 96-well plate. Spectra were recorded in the precision mode and corresponding absorbance values were exported with Thermo Scientific SkanIt Software (Thermo Fisher Scientific). Background was measured as of the respective buffer and subsequently subtracted from absorbance values of the samples.

Native Top-down MS/MS on QE-UHMR Mass Spectrometer

B-PE at concentration of ~2 µM was introduced into Q Exactive mass spectrometer with Ultra High Mass Range (QE-UHMR, Thermo Fisher Scientific, Bremen, Germany) via in-house pulled gold-coated borosilicate capillaries. Sample was sprayed at capillary voltage set to 1.3 kV in positive ion mode. The following mass spectrometer parameters were used: collision gas – Nitrogen; AGC mode – fixed; noise level – 2. Ion transmission settings were as follows: S-lens voltage – 25 V, inject flatapole offset – 10 V, bent flatapole DC – 4 V, gate lens voltage – 3. Resolution setting was 8,750 (at 200 *m/z*) and ion injection time was set to 100 ms. Instrument calibration was performed using cesium iodide clusters up to 11,000 *m/z*. Scan mass range was between 300 and 20,000 Th for all experiments. For measurements of intact complex, source trapping voltage was set to 25 V and HCD voltage was defined at 10 V. For native MS/MS experiments peaks of interest were isolated with 8-10 Th width, ion injection time was increased to 500 ms, and HCD voltage was elevated to 150 V. Each spectrum was obtained by averaging ~100 microscans in the time domain. Pseudo-MS3 analysis of B-PE is described in Supplemental Experimental Procedures.

Identification of B-PE Proteoforms

First, the gamma subunits were identified through matching of the sequence tags observed in full MS. For these filtered distinct sequences a database of all possible sequence truncations, up to the index of detected sequence tag, were created with addition of variable number of chromophores, up to the number of cysteines. Then, these masses of created proteoforms were matched with 2 Th tolerance to the mass features in technical triplicate of LC-MS runs. All the proteoforms matched in at least two out of three runs were manually verified with information available from bottom-up LC-MS/MS regarding the maximum number of detected chromophores per gamma subunit and respective sequence coverage provided by detected peptides. The most abundant proteoform of γ subunit was investigated by direct injection and EThcD MS/MS on Thermo Scientific Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (see Supplemental Experimental Procedures).

Data Analysis

Raw bottom-up LC-MS/MS data was analyzed with Proteome Discoverer 2.2 (Thermo Fisher Scientific) equipped with Byonic nodes (Protein Metrics, Cupertino, USA). Following parameters were used for database search. Protease: Trypsin (full). Variable modifications: Met oxidation; Cys carbamidomethyl; Cys chromophorylations of 586.279 and 590.31 Da. Protein sequence database was generated based on recently published genome of *P. purpureum* (*P. cruentum*)[31]. Sequence alignment was performed in R with the use of "msa" package[38] and MUSCLE algorithm[34]. Top-down LC-MS raw files were deconvoluted by

Sliding Window ReSpect algorithm available in Protein Deconvolution 4.0 software package (Thermo Fisher Scientific). Zero charged mass distribution profiles were obtained from raw native mass spectra with UniDec[39]. Structural visualization of phycoerythrin complexes was done in PyMOL (Schrödinger). Chemical structures of bilin molecules were drawn in ChemDraw (PerkinElmer). All additional data analysis was performed in R; hierarchical clustering was performed using algorithm that implements Ward's criterion[40]; data was visualized with ggplot2 package[41].

DATA AND SOFTWARE AVAILABILITY

The data have been deposited to the ProteomeXchange Consortium via the PRIDE[42] partner repository with the dataset identifier PXD011275. The native MS data relevant to the study is available upon request.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Figures S1-S9, Table S1, and Supplemental Experimental Procedures.

Supporting Data 1. Detailed Table of Proteoforms of B-phycoerythrin Subunits.

ACKNOWLEDGMENTS

The Netherlands Organization for Scientific Research (NWO) supported this research through funding of the large-scale proteomics facility *Proteins@Work* (project 184.032.201) embedded in the Netherlands Proteomics Centre, and through the Spinoza Award SPI.2017.028 for AJRH. Additional support came through the European Union Horizon 2020 program FET-OPEN project MSmed (Project 686547), and the European Union Horizon 2020 program INFRAIA project Epic-XS (Project 823839). We thank Aline Tschanz for help in acquiring preliminary data for the project.

AUTHOR CONTRIBUTIONS

Conceptualization, S.T., A.C.L., and A.J.R.H.; Methodology, S.T., A.C.L., and A.J.R.H.; Investigation, S.T. and M.H.; Software, S.T. and R.A.S.; Formal Analysis, S.T.; Visualization, S.T.; Writing – Original Draft, S.T., A.C.L., and A.J.R.H.; Writing – Review & Editing, S.T., M.H., A.C.L., A.J.R.H.; Funding Acquisition, A.J.R.H.; Resources, A.J.R.H.; Supervision, A.C.L., R.A.S., and A.J.R.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES AND NOTES

1. Grossman, A.R., Schaefer, M.R., Chiang, G.G., and Collier, J.L. (1993). The phycobilisome, a light-harvesting complex responsive to environmental conditions. Microbiol. Rev. *57*, 725–49. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=372933&tool=pmcentrez&rendertype=abstract.
2. Zhang, J., Ma, J., Liu, D., Qin, S., Sun, S., Zhao, J., and Sui, S.F. (2017). Structure of phycobilisome from the red alga Griffithsia pacifica. Nature *551*, 57–63. Available at: http://dx.doi.org/10.1038/nature24278.
3. Adir, N. (2005). Elucidation of the molecular structures of components of the phycobilisome: Reconstructing a giant. Photosynth. Res. *85*, 15–32.
4. Murakami, A., Mimuro, M., Ohki, K., and Fujita, Y. (1981). Absorption spectrum of allophycocyanin isolated from Anabaena cylindrica: Variation of the absorption spectrum induced by changes of the physico-chemical environment. J. Biochem. *89*, 79–86.
5. Glazer, A.N., Fang, S., and Brown, D.M. (1973). Spectroscopic properties of C-phycocyanin and of its alpha and beta subunits. J. Biol. Chem. *248*, 5679–5685.
6. Glazer, A.N., and Hixson, C.S. (1977). Subunit Structure and Chromophore Composition of Rhodophytan Phycoerythrins. J. Biol. Chem. *252*, 32–42.
7. Glazer, A.N. (1989). Light guides. Directional energy transfer in a photosynthetic antenna. J. Biol. Chem. *264*, 1–4.
8. Nagy, J.O., Bishop, J.E., Klotz, A. V., Glazer, A.N., and Rapoport, H. (1985). Bilin attachment sites in the α, ß, and γ subunits of R-phycoerythrin. Structural studies on singly and doubly linked phycourobilins. J. Biol. Chem. *260*, 4864–4868.
9. Oi, V.T., Glazer, A.N., and Stryer, L. (1982). Fluorescent phycobiliprotein conjugates for analyses of cells and molecules. J. Cell Biol. *93*, 981–986.

10. Torres-acosta, M.A., Monterrey, D., Monterrey, C., Eugenio, A., Sada, G., Ruiz-ruiz, F., Monterrey, D., Monterrey, C., Eugenio, A., Sada, G., *et al.* (2016). Economic Analysis of Pilot-Scale Production of B-Phycoerythrin. Biotechnol Prog. *32*, 1472–1479.
11. Fleurence, J. (2003). R-Phycoerythrin from red macroalgae: Strategies for extraction and potential application in Biotechnology. Appl. Biotechnol. Food Sci. Policy *1*, 63–68.
12. Tang, Z., Zhao, J., Ju, B., Li, W., Wen, S., Pu, Y., and Qin, S. (2016). One-step chromatographic procedure for purification of B-phycoerythrin from Porphyridium cruentum. Protein Expr. Purif. *123*, 70–74. Available at: http://dx.doi.org/10.1016/j.pep.2016.01.018.
13. Bermejo, R., Talavera, E.M., and Alvarez-Pez, J.M. (2001). Chromatographic purification and characterization of B-phycoerythrin from Porphyridium cruentum - Semipreparative high-performance liquid chromatographic separation and characterization of its subunits. J. Chromatogr. A *917*, 135–145.
14. Munier, M., Jubeau, S., Wijaya, A., Morançais, M., Dumay, J., Marchal, L., Jaouen, P., and Fleurence, J. (2014). Physicochemical factors affecting the stability of two pigments: R-phycoerythrin of Grateloupia turuturu and B-phycoerythrin of Porphyridium cruentum. Food Chem. *150*, 400–407. Available at: http://dx.doi.org/10.1016/j.foodchem.2013.10.113.
15. Leney, A.C., Tschanz, A., and Heck, A.J.R. (2018). Connecting color with assembly in the fluorescent B-phycoerythrin protein complex. FEBS J. *285*, 178–187. Available at: http://doi.wiley.com/10.1111/febs.14331 [Accessed September 16, 2018].
16. Redlinger, T., and Gantt, E. (1981). Phycobilisome structure of Porphyridium cruentum: polypeptide composition. Plant Physiol. *68*, 1375–1379.
17. Ficner, R., and Huber, R. (1993). Refined crystal structure of phycoerythrin from Porphyridium cruentum at 0.23-nm resolution and localization of the γ subunit. FEBS J. *218*, 103–106.
18. Sepúlveda-ugarte, J., Brunet, J.E., Matamala, A.R., Martínez-oyanedel, J., and Bunster, M. (2011). Spectroscopic parameters of phycoerythrobilin and phycourobilin on phycoerythrin from Gracilaria chilensis. J. Photochem. Photobiol. A Chem. *219*, 211–216.
19. Lundell, D.J., Glazergt, A.N., I, R.J.D., and Douglas, M. (1984). Bilin Attachment Sites in the α and ß subunits of B-phycoerythrin. J. Biol. Chem. *259*, 5472–5480.
20. Liu, L.N., Elmalk, A.T., Aartsma, T.J., Thomas, J.C., Lamers, G.E.M., Zhou, B.C., and Zhang, Y.Z. (2008). Light-induced energetic decoupling as a mechanism for phycobilisome-related energy dissipation in red algae: A single molecule study. PLoS One *3*, e3134.
21. Swanson, R. V., and Glazer, A.N. (1990). Separation of phycobiliprotein subunits by reverse-phase high-pressure liquid chromatography. Anal. Biochem. *188*, 295–299.
22. Klotz, A. V., and Glazer, A.N. (1985). Characterization of the bilin attachment sites in R-phycoerythrin. J. Biol. Chem. *260*, 4856–4863.
23. Skinner, O.S., Haverland, N.A., Fornelli, L., Melani, R.D., Vale, L.H.F. Do, Seckler, H.S., Doubleday, P.F., Schachner, L.F., Srzentic, K., Kelleher, N.L., *et al.* (2018). Top-down characterization of endogenous protein complexes with native proteomics. Nat. Chem. Biol. *14*, 36–41.
24. Wu, D., Struwe, W.B., Harvey, D.J., Ferguson, M.A.J., and Robinson, C. V. (2018). N-glycan microheterogeneity regulates interactions of plasma proteins. Proc. Natl. Acad. Sci. *115*, 8763–8768. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1807439115.
25. Franc, V., Zhu, J., and Heck, A.J.R. (2018). Comprehensive Proteoform Characterization of Plasma Complement Component C8αβγ by Hybrid Mass Spectrometry Approaches. J. Am. Soc. Mass Spectrom. *29*, 1099–1110. Available at: http://link.springer.com/10.1007/s13361-018-1901-6.
26. Vásquez-Suárez, A., Lobos-González, F., Cronshaw, A., Sepúlveda-Ugarte, J., Figueroa, M., Dagnino-Leone, J., Bunster, M., and Martínez-Oyanedel, J. (2018). The γ33 subunit of R-phycoerythrin from Gracilaria chilensis has a typical double linked phycourobilin similar to γ subunit. PLoS One *13*, e0195656. Available at: http://dx.plos.org/10.1371/journal.pone.0195656 [Accessed August 24, 2018].
27. Nair, D., Krishna, J.G., Panikkar, M.V.N., Nair, B.G., Pai, J.G., and Nair, S.S. (2018). Identification, purification, biochemical and mass spectrometric characterization of novel phycobiliproteins from a marine red alga, Centroceras clavulatum. Int. J. Biol. Macromol. *114*, 679–691.
28. Wohlschlager, T., Scheffler, K., Forstenlehner, I.C., Skala, W., Senn, S., Damoc, E., Holzmann, J., and Huber, C.G. (2018). Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. Nat. Commun. *9*, 1–9.
29. Franc, V., Yang, Y., and Heck, A.J.R. (2017). Proteoform profile mapping of the human serum Complement component C9 reveals unexpected new features of N- , O- and C-glycosylation. Anal. Chem. *89*, 3483–3491.
30. Yang, Y., Liu, F., Franc, V., Halim, L.A., Schellekens, H., and Heck, A.J.R. (2016). Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. Nat. Commun. *7*, 1–10. Available at: http://dx.doi.org/10.1038/ncomms13397.

31. Bhattacharya, D., Price, D.C., Chan, C.X., Qiu, H., Rose, N., Ball, S., Weber, A.P.M., Cecilia Arias, M., Henrissat, B., Coutinho, P.M., *et al.* (2013). Genome of the red alga Porphyridium purpureum. Nat. Commun. *4*, 1941. Available at: http://www.nature.com/doifinder/10.1038/ncomms2931.
32. Liu, L.N., Chen, X.L., Zhang, Y.Z., and Zhou, B.C. (2005). Characterization, structure and function of linker polypeptides in phycobilisomes of cyanobacteria and red algae: An overview. Biochim. Biophys. Acta - Bioenerg. *1708*, 133–142.
33. Apt, K.E., Hoffman, N.E., and Grossman, A.R. (1993). The γ subunit of R-phycoerythrin and its possible mode of transport into the plastid of Red algae. J Biol Chem. *268*, 16208–15.
34. Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.
35. Blanksby, S.J., and Ellison, G.B. (2003). Bond dissociation energies of organic molecules. Acc. Chem. Res. *36*, 255–263.
36. Mancini, J.A., Sheehan, M., Kodali, G., Chow, B.Y., Bryant, D.A., Dutton, P.L., and Moser, C.C. (2018). De novo synthetic biliprotein design, assembly and excitation energy transfer. J. R. Soc. Interface *15*.
37. Kelstrup, C.D., Bekker-Jensen, D.B., Arrey, T.N., Hogrebe, A., Harder, A., and Olsen, J. V. (2017). Performance evaluation of the Q Exactive HF-X for shotgun proteomics. J. Proteome Res. *17*, 727–738. Available at: http://pubs.acs.org/doi/abs/10.1021/acs.jproteome.7b00602.
38. Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., and Hochreiter, S. (2015). Msa: An R package for multiple sequence alignment. Bioinformatics *31*, 3997–3999.
39. Marty, M.T., Baldwin, A.J., Marklund, E.G., Hochberg, G.K.A., Benesch, J.L.P., and Robinson, C. V. (2015). Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. Anal. Chem. *87*, 4370–4376. Available at: http://pubs.acs.org/doi/10.1021/acs.analchem.5b00140 [Accessed August 24, 2018].
40. Murtagh, F., and Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J. Classif. *31*, 274–295.
41. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (Springer New York).
42. Vizcaino, J.A., Csordas, A., Griss, J., Lavidas, I., Mayer, G., Perez-riverol, Y., Reisinger, F., Ternent, T., Xu, Q., Wang, R., *et al.* (2016). 2016 update of the PRIDE database and its related tools. *44*, 447–456.