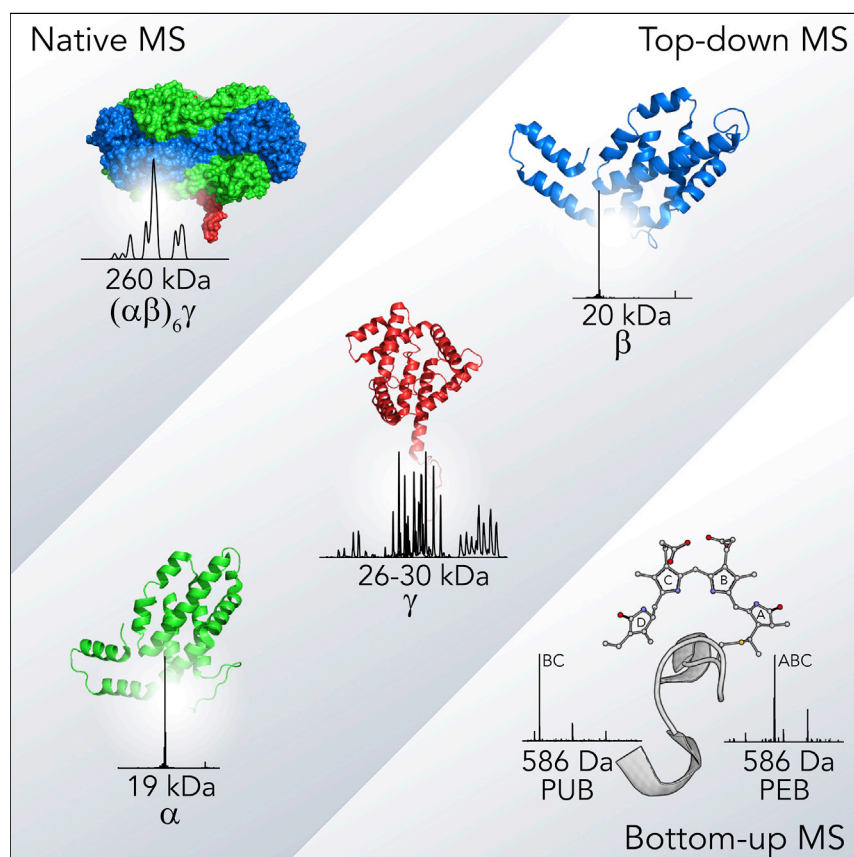


## Article

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



B-phycoerythrin is an extremely heterogeneous protein assembly comprising 13 subunits, all decorated with numerous chromophores. By combining different tiers of mass spectrometric analysis, native MS, top-down MS, and bottom-up MS, Tamara et al. characterize B-phycoerythrin in great detail. Integrating the data allowed them to identify all co-occurring proteoforms within B-phycoerythrin and identify the chemical nature and exact localization of all chromophores they harbor. Knowledge of these structural heterogeneities will be beneficial for the future design of even more efficient light-harvesting systems.

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

a.leney@bham.ac.uk (A.C.L.)  
a.j.r.heck@uu.nl (A.J.R.H.)

## HIGHLIGHTS

B-phycoerythrin assembly is structurally and chemically exceptionally heterogeneous

Complex heterogeneity unraveled by combining different tiers of mass spectrometry

The  $\gamma$  subunit is present in 4 distinct isoforms carrying 3–5 chromophores each

MS/MS allows unambiguous distinction of attached isobaric PEB and PUB chromophores

## Article

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

Sem Tamara,<sup>1,2</sup> Max Hoek,<sup>1,2</sup> Richard A. Scheltema,<sup>1,2</sup> Aneika C. Leney,<sup>3,\*</sup> and Albert J.R. Heck<sup>1,2,4,\*</sup>

## SUMMARY

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, which represents 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 ( $\alpha\beta$ )<sub>12</sub> arranged in a ring with the central cavity housing a linker ( $\gamma$ ) subunit, which is crucial for stabilizing B-phycoerythrin within the phycobilisome. Using top-down MS, we unravel the heterogeneity in the  $\gamma$  proteoforms, characterizing the distinct  $\gamma$  chains and multiple isobaric chromophores they harbor. Our data highlight the key role  $\gamma$  plays in phycobilisome organization that enables optimal light transmission.

## INTRODUCTION

Phycobilisomes are large light-harvesting antennas that facilitate the conversion of light into chemical energy in different species of cyanobacteria and red algae.<sup>1,2</sup> These MDa protein assemblies are formed by a morphologically distinct core complex and rod-like assemblies that are attached to the core.<sup>3</sup> Both structural units consist of stacked disc-shaped phycobiliproteins (PBPs), which are themselves multi-chain protein complexes with distinct photochemical properties. The core of the phycobilisome primarily comprises allophycocyanin (APC;  $\lambda_{\max}$  651 nm),<sup>4</sup> whereas the rods incorporate phycocyanin (PC;  $\lambda_{\max}$  620 nm)<sup>5</sup> and phycoerythrin (PE;  $\lambda_{\max}$  565 nm)<sup>6</sup> 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.<sup>7</sup> 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.<sup>8</sup>

The PE family of PBPs is 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.<sup>9</sup> As such, PEs have numerous biotechnological applications as dyes and fluorescent tags.<sup>10,11</sup> One of the most studied PEs is B-phycoerythrin (B-PE; Q = 0.98),<sup>12–15</sup> which is the most abundant PBP (~42% of all colorant proteins) in the red algae *Porphyridium cruentum*.<sup>16</sup> B-PE is known to be a hetero-13-mer that contains six  $\alpha$ , six  $\beta$ , and one linker-protein subunit termed

## The Bigger Picture

Some of the most efficient light-harvesting machineries present on earth are found in red algae and cyanobacteria. These systems, termed phycobilisomes, comprise numerous proteins decorated with a plethora of chromophores. The precise arrangement of all proteins and chromophores in the phycobilisome assembly form the basis for extremely efficient energy transfer. Here, we combine different mass spectrometric methods, enabling the structural investigation of all components of the B-phycoerythrin sub-complex in a highly detailed manner. This includes identifying all proteoforms present in the assembly, as well as distinguishing the various (isobaric) chromophores they harbor. Together, this information leads to fundamental insights into the arrangement and chemical heterogeneity of the phycobilisome. Better understanding of the architecture of this complex is essential for the future design of even more efficient light-harvesting machineries.

$\gamma$ .<sup>6</sup> 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  $\gamma$  subunit.<sup>17</sup> The B-PE complex has two types of bilin prosthetic groups that are covalently bound to cysteine residues: phycoerythrobilin (PEB;  $\lambda_{\max}$  550 nm) and phycourobilin (PUB;  $\lambda_{\max}$  498 nm).<sup>18</sup> 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  $\alpha$  chain carries 2 PEB prosthetic groups, whereas the 18.5 kDa  $\beta$  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  $\beta$  chain is connected through two thioester linkages.<sup>19</sup> Compared to the available knowledge about the  $\alpha$  and  $\beta$  chains, the nature of the  $\gamma$  chain and the chromophores it carries have so far remained much more elusive.

The  $\gamma$  chain is important because it stabilizes the tertiary structure of PEs by holding the discs of  $(\alpha\beta)_3$  hexamers together.<sup>15,17</sup> Moreover, the bilin prosthetic groups the  $\gamma$  chain harbors enhance the light-absorbance properties without increasing the spacing of PBPs.<sup>2</sup> Additionally, it has been proposed that the  $\gamma$  chain provides energetic decoupling protecting the photosynthetic reaction center from damage induced by excessive photoexcitation.<sup>20</sup> Initially, in biochemical studies of PEs, the  $\gamma$  subunit was identified as a single band on sodium dodecyl sulfate (SDS) gels and was assumed to be a single protein.<sup>6</sup> Later, reversed-phase liquid chromatography (RPLC) revealed that the  $\gamma$  subunit is represented by at least three distinct polypeptide chains in B-PE assemblies,<sup>16,21</sup> however, the exact amino acid sequences and positions of attached bilins were not determined. Overall, the  $\gamma$  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.<sup>6</sup> In the related R-PE complex, the  $\gamma$  subunit was long proposed to harbor 4 chromophores as well; however, five distinct chromophorylated peptides were detected, which was rationalized by the presence of several distinct  $\gamma$  subunits.<sup>8,22</sup> Recently, a structural model was reported for the entire phycobilisome from the red algae *Griffithsia pacifica* based on cryoelectron microscopy (cryo-EM) data revealing more details on the structure and conformation of the  $\gamma$ -linker subunit.<sup>2</sup> Predicted structures outlined the presence of a chromophore-binding domain on  $\gamma$  that could carry up to 5 bilin molecules. However, because of the variations of the  $\gamma$ -subunit sequences within single and, in particular, different algae strains, it is likely that not all  $\gamma$  subunits are identically modified. Moreover, because extensive “class” averaging was performed to obtain the cryo-EM images, heterogeneity in these  $\gamma$ -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  $\gamma$  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.<sup>23–25</sup> Because of their difference in mass, proteoforms and the modifications contained within them 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  $\gamma$  subunit in R-phycoerythrin.<sup>26,27</sup> The combination of multiple MS approaches provides complementary information that is not obtainable from a single MS method, which proved to be advantageous for the analysis of highly heterogeneous proteins and protein complexes.<sup>25,28–30</sup> 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

<sup>1</sup>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

<sup>2</sup>Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, the Netherlands

<sup>3</sup>School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

<sup>4</sup>Lead Contact

\*Correspondence: [a.leney@bham.ac.uk](mailto:a.leney@bham.ac.uk) (A.C.L.), [a.j.r.heck@uu.nl](mailto:a.j.r.heck@uu.nl) (A.J.R.H.)

<https://doi.org/10.1016/j.chempr.2019.03.006>

co-occurrence of multiple variants of the B-PE assembly and link each of these variants to distinct proteoforms of the  $\alpha$ ,  $\beta$ , and  $\gamma$  subunits. In our work, four distinct polypeptide chains of the  $\gamma$  subunit are identified (one more than previously reported for B-PE from *P. cruentum*<sup>13</sup>), 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*<sup>21</sup>). 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  $\gamma$  chains. Such information is relevant because, 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 PBPs in a phycobilisome are influenced by a linker protein and the number and/or type of carried bilin molecules,<sup>2</sup> different  $\gamma$  subunits define ordering of B-PE within phycobilisome rods.

## RESULTS

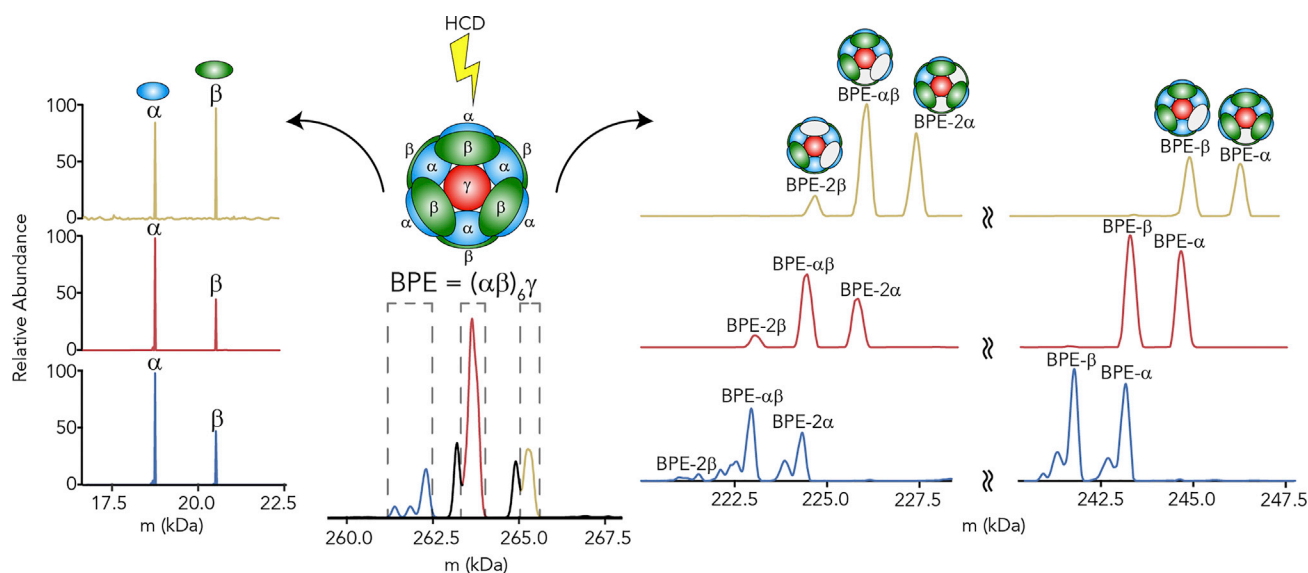
### Heterogeneity of B-Phycoerythrin Probed by Native MS

As a first step to unravelling the structural heterogeneity within B-PE, we measured the 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.<sup>15</sup> 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  $\alpha$  and  $\beta$  subunits and residual complexes of B-PE wherein  $\alpha$ ,  $\beta$ , 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  $\alpha$  and  $\beta$  subunits always had identical masses while the residual high molecular weight fragment complexes, formed by the loss of an  $\alpha$  and/or  $\beta$  subunit, exhibited clear mass differences that, thus, can be attributed to the presence of different forms of the  $\gamma$  subunit.

The native top-down dissociation of B-PE was found to be limited to the ejection of maximally two subunits. We did not detect the  $\gamma$  subunit to be ejected. This may have multiple reasons; namely its lower stoichiometry, broader structural heterogeneity, and less topological accessibility because the  $\gamma$  subunit is buried inside the cavity of the  $(\alpha\beta)_3$  hetero-hexamers. If  $\gamma$  is present in a large number of isoforms and proteoforms, its ion signals would spread over numerous peaks, hampering detection. In addition, we performed pseudo-MS3 experiments (Supplemental Experimental Procedures) to test whether it is possible to detect the  $\gamma$  subunit detaching from the residual B-PE complexes formed upon collisional activation. For this, following in-source activation, residual complexes corresponding to B-PE assembly lacking  $2\alpha$ ,  $\alpha\beta$ , or  $2\beta$  subunits were mass-selected and fragmented with higher-energy collisional dissociation (HCD) (Figure S1B). Alongside the  $\alpha$  and  $\beta$  monomeric products of dissociation in this experiment, we observed dimers and trimers of  $\alpha$  and/or  $\beta$  subunits; however, even then, we did not detect intact  $\gamma$  subunits (Figures S1C and S1D).

### Characteristics of Proteins Comprising the B-Phycoerythrin Assemblies

To identify the heterogeneity behind the  $\gamma$  subunit as revealed by native MS (here and earlier by Leney et al.<sup>15</sup>), we next denatured B-PE, digested it into peptides,

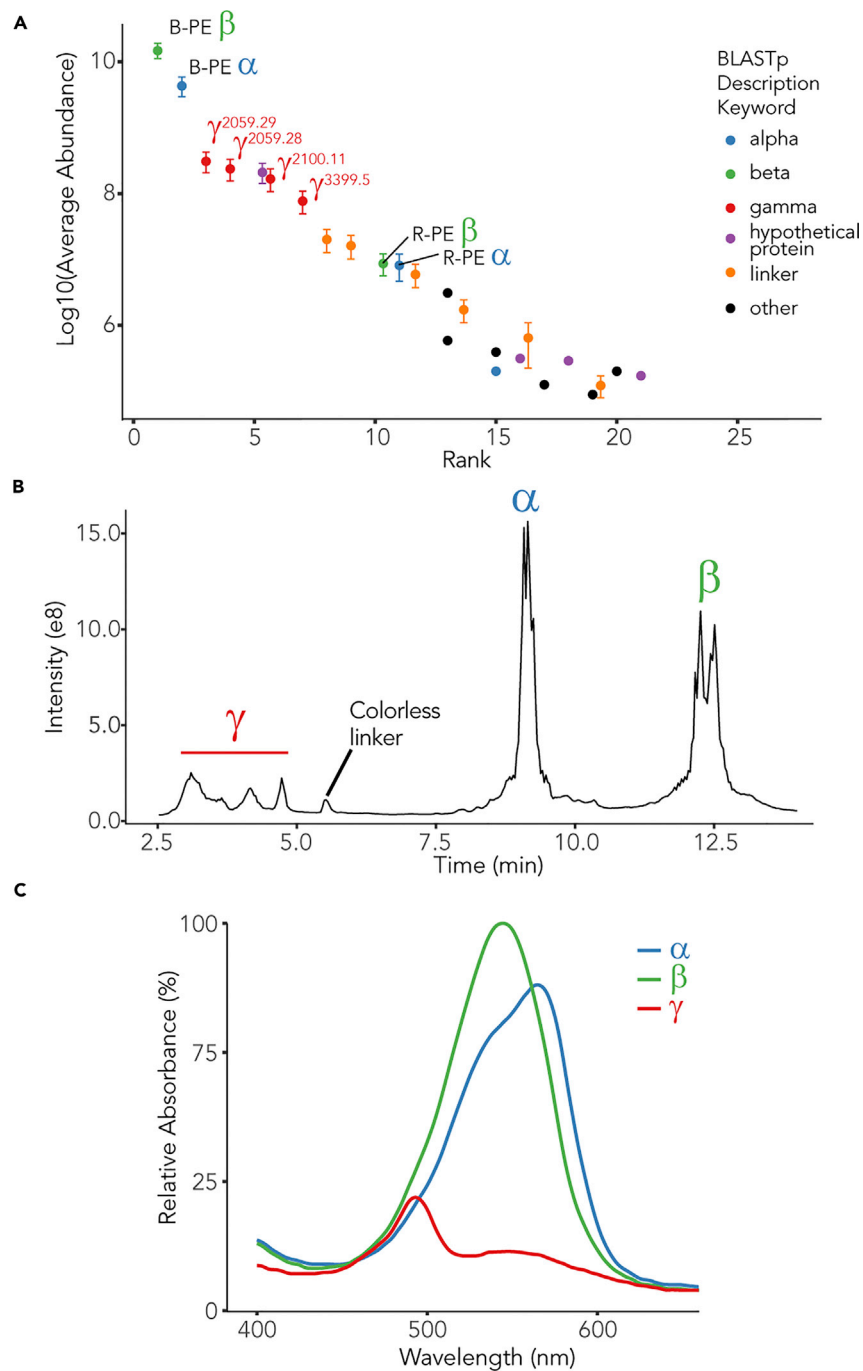


**Figure 1. Native Top-Down MS/MS of B-Phycocerythrin Assembly Variants**

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  $\alpha$  and  $\beta$  monomeric subunits (left) and the concomitant formation of residual fragmented complexes (right) missing a single  $\alpha$  or  $\beta$  subunit or combinations of subunits ( $2\alpha$ ,  $2\beta$ , or  $\alpha\beta$ ).

and analyzed the resulting peptides by bottom-up liquid chromatography-tandem mass spectrometry (LC-MS/MS). The data revealed that the most abundant proteins in the sample were as expected the  $\alpha$  and  $\beta$  chains of B-PE. Identifying any  $\gamma$  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.<sup>31</sup> By comparing the peptides identified with this dedicated protein database, the presence of 4 different  $\gamma$  subunits was revealed in the list of the most abundant identified proteins (Figure 2A; Table S1). Next to these, we also detected peptides originating from other linker proteins 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.<sup>32</sup> 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  $\gamma$  subunits within B-PE, we next denatured B-PE and separated the intact proteins using reversed-phase high-pressure liquid chromatography (HPLC). Consistent with the bottom-up results, the data showed two abundant signals corresponding to the  $\alpha$  (~9-min retention time) and  $\beta$  subunits (12.5-min retention time), as well as several lower abundant peaks with shorter retention times (2.5–5 min) (Figure 2B). Peak splitting observed for the  $\beta$  subunit was attributed to the shifted retention time of oxidized proteoforms. Additionally, we observed a peak at a retention time of 5.5 min that was assigned to the colorless linker protein (Figure S2; fraction A06). The shorter retention times of the  $\gamma$  subunits can be rationalized by these proteins harboring more hydrophilic residues than the  $\alpha$  and  $\beta$  subunits. To verify whether the eluting proteins are chromophorylated subunits of B-PE, we measured the absorbance spectra of the fractions (Figure 2C).



**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 the respective peptides in LC-MS. The error bars 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 RPLC, corresponding to the  $\alpha$  (blue),  $\beta$  (green), and four  $\gamma$  (red) subunits.

Clearly, three distinct absorption profiles were observed that resembled previously reported absorbance spectra of the B-PE subunits.<sup>13</sup> Additionally, we measured the absorbance for fractions corresponding to different  $\gamma$  chains (Figure S2; fractions A02–A04), whereby we observed that all these fractions resulted in similar 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  $\gamma$  subunits.

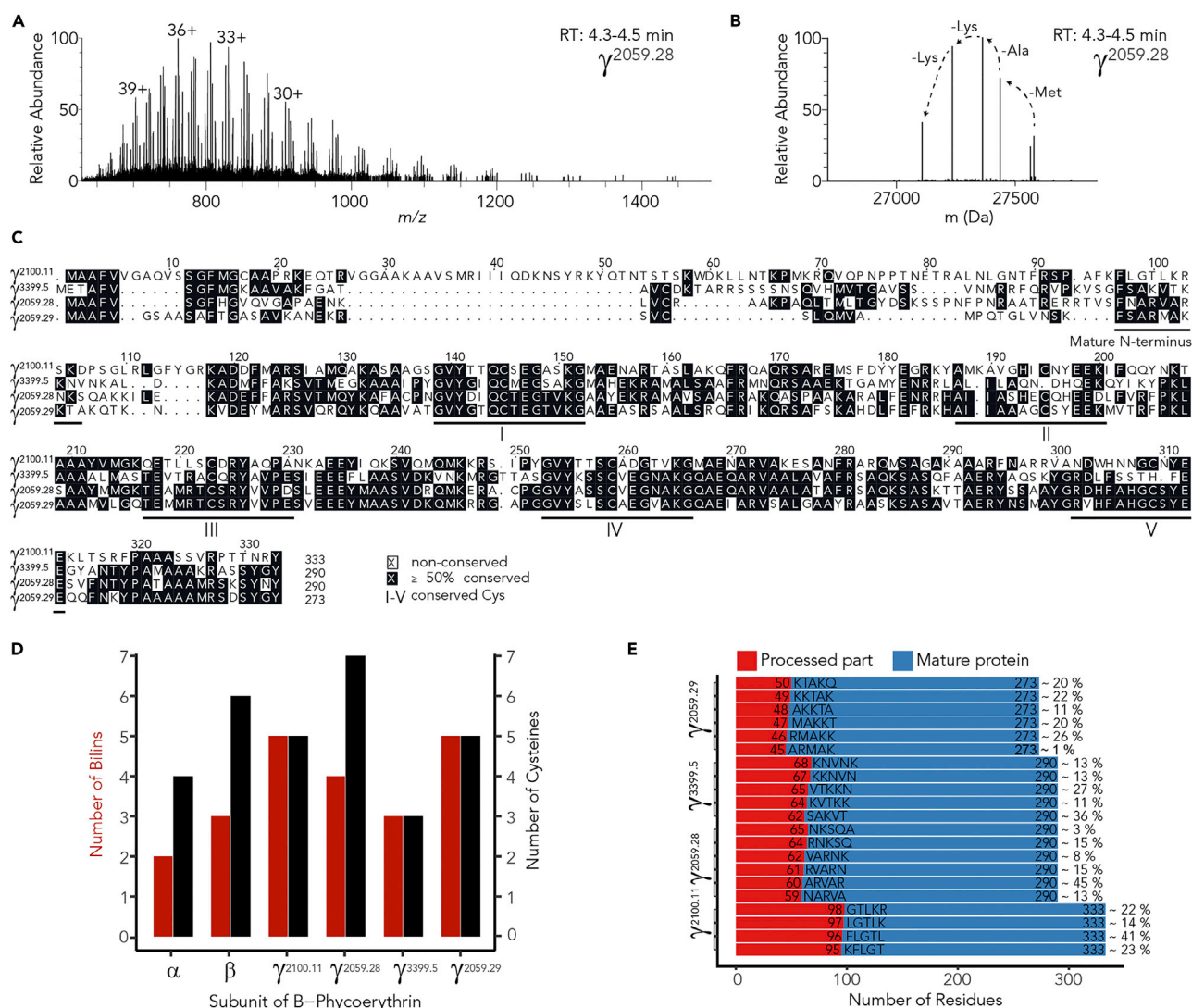
### Characterizing $\gamma$ Subunit with “Sequence Tags” and Mass Matching

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  $\gamma$  proteoforms studied here harbor various bilin modifications. These chromophores heavily influence and complicate the observed dissociation patterns (Figure S3A). In addition, co-isolation of co-eluting proteoforms further limit the straightforward retrieval of sequence information. Therefore, we here characterized the  $\gamma$  proteins using an alternative approach, wherein the proteoforms were identified by mass matching to the theoretical proteoform masses that were further verified by using bottom-up LC-MS/MS.

Using fast, low resolution (7,500 at 200  $m/z$ ) recording of mass spectra, several  $\gamma$  proteoforms were successfully resolved by top-down LC-MS (Figure S3B; Data S1), consistent with the broad elution peaks observed in the total ion chromatogram corresponding to the  $\gamma$  subunits (Figure 2B). The intact masses alone provide significant insight into the heterogeneity present within the  $\gamma$  subunit with multiple molecular weights being identified for each proteoform (Figure S3C). Elucidating their sequences from intact masses only, however, is challenging because of two reasons. First, the protein sequences of B-PE  $\gamma$  subunits from *P. cruentum* are missing from the conventionally used protein databases (e.g., UniProt). Second, information on the bilin content and localization in the  $\gamma$  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 numbers 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  $\gamma$  proteoforms identified, post-translational sequence processing events could have occurred.

While MS scans of the  $\alpha$  and  $\beta$  subunits were dominated by a single most abundant proteoform (Figure S4), for the  $\gamma$  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 (Figures 3A and 3B). Excitingly, by searching these “sequence tags” against the custom database of the  $\gamma$  subunit proteoforms, we identified one of the  $\gamma$  subunits, which we named  $\gamma^{2059.29}$  wherein the superscript refers to the contig number annotated in the DNA sequencing.<sup>31</sup> The observed position of the “sequence tag” indicated an N-terminal processing of this chain, which previously was proposed for  $\gamma$  subunits of R-PE, based on the fact that they require a transit peptide for transfer into the chloroplasts.<sup>33</sup> Interestingly, upon further



**Figure 3. Determination of the Pallet of  $\gamma$  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) Deconvoluted 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  $\gamma$  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  $\gamma$  subunit.

analysis, top-down LC-MS revealed truncated sequence variants for all the  $\gamma$  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  $\gamma$  subunits were detected by bottom-up MS (data not shown).

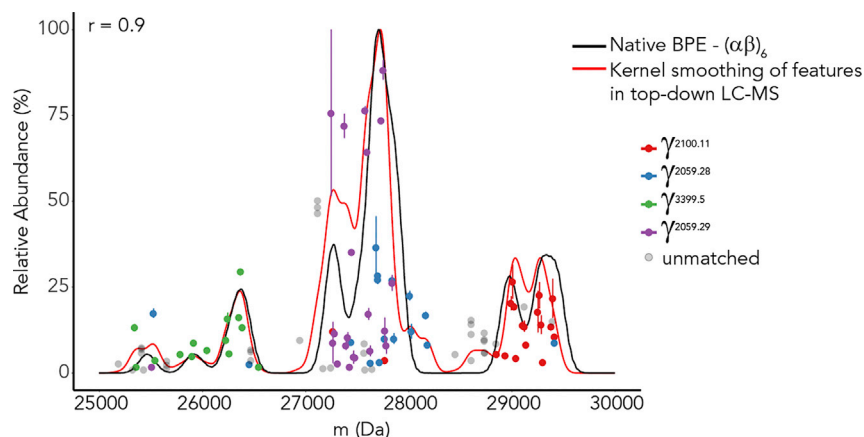
To investigate  $\gamma$  protein processing and chromophorylation sites, the sequences of the  $\gamma$  subunits were aligned, making use of the Multiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm.<sup>34</sup> This alignment revealed that all the  $\gamma$

subunits contain a conserved region, which has been recently pointed out as the chromophore-binding domain of the  $\gamma$  subunits.<sup>2</sup> In three out of four  $\gamma$  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 a total of 5 conserved cysteines) (Figure 3C). We extended our custom database of theoretical proteoforms with truncated forms of  $\gamma$ , 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 numbers 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 MS can provide by 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  $\gamma$  proteoforms, with the data revealing both the extent of protein processing and the variable bilin occupancies for each  $\gamma$  subunit (Figures 3D and 3E). Two out of four  $\gamma$  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  $\alpha$  and  $\beta$  subunits predominantly carry 2 and 3 bilins per subunit, respectively, resulting in half of the total number of cysteines being occupied (Figure 3D). The heterogeneity observed for the  $\gamma$  subunit in our work for B-PE from *P. cruentum* is in agreement with the cryo-EM data recently reported for the phycobilisome from *G. pacifica*, which has R-PE as a primary type of PE. For *G. pacifica*, 5 distinct isoforms of the  $\gamma$  subunit were detected, four of which harbored 5 chromophores, and one had 4 bilin molecules attached.<sup>2</sup> The complete list of matched proteoforms of  $\alpha$ ,  $\beta$ , and the various  $\gamma$  subunits of B-PE from *P. cruentum* can be found in Data S1. Additionally, to verify the most prominent  $\gamma$  proteoform within the B-PE assembly, we targeted the RPLC fraction corresponding to the  $\gamma^{2059.29}$  isoform and performed electron-transfer/higher-energy collision dissociation (ETHcD) MS/MS. All 5 predicted bilins with isobaric masses of 586.279 Da were detected by corresponding fragments and could all be positioned at the conserved chromophorylation sites (Figure S5; Supplemental Experimental Procedures).

### Reconstruction of the Native Mass Spectrum of B-PE from the Qualitative and Quantitative Data on All $\alpha$ , $\beta$ , and $\gamma$ 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<sup>30</sup> 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  $\gamma$  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  $\gamma$  subunits that participate in formation of different B-PE variants have been explicitly and correctly identified by our top-down MS approach. Additionally, it confirms that the  $\gamma$  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 (Figures 4 and S6).

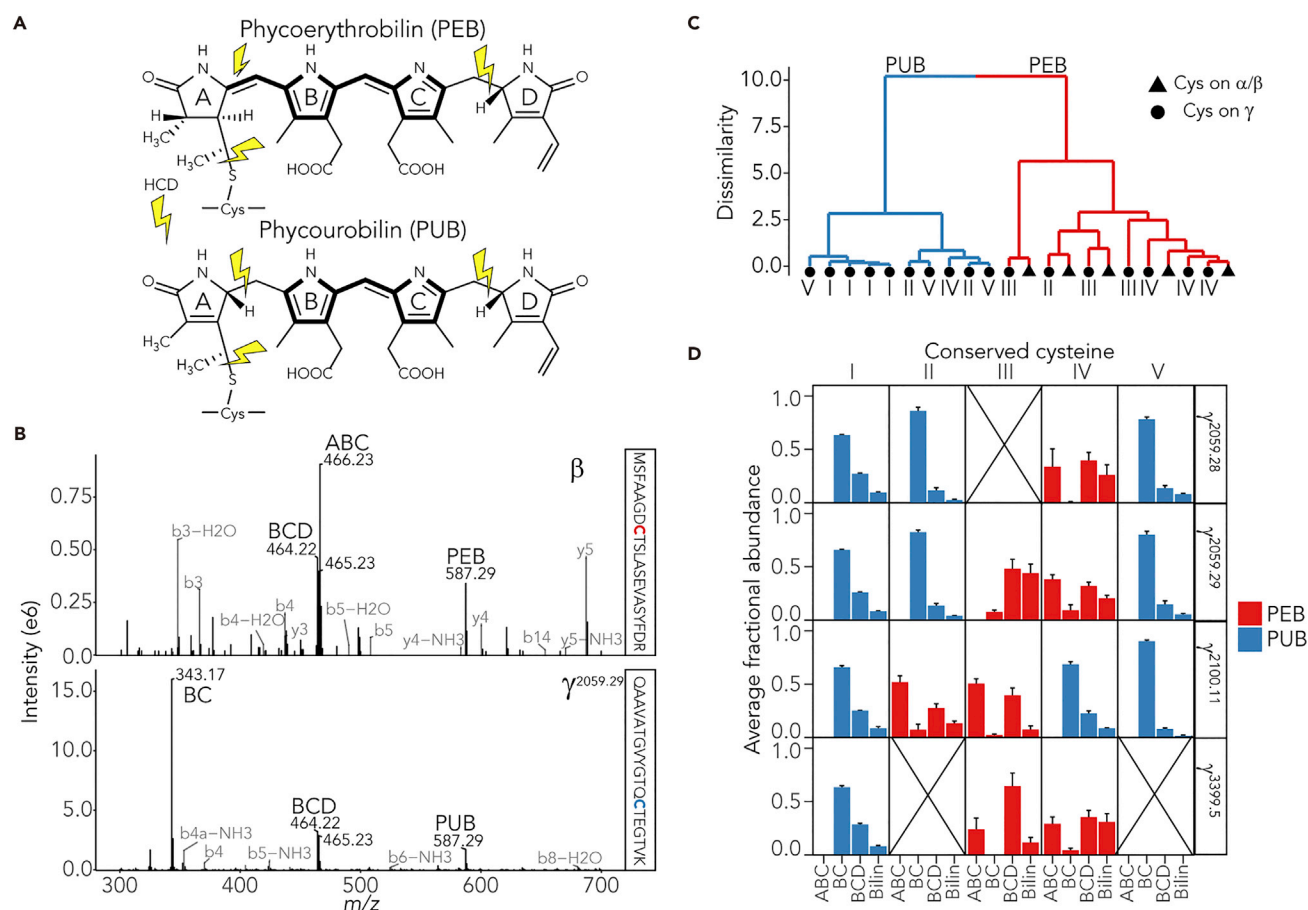


**Figure 4. Reconstruction of the Features Observed in the Native Mass Spectrum of the Intact B-Phycoerythrin Assembly Based on the Proteoforms Detected in the Top-Down LC-MS Analysis**

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  $\gamma$  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 (Figures 2C and S7) indicate the presence of two types of chromophores, namely, PEB (absorbance maximum at 550 nm) and PUB (absorbance maximum at 498 nm). In the PubChem database, the molar mass of PEB (CID 5289229) is 586.689 g/mol, whereas 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 a 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 a 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 considering that the chromophores have different  $\pi$  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),<sup>35</sup> we hypothesized that upon MS/MS, the extended  $\pi$  conjugation system of PEB should prevent the formation of fragments containing two inner pyrroles (annotated as BC in Figures 5A and 5B), 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<sup>+</sup> and 464.22<sup>+</sup>) or two-pyrrole ( $m/z$  343.17<sup>+</sup>) 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<sup>+</sup> annotated as ABC in Figures 5A and 5B). This fragmentation pattern agrees with the proposed conjugation system in the PEB molecule. Notably, the slightly higher abundance of the tri-pyrrole



**Figure 5. Differentiating between Isobaric Phycoerythrobilin and Phycourobilin 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.

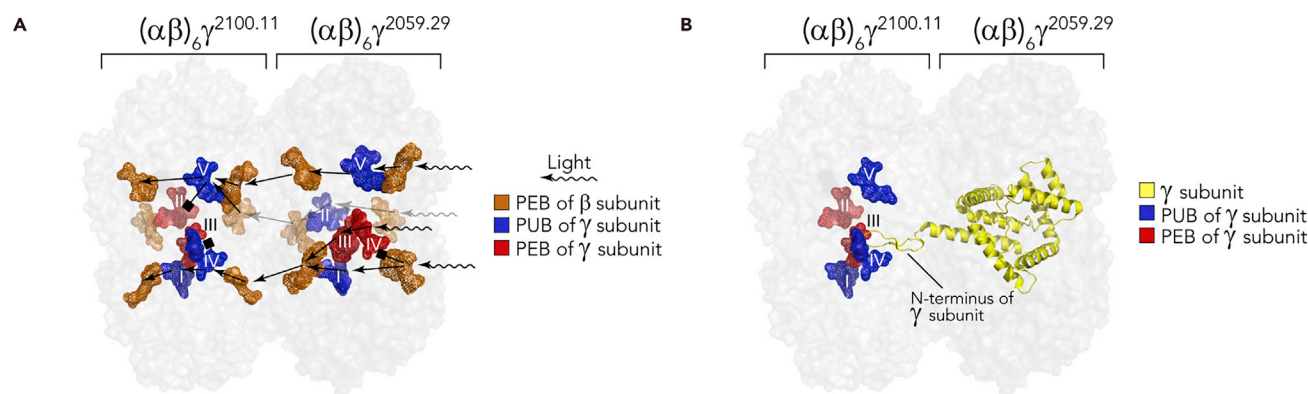
(B) Examples of tandem mass spectra of peptides chromophorylated with PEB (β subunit) or PUB (γ<sup>2059.29</sup> subunit).

(C) Dendrogram representing the hierarchical clustering of the chromophorylation sites based on the abundances of four characteristic fragments (ABC, BC, BCD, and Bilin). Each dot or triangle represents the chromophorylation site on the 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.

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 the dissociation energies of the respective bonds (C-S and C-C single bonds for ABC; C=C double bond for BCD).<sup>35</sup>

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 and 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 the expected presence of two bilin types suggested by the absorbance profiles of B-PE subunits (Figure 2B). Distinctively, doubly linked PEB on β subunit (Cys50-Cys61) produced fewer ABC fragments 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.



**Figure 6.  $\gamma$  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: 5Y6P, phycobilisome from *G. pacifica*). The bilin chromophores on the  $\gamma$  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 gray transparent surface. The black lines schematically represent the shortest energy transfer pathways through the phycobilisome complex.

(B) N terminus of the  $\gamma$  subunit from one phycoerythrin lies within the complex in close proximity to bilin groups on conserved cysteines III and IV (Figure 3C) of the  $\gamma$  subunit from a neighboring phycoerythrin complex.

In the  $\gamma$  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  $\gamma$  subunits connect with those of the  $\beta$  subunit to allow efficient energy transfer within rods of the phycobilisome.<sup>2</sup> Using the recently published structural model of the phycobilisome from *G. pacifica*, we color-coded in Figure 6, bilins of the  $\gamma$  and  $\beta$  subunits of PEs 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  $\gamma$  subunit within B-PE in *P. cruentum* also connect with the PEB molecules on the  $\beta$  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  $\gamma$  subunits are participating in the shortest energy transfer pathways indicated by the black arrow lines in Figure 6A. In a phycobilisome, such a layout of chromophores in the PE, 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 transfers within PEs in the proximal parts of the rods of a phycobilisome. To a lesser extent, PUB groups of the  $\gamma$  subunits might act as light quenchers by absorbing less energetic emission from excited PEB molecules. The overall structure of stacked PE complexes reveals that the  $\gamma$  subunits from one PE complex link to another via its N terminus, being in close proximity to bilin groups on conserved cysteines III and IV of the  $\gamma$  subunit from subsequent PE (Figure 6B). Diverse bilin combinations at these positions (Figure 5D) may regulate the ordering of B-PE complexes in the rods of the phycobilisome.

## DISCUSSION

The distinct photochemical properties of each PBP 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 because of 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 PBP 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. Finally, 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 a mixture composed of six  $\alpha$ , six  $\beta$ , and one of four distinct  $\gamma$  subunits:  $\gamma^{2059.29}$ ,  $\gamma^{2059.28}$ ,  $\gamma^{3399.5}$ , and  $\gamma^{2100.11}$  whereby an average of 35 bilin molecules decorate 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  $\gamma$  subunit; the subunit that links B-PEs together and is essential for its stability. Our results indicate that 4 distinct  $\gamma$  subunits are present, which is in agreement with the cryo-EM structure whereby  $\gamma$  subunits are required to link the individual PE complexes in the rods of the phycobilisome.<sup>2</sup> Furthermore, the different N-terminal regions of the distinct  $\gamma$  subunits that participate in linker-linker contacts might influence the absorbance and emission properties of the involved prosthetic groups because the protein microenvironment influences the fluorophore-mediated light transmission.<sup>36</sup> Additionally, we show the most abundant  $\gamma$  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  $\gamma$  subunits because  $\gamma^{2059.28}$  and  $\gamma^{3399.5}$  carry 4 and 3 bilins, respectively. The  $\gamma$  subunit is the only PBP within PE to contain 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  $\gamma$  subunits wherein two chromophore positions that participate in linking of PEs demonstrate unique chromophorylation patterns. Thus, we speculate that varying chromophorylation patterns and distinct primary structures of linker subunits drive joining and ordering of PEs for efficient light transmission throughout the rods of phycobilisomes. Overall, we expect that the detailed molecular knowledge gathered here will provide a 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-PE was reduced with 20 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) at room temperature for 30 min and alkylated with 20 mM chloroacetamide for 30 min in the dark. Digestion of proteins was performed 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 a pH > 7 prior to LC-MS/MS in order to prevent interconversion between PUB and PEB.

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  $\mu\text{g}/\mu\text{L}$  final concentration.

B-PE sample was prepared for native MS experiments through 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 of 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  $\mu\text{m} \times 20$  mm trap column (in-house packed with ReproSil-Pur C18-AQ, 3  $\mu\text{m}$ ) (Dr. Maisch, GmbH, Ammerbuch-Entringen, Germany) coupled to a 50  $\mu\text{m} \times 500$  mm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7  $\mu\text{m}$ ) (Agilent Technologies, Amstelveen, the Netherlands). A 2–5  $\mu\text{L}$  injection of peptides was used, corresponding to  $\sim 0.05$   $\mu\text{g}$  of material. The LC-MS/MS run time was set to 40 min with a flow rate of 300 nL/min. Mobile phases A (0.1% formic acid in water) and B (0.1% formic acid in 80% ACN) were used for a gradient time of 35 min: 13%–44% B for 20 min, and 44%–100% B over 3 min. 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 the resolution setting set to 60,000 (200  $m/z$ ) and at a scan mass range of 375–2,000  $m/z$ . 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 200–2,000  $m/z$ . AGC target was set to 5e4 with a maximum injection time defined at 54 ms. 1  $\mu\text{scan}$  was acquired both for full MS and dd-MS/MS scans. The data-dependent method was set to isolation and fragmentation for the cycle time set to 5 s. Parameters for isolation and 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  $\times$  50 mm column. 2  $\mu\text{g}$  material was loaded on the column heated to 80°C. LC-MS runtime was set to 22 min with a flow rate of 250  $\mu\text{L}/\text{min}$ . Gradient elution was performed using mobile phases A (0.1% formic acid in water) and B (0.1% formic acid in ACN): 25%–46% B for 14 min.

All top-down MS experiments were performed on a Thermo Scientific Q Exactive HF-X instrument (Thermo Fisher Scientific, Bremen, Germany).<sup>37</sup> LC-MS data were collected with the 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 150–2,000  $m/z$  with the AGC target set to 3e6. The maximum of injection time was defined at 16 ms with 1  $\mu\text{scan}$  recorded.

#### Absorbance Measurements

B-PE subunits separated with RPLC were collected at the time of elution following loading of 25  $\mu\text{g}$  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  $\mu$ L corresponding to each subunit were loaded into a 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 a concentration of  $\sim 2 \mu$ M 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. The sample was sprayed at a 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. The resolution setting was 8,750 (at 200  $m/z$ ), and the 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  $\sim 100$  microscans in the time domain. Pseudo-MS3 analysis of B-PE complexes that had already ejected one or more subunits, are described in the [Supplemental Experimental Procedures](#).

#### Identification of B-PE Proteoforms

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

#### Data Analysis

Raw bottom-up LC-MS/MS data were analyzed with Proteome Discoverer 2.2 (Thermo Fisher Scientific) equipped with Byonic nodes (Protein Metrics, Cupertino, USA). The following parameters were used for the database search. Protease: Trypsin (full). Variable modifications: Met oxidation; Cys carbamidomethyl; Cys chromophorylations of 586.279 and 590.31 Da. A protein sequence database was generated based on the recently published genome of *P. purpureum* (*P. cruentum*).<sup>31</sup> Sequence alignment was performed in R with the use of an "msa" package,<sup>38</sup> and the MUSCLE algorithm.<sup>34</sup> Top-down LC-MS raw files were deconvoluted with the Sliding Window ReSpect algorithm available in the Protein Deconvolution 4.0 software package (Thermo Fisher Scientific). Zero charged mass distribution profiles were obtained

from raw native mass spectra with UniDec.<sup>39</sup> Structural visualization of PE 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 an algorithm that implements Ward's criterion.<sup>40</sup> Data were visualized with ggplot2 package.<sup>41</sup>

## DATA AND SOFTWARE AVAILABILITY

The data reported in this paper have been deposited to the ProteomeXchange Consortium via the PRIDE<sup>42</sup> partner repository. The accession number for the data is ProteomeXchange: PXD011275. The native MS data relevant to the study are available upon request.

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.chempr.2019.03.006>.

## 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 A.J.R.H. 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., and 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.

Received: December 20, 2018

Revised: February 13, 2019

Accepted: March 12, 2019

Published: May 9, 2019

## 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, 725–749.
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.
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. Bio.Chem.* 89, 79–86.
5. Glazer, A.N., Fang, S., and Brown, D.M. (1973). Spectroscopic properties of C-phycoerythrin 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. Porphyrinoid cruentum B-phycoerythrin and b-phycoerythrin. *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  $\alpha$ ,  $\beta$ , and  $\gamma$  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., Ruiz-Ruiz, F., Aguilar-Yáñez, J.M., Benavides, J., and Rito-Palomares, M. (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.* 1, 63–68.
  12. Tang, Z., Jilu 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.
  13. Bermejo, R., Talavera, E.M., and Alvarez-Pez, J.M. (2001). Chromatographic purification and characterization of B-phycoerythrin from *Porphyridium cruentum*. *J. Chromatogr. A* 917, 135–145.
  14. Munier, M., Jubeau, S., Wijaya, A., Morancais, 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.
  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.
  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  $\gamma$  subunit. *Eur. J. Biochem.* 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., DeLange, R.J., and Brown, D.M. (1984). Bilin attachment sites in the  $\alpha$  and  $\beta$  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., Do Vale, L.H.F., Seckler, H.S., Doubleday, P.F., Schachner, L.F., Szentić, 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. USA* 115, 8763–8768.
  25. Franc, V., Zhu, J., and Heck, A.J.R. (2018). Comprehensive proteoform characterization of plasma complement component C8 $\alpha\beta\gamma$  by hybrid mass spectrometry approaches. *J. Am. Soc. Mass Spectrom.* 29, 1099–1110.
  26. Vázquez-Suárez, A., Lobos-González, F., Cronshaw, A., Sepúlveda-Ugarte, J., Figueroa, M., Dagnino-Leone, J., Bunster, M., and Martínez-Oyanedel, J. (2018). The  $\gamma$ 33 subunit of R-phycoerythrin from *Gracilaria chilensis* has a typical double linked phycourobilin similar to  $\gamma$  subunit. *PLoS One* 13, e0195656.
  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. Wohlschlagel, 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, 1713.
  29. Franc, V., Yang, Y., and Heck, A.J.R. (2017). Proteoform profile mapping of the human serum complement component C9 revealing 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, 13397.
  31. Bhattacharya, D., Price, D.C., Chan, C.X., Qiu, H., Rose, N., Ball, S., Weber, A.P.M., Arias, M.C., Henrissat, B., Coutinho, P.M., et al. (2013). Genome of the red alga *Porphyridium purpureum*. *Nat. Commun.* 4, 1941.
  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* 1708, 133–142.
  33. Apt, K.E., Hoffman, N.E., and Grossman, A.R. (1993). The  $\gamma$  subunit of R-phycoerythrin and its possible mode of transport into the plastid of red algae. *J. Biol. Chem.* 268, 16208–16215.
  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. (2018). Performance evaluation of the Q Exactive HF-X for shotgun proteomics. *J. Proteome Res.* 17, 727–738.
  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.
  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).
  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. *Nucleic Acids Res.* 44, 447–456.