



# Polysaccharide associated protein (PSAP) from the green microalga *Botryococcus braunii* is a unique extracellular matrix hydroxyproline-rich glycoprotein

Mehmet Tatli<sup>a</sup>, Mayumi Ishihara<sup>b</sup>, Christian Heiss<sup>b</sup>, Daniel R. Browne<sup>a</sup>, Lawrence J. Dangott<sup>a,c</sup>, Stanislav Vitha<sup>d</sup>, Parastoo Azadi<sup>b</sup>, Timothy P. Devarenne<sup>a,\*</sup>

<sup>a</sup> Department of Biochemistry & Biophysics, Texas A&M University, College Station, TX 77843, United States

<sup>b</sup> Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, United States

<sup>c</sup> Protein Chemistry Lab, Department of Biochemistry & Biophysics, Texas A&M University, College Station, TX 77843, United States

<sup>d</sup> Microscopy & Imaging Center, Texas A&M University, College Station, TX 77843, United States



## ARTICLE INFO

### Keywords:

*Botryococcus braunii*  
Extracellular matrix proteins  
Hydroxyproline-rich glycoprotein  
Polysaccharide associated protein  
Protein glycosylation

## ABSTRACT

The green colonial microalga *Botryococcus braunii* produces large amounts of liquid hydrocarbons that can be converted into transportation fuels. Colony cells are held together by a complex extracellular matrix (ECM) made up of a cross-linked long-chain hydrocarbon network around which liquid hydrocarbons are stored, a retaining wall for holding hydrocarbons within the cross-linked hydrocarbon network, and a polysaccharide fibrillar sheath radiating from the retaining wall and surrounding the entire colony. Analysis of “shells” shed from cell apical regions during cell division and containing the retaining wall and polysaccharide fibers shows association of a single protein where the fibers meet the retaining wall, suggesting involvement of this protein in polysaccharide fiber formation. Here we use peptide mass fingerprinting and bioinformatics to identify this protein called polysaccharide associated protein (PSAP). PSAP does not show similarity to any protein in databases, but contains several Proline-rich domains. Staining studies confirm PSAP as a glycoprotein, and mass spectrometry analysis identified ten N-linked glycosylation sites comprising seven different glycans containing mainly mannose and N-acetylglucosamine. Three of these glycans also contain fucose, with one of these glycans being unusual since it also contains arabinose. Additionally, four hydroxyproline residues have short O-linked glycans of mainly arabinose and galactose, with one also containing a 6-deoxyhexose. PSAP secretion and localization to shell material is confirmed using western blot analysis and microscopy. These studies indicate PSAP contains unique glycans and suggest its involvement in ECM polysaccharide fiber biosynthesis.

## 1. Introduction

An extracellular matrix (ECM) is a common feature in multicellular eukaryotic organisms and offers a range of functions from structural support to cell-to-cell communication [1,2]. The ECM in algae is used to hold cells into a multicellular organization that can contain differentiated cell types or all non-differentiated cells [3,4]. This multicellularity due to ECM development in green algae is thought to have been the first step in the evolution of land plants [5,6]. For colonial green microalgae, the ECM is integral to cell survival by offering a scaffold for holding cells into a colony, helping to regulate cell development and differentiation, and acting as a reservoir for molecules secreted from the cells [7–10].

In terms of colony formation, some green microalgae form

interconnected chains of cells through an ECM consisting mainly of the cell wall, while in others separate cells are implanted into an ECM made up of the cell wall plus a complex structure extending beyond the cell wall [4,8,11,12]. In typical descriptions, the ECM consists of all material outside the cell membrane, including the cell wall [4]. For simplicity, our discussions of the ECM in this study will refer only to material exterior to the cell wall.

The ECM in green microalgae can consist of many molecules ranging from glycoproteins to hydrocarbons [4,8,11,12], however, the most common component is glycoproteins that are often cross-linked with each other [4]. One of the best studied green microalgal ECMs is that of *Volvox carteri*, which consists solely of glycoproteins, can hold over 2000 cells differentiated mainly into somatic cells with a few reproductive gonidia cells, and is highly enriched in 4-hydroxyproline

\* Corresponding author at: Department of Biochemistry & Biophysics, 2128 TAMU, Texas A&M University, College Station, TX 77843-2128, United States.  
E-mail address: [tpd8@tamu.edu](mailto:tpd8@tamu.edu) (T.P. Devarenne).

(Hyp) rich glycoproteins (HRGPs) that are cross-linked for stabilization [4,9,10,13]. The use of HRGPs is a common feature in the ECMs and cell walls of both algae and land plants, and they are commonly O-glycosylated at Ser, Thr, and Hyp residues with arabinose and galactose [14–17].

The green colony-forming microalga *Botryococcus braunii* is well known as a producer of hydrocarbon oils that can be converted into combustion engine fuels [8,18]. The three chemical races of *B. braunii* known as A, B, and L are differentiated by the type of hydrocarbons produced; odd carbon number alkenes in the A race [19–22], triterpenoid botryococenes and methylsqualenes in the B race [23–25], and the tetraterpenoid lycopadiene in the L race [26–28]. In all three races an ECM holds approximately 100–200 undifferentiated cells into a colony, and the hydrocarbons are biosynthesized inside the cells for secretion and storage in the ECM [8,18].

The *B. braunii* ECM shows some striking differences from the ECMs of other green microalgae such as *V. carteri*. For example, the main component of the *B. braunii* ECM is an intricate network of long chain polyacetal hydrocarbons that are cross-linked through epoxide bridges to monomers of the liquid hydrocarbons in each race [29–31]. Additionally, the polyacetal hydrocarbons can be threaded through large macrocyclic carbon rings [32]. These two types of linkages offer both a strong chemical bond (epoxide bridges) for structural integrity and mechanical linkages (carbon rings) for extensibility [8,29,32]. The spaces within this cross-linked hydrocarbon network are filled with the liquid hydrocarbon oils produced by each race of *B. braunii* [8,18].

In the B race of *B. braunii* the ECM also contains a retaining wall located near the outer edge of the ECM that encompasses the cross-linked hydrocarbon network [12]. As the name implies, the retaining wall holds the liquid hydrocarbons around the cells and within the spaces created by the cross-linked hydrocarbon network [12]. Radiating outward from the retaining wall is a network of 2–3  $\mu\text{m}$  long polysaccharide fibers that completely circumscribe the colony [12,33–35]. These fibers consist mainly of galactose and arabinose with uncommonly abundant (1  $\rightarrow$  2), (1  $\rightarrow$  3) and (1  $\rightarrow$  2), (1  $\rightarrow$  4) branching connections between the sugars, possibly to limit degradation by bacteria cohabitating the environment [12].

Protein granules are localized to the base of these polysaccharide fibers where the fibers attach to the retaining wall, and it has been suggested the protein in these granules may be involved in fiber biosynthesis [12]. During *B. braunii* cell division, portions of the ECM covering the apex of each cell and containing the retaining wall, polysaccharide fibers, and protein granules are broken off from the ECM and shed into the culture media [12,33,36]. These fragments have the appearance of cup-shaped “shells”, are easily collected from the media, and analysis indicates the associated protein granules consist of a single protein species [12].

Here we identify this protein, term it polysaccharide associated protein (PSAP), confirm its localization to the ECM by association with shell material, characterize it as an HRGP, identify the PSAP glycosylations, and propose possible PSAP functions in *B. braunii* race B ECM polysaccharide fibril formation.

## 2. Materials and methods

### 2.1. Identifying the PSAP full length cDNA

A partial PSAP sequence was initially identified by peptide mass fingerprinting using the following approach. PSAP protein was extracted from 3.0 mg of lyophilized shell material by adding 200  $\mu\text{l}$  of extraction buffer (1 mM NaCl, 10 mM Tris, pH 6.8, 1  $\times$  protease inhibitors) and 200  $\mu\text{l}$  of 2  $\times$  SDS-PAGE sample buffer to the shell material, and placing the sample at 95  $^{\circ}\text{C}$  for 5 min. A 4–15% gradient SDS-PAGE gel was used to separate 20  $\mu\text{l}$  of the extract, the protein bands visualized using Coomassie blue, the PSAP band excised from the gel, and the samples in-gel digested with trypsin overnight. The resulting

peptides were separated and analyzed for mass and sequence identity by liquid chromatography-mass spectrometry (LC-MS/MS) at the University of Texas at San Antonio Health Science Center Institutional Mass Spectrometry Core Laboratory. The analysis for peptide sequence identity was done with MASCOT and X!Tandem software, and the results were summarized and validated using Scaffold v4 (Proteome Science). The generated sequences were searched against locally entered sequences concatenated to the Swiss-Prot database.

The identified peptides were queried against the *B. braunii* race B transcriptome [37] using TBLASTN and contigs 11859, 10353, and 43184 were found to contain matches to the peptides. These contigs were then queried against the race B genome [38] and the 401 kb scaffold 141 was identified to contain sequence matches to the contigs. To determine the full gene structure of PSAP, the race B RNA-seq reads were aligned to scaffold 141 using TopHat [39] and assembled into transcripts using Cufflinks [40]. This approach allowed identification of the 5'- and 3'-UTRs, the intron-exon boundaries, and the open reading frame for PSAP within the scaffold. The PSAP cDNA sequence has been deposited in GenBank, accession number MF36074.

Conserved domains within the PSAP protein were identified using the MyHits protein motif scan tool [41], the PSAP signal peptide was identified using the TatP [42] and Phobius [43] signal peptide prediction tools, and secretion pathway prediction was analyzed using the YLoc subcellular prediction tool [44,45].

### 2.2. In-gel glycoprotein staining

PSAP protein was extracted from shell material as described above. The extract was then separated by 8% SDS-PAGE as described previously [12]. The glycosylation status of PSAP in the gel was then analyzed using the Pro-Q Emerald 300 glycoprotein gel and blot stain kit (ThermoFisher) following the including instructions. Briefly, the SDS-PAGE gel containing PSAP was immersed in the fixing solution at room temperature for 45 min followed by one wash in the wash solution for 15 min with gentle agitation. The gel was then incubated in the oxidizing solution for 30 min and washed for 15 min. For staining, the gel was immersed in Pro-Q Emerald 300 staining solution while agitating for 2 h followed by washing twice for 15 min. The gel was then visualized under UV light using a Bio-Rad ChemiDoc XRS.

### 2.3. PSAP N-linked glycan site mapping

N-linked site mapping for PSAP was determined based on previously described methods [46]. Coomassie blue-stained SDS-PAGE gel slices corresponding to PSAP were cut into  $\sim 1\text{ mm}^3$  pieces, and destained alternately with 40 mM ammonium bicarbonate and 100% acetonitrile until the color turned clear. This was followed by re-swelling the gel slices in 10 mM DTT and 50 mM ammonium bicarbonate at 55  $^{\circ}\text{C}$  for 1 h. The DTT solution was exchanged with 55 mM iodoacetamide and incubated in the dark for 45 min followed by washing twice alternately with 40 mM ammonium bicarbonate and 100% acetonitrile. The dehydrated gel was re-swelled with chymotrypsin in 50 mM ammonium bicarbonate on ice for 45 min, followed by protein digestion at 37  $^{\circ}\text{C}$  overnight. Peptides and the glycopeptides were extracted from the gel slices with successive additions of 20% acetonitrile in 5% formic acid, 50% acetonitrile in 5% formic acid, and 80% acetonitrile in 5% formic acid. The three collected solutions were dried and combined into one tube. The chymotrypsin in the sample was inactivated, and deglycosylation was carried out with 2  $\mu\text{l}$  of PNGase A (Calbiochem) in 36  $\mu\text{l}$  of  $^{18}\text{O}$ -labeled water ( $\text{H}_2^{18}\text{O}$ ) and 2  $\mu\text{l}$  of 0.5 M citrate phosphate buffer pH 5.0. The sample was dried and resuspended in nanopure  $\text{dH}_2\text{O}$  with protease(s) to remove any possible C-terminal incorporation of  $^{18}\text{O}$  from residual protease activity [47]. The sample was then dried and analyzed by mass spectrometry.

The peptides were analyzed by LC-MS/MS as previously described [46]. Briefly, an LTQ Orbitrap XL mass spectrometer (ThermoFisher)

equipped with a nanospray ion source was used. One third of the labeled peptides was resuspended in 100  $\mu$ l mobile phase A (0.1% formic acid in dH<sub>2</sub>O) and filtered with 0.2  $\mu$ m filters (Nanosep, PALL). The sample was then loaded onto a nanospray tapered capillary column/emitter (360  $\times$  75  $\times$  15  $\mu$ m, PicoFrit, New Objective) self-packed with C<sub>18</sub> reverse-phase resin (10.5 cm, MICHROM Bioresources Inc.) in a nitrogen pressure bomb for 10 min at 1000 psi (~5  $\mu$ l load) and then separated via a 160 min linear gradient of increasing mobile phase B (80% acetonitrile, 0.1% formic acid in dH<sub>2</sub>O) at a flow rate of ~500 nL/min with direct injection into the mass spectrometer. The LTQ was run in the automatic mode collecting an MS scan followed by data dependent MS/MS scans of the six highest abundant precursor ions.

The resulting data were analyzed manually as well as via a software search. For the software search, the raw data was analyzed against the PSAP protein sequence using the TurboSequest algorithm (Proteome Discoverer 1.4, ThermoFisher). The SEQUEST parameters were set to allow 10.0 ppm of precursor ion mass tolerance and 0.8 Da of fragment ion tolerance with monoisotopic mass. Digested peptides were allowed with up to three missed internal cleavage sites. The differential modifications of 57.02146 Da, 15.9949 Da and 2.98826 Da were allowed for alkylated cysteine, oxidation of methionine, and <sup>18</sup>O-labeled aspartic acid, respectively.

#### 2.4. Free N-linked glycan preparation for linkage analysis

Free N-linked glycans were prepared from shell material based on previous studies [48] with slight modifications as follows. Shell material was homogenized using a dounce homogenizer with methanol:water (1:1 v/v) on ice. Lipids were extracted by adjusting the solvent mixture to give chloroform:methanol:water (4:8:3 v/v). The insoluble proteinaceous material was collected by centrifugation and the extraction procedure was repeated three times. The final pellet of insoluble protein was washed with cold acetone:water (4:1 v/v) two times, washed with cold acetone once, and dried under nitrogen. The sample was then reduced with 5 mM DTT for 1 h at 55 °C and carboxyamidomethylated with 15 mM iodoacetamide in the dark at room temperature for 45 min. The modified protein was dialyzed with a 4 kDa cutoff membrane (Millipore) against nanopure dH<sub>2</sub>O at 4 °C overnight, dried in a SpeedVac concentrator, dissolved in 50 mM ammonium bicarbonate, and digested with chymotrypsin at 37 °C overnight. The sample was then heated at 100 °C for 5 min to inactivate chymotrypsin, centrifuged at 2000  $\times$  g at 4 °C for 15 min, and the supernatant collected and dried. The sample was then passed through a C<sub>18</sub> Sep-Pak cartridge, washed with 5% acetic acid to remove contaminants (salts, free sugar, etc.), the peptides and glycopeptides eluted in series with 20% isopropanol in 5% acetic acid, 40% isopropanol in 5% acetic acid, and 100% isopropanol. All eluates were then combined. The N-glycans were released with PNGase A (Calbiochem) at 37 °C overnight. After digestion, the sample was passed through a C<sub>18</sub> Sep-Pak cartridge, the released N-glycans eluted with 5% acetic acid, and dried by lyophilization.

#### 2.5. N-linked glycan linkage analysis

For determination of sugar linkages, the free N-linked glycans were converted to partially methylated alditol acetates (PMAAs) as previously described [49]. Briefly, the samples were fully permethylated as previously described [50], hydrolyzed with HCl:water:acetic acid (0.5:1.5:8 v/v) at 80 °C for 18 h, reduced with 1% NaBD<sub>4</sub> in 30 mM NaOH overnight, and acetylated with acetic anhydride/pyridine (1:1, v/v) at 100 °C for 15 min. The PMAAs were then separated and analyzed by gas chromatography–mass spectrometry (GC–MS) using an Agilent 7890A GC interfaced to a 5975C mass selective detector (MSD) in electron impact ionization mode. The GC separation of the PMAAs was performed on a 30 m SP2331 fused silica capillary column (Supelco) for neutral sugar derivatives, whereas a 30 m DB-1 fused silica

capillary column (Agilent) was used for neutral and amino sugar derivatives.

#### 2.6. N-linked oligosaccharide profiling

MS profiling of N-linked oligosaccharides was performed as described previously [51,52]. Briefly, the released N-glycans were permethylated and analyzed directly on MS instruments to identify size and structural details. Matrix-assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF-MS) was performed in the reflector positive ion mode using  $\alpha$ -dihydroxybenzoic acid (DHBA, 20 mg/ml solution in 50% methanol:water) as a matrix. The spectrum was obtained using an AB SCIEX TOF/TOF 5800 (AB SCIEX). Nanospray ionization MS/MS (NSI-MSn) analysis was performed using an LTQ Orbitrap XL mass spectrometer (ThermoFisher) equipped with a nanospray ion source. The permethylated samples were dissolved in 1 mM NaOH in 50% methanol and infused into the instrument at a constant flow rate of 0.5  $\mu$ l/min. A full Fourier transform mass spectrometry (FTMS) spectrum was collected at a resolution of 30,000. The capillary temperature was set at 210 °C and MS analysis was performed in the positive ion mode. For total ion mapping (automated MS/MS analysis), the *m/z* range of 600 to 2000 was scanned in ITMS mode in successive 2.8 mass unit windows that overlapped the preceding window by 2 mass units.

#### 2.7. Hydroxyproline O-linked glycan analysis

For the analysis of hydroxyproline O-glycosylation, PSAP was extracted from 6 mg of shell material by boiling the material with SDS-PAGE buffer as described above. After protein extraction, Cys residues were carbamidomethylated and the protein was precipitated with acetone. The protein precipitate was washed with chloroform:methanol (2:1 v/v) to remove lipids, resuspended in Tris-HCl pH 8.2 with 10 mM CaCl<sub>2</sub>, and digested with Pronase (Roche) at 37 °C for 48 h. The sample was dried, an aliquot dissolved in 200  $\mu$ l of mobile phase A (0.1% formic acid in dH<sub>2</sub>O), and the sample passed through a 0.2  $\mu$ m filter (Nanosep, PALL).

For the LC-MS analysis of the pronase digest, the sample was separated on a nano-C<sub>18</sub> column (Acclaim pepMap RSLC, 75  $\mu$ m  $\times$  150 mm, 2  $\mu$ m particle size) via a 80 min gradient of increasing mobile phase B (80% acetonitrile, 0.1% formic acid in dH<sub>2</sub>O) at a flow rate of 300 nL/min into the mass spectrometer. MS analysis was performed on an Orbitrap-Fusion equipped with an EASY nanospray source and Ultimate3000 autosampler LC system (ThermoFisher). For online MS detection, full MS data was first collected at a resolution of 600,000 in FT mode. MS/MS-CID data and MS/MS-HCD activation data (both in FT mode) were obtained for each precursor ion by data dependent scan (top speed scan, 3 s). The resulting data were analyzed manually.

#### 2.8. Functional analysis of the PSAP secretion signal peptide

The synthetic gene product, SP<sub>PSAP</sub>-GFP, containing the PSAP signal peptide (SP) fused to the N-terminus of GFP was synthesized with *S. cerevisiae* codon optimization (ThermoFisher), inserted into the pMA-T plasmid using the *EcoRI* and *NotI* restriction sites, and subcloned into pESC-TRP using the same restriction sites for expression under the control of the inducible GAL10 promoter. Additionally, a GFP:pESC-TRP construct was generated for control experiments. The constructs were transformed into *S. cerevisiae* strain CKY457 (*MATa*, *leu2Δ1*, *ura3-52*, *trp1Δ63*, *his3Δ200* and *lys2-1288*) via lithium acetate transformation, positive transformants were grown at 30 °C in 40 ml SC-TRP media to O.D<sub>600</sub> = 0.6, protein expression induced by replacing media with fresh media containing 2% galactose, and the cells grown for an additional 8 h. The cells were then centrifuged at 1500  $\times$  g for 30 min at room temperature and both the cell pellet and supernatant (media)

were saved for protein extraction. To extract protein from the cells, the cell pellet was resuspended in 600  $\mu$ l lysis buffer (0.1 M NaOH, 2%  $\beta$ -mercaptoethanol, 2% SDS, and 0.05 M EDTA), incubated at 95 °C for 10 min, 4.56  $\mu$ l of 4 M acetic acid added, incubated at 95 °C for 10 min, 150  $\mu$ l of loading dye added (0.25 M Tris-HCl pH 6.8, 50% glycerol, and 1  $\times$  bromophenol blue), incubated at 95 °C for 10 min, and centrifuged at 3380  $\times$  g for 2 min at room temperature. The collected media for each sample was run through a 0.22  $\mu$ m filter (Millipore Express PLUS Membrane, EMD Millipore) to eliminate any *S. cerevisiae* cells in the media and then lyophilized. For protein extraction from the lyophilized media, 2 ml dH<sub>2</sub>O was added to each sample, the samples run through a 10 kDa cutoff centrifugal filter (Amicon Ultra, EMD Millipore) to eliminate reagents and concentrate the samples to ~60  $\mu$ l, and 20  $\mu$ l of 4  $\times$  SDS-PAGE sample buffer added to each sample. Twenty-five microliters of the cell protein extract and all of the media protein extract were separated by 12% SDS-PAGE, transferred to a PVDF membrane, and analyzed by western blot following standard procedures. The  $\alpha$ GFP-HRP antibody (1:3000 for cell protein extracts and 1:1500 for media protein extracts, Santa Cruz Biotechnology) was incubated with the membrane at 4 °C overnight with rotation. Detection and visualization were carried out using the Amersham ECL Prime Western Blotting Detection Reagent and an Amersham Imager 600RGB, respectively.

## 2.9. Lectin blot

PSAP was extracted from shell material, separated by SDS-PAGE, transferred to a PVDF membrane as described above, and detection using lectins followed standard western blot procedures. The biotinylated lectin (10  $\mu$ g) was first incubated with the membrane for 30 min at room temperature, followed by treatment with the VectorLabs VECTASTAIN Elite ABC HRP Kit following included instructions. Detection and visualization were carried out using the Amersham ECL Prime Western Blotting Detection Reagent and an Amersham Imager 600RGB, respectively. The biotinylated lectins used for this analysis were *Hippeastrum* hybrid lectin (HHL), *Lycopersicon esculentum* lectin (LEL), and *Sambucus nigra* lectin (SNA), all from VectorLabs.

## 2.10. PSAP peptide antibody generation and western blot analysis

A polyclonal PSAP specific antibody was raised in rabbit by GenScript using the PSAP peptide 854-TGVRRIIAPPGTGFGD-869 with a Cys added at the N-terminus for conjugation to a carrier protein. Specificity of the antibody was tested by western blot using PSAP protein extracted from shell material as described above and a *B. braunii* total protein extract. The total protein extract was prepared by adding 350  $\mu$ l of the extraction buffer (50 mM MOPS pH 7.3, 20 mM MgCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol, 5 mM EGTA, 20% glycerol) to 200 mg of race B tissue, mixing well, bead beating for 10 min, and removing the cell debris by centrifuging at 10,000  $\times$  g for 10 min at 4 °C. Following 12% SDS-PAGE, proteins were transferred to a PVDF membrane and western blot analysis followed standard procedures. The PSAP antibody (1:500) was incubated with the membrane at 4 °C overnight with rotation, followed by incubation with a goat anti-rabbit-HRP secondary antibody (1:3000; ThermoScientific) for 2 h at room temperature. Detection and visualization were carried out using the Amersham ECL Prime Western Blotting Detection Reagent and an Amersham Imager 600RGB, respectively.

## 2.11. PSAP localization to shell material

Approximately 100  $\mu$ g of lyophilized shell material was rehydrated in 200  $\mu$ l PBS, centrifuged at 15,000  $\times$  g, and the PBS discarded. The shell material was then blocked with 5% dry milk for 30 min, washed with PBS three times, and incubated with the biotinylated HHL (1:5000, 30 min, 25 °C), LEL (1:250, 30 min, 25 °C), and SNA (1:500,

30 min, 25 °C) lectins, and the PSAP antibody (1:100, 12 h, 4 °C). The shell material was then washed three times for 5 min at 25 °C with PBS + 0.05% Tween 20 (PBST).

Detection of the biotinylated lectins was carried out by incubating the shell material with a streptavidin-Texas red conjugate (VectorLabs) in PBS (1:1000) for 30 min in the dark, washing three times with PBST, and imaging on a Nikon Eclipse Ti inverted epifluorescence microscope using a 100  $\times$  objective (Plan Fluo, NA 1.40, oil immersion) with a 2.5 TV relay lens. A mercury lamp was used as the light source (X-Cite 200 DC, Excelitas Technologies) within a cage incubator (InVivo Scientific) at 30 °C, and images acquired using a cooled EMCCD (electron multiplying charge-coupled device) camera (iXon3 897, Andor, Belfast, Ireland). The images were acquired as phase-contrast (200 ms) and Texas red cube (200 ms).

The HHL and LEL lectins were also detected using diaminobenzidine (DAB). Following incubation with the lectins as above, the shell material was washed three times for 5 min at 25 °C with PBST. Ten microliters of Reagent A and ten microliters of Reagent B from the VECTASTAIN Elite ABC HRP Kit were added to 1 ml of PBS buffer, mixed well with the shell material, and allowed to stand for 30 min. The shell material was then washed three times with PBST, 550  $\mu$ l of a DAB solution (25  $\mu$ l of 1% DAB in dH<sub>2</sub>O, 25  $\mu$ l of 0.3% H<sub>2</sub>O<sub>2</sub>, 500  $\mu$ l PBS, pH 7.2) was added, mixed well, and incubated for 3 min or 10 min at 25 °C. The shell material was then visualized using a Zeiss Axiophot light microscope equipped with a Plan Neofluar 100/1.3 oil immersion objective and a Coolsnap CF monochrome CCD camera (Photometrics) controlled by MetaView version 5.2 software (Molecular Devices) at the Texas A&M University Microscopy and Imaging Center.

For detection of the PSAP antibody, following incubation with the shell material as described above the sample was washed three times for 5 min at 25 °C with PBST. The sample was then incubated with a mouse anti-rabbit-Texas red conjugated secondary antibody (1:1000) for 2 h at room temperature followed by three washes with PBST. The shell material was then imaged using the fluorescence microscope described above.

To visualize unstained shell material under transmission electron microscopy, lyophilized shell material was resuspended in water, applied to a carbon support film on an EM grid, dried, and imaged on the 1200 EX transmission electron microscope (JEOL USA, Inc.). Images were captured on a 15C CCD camera (SIA).

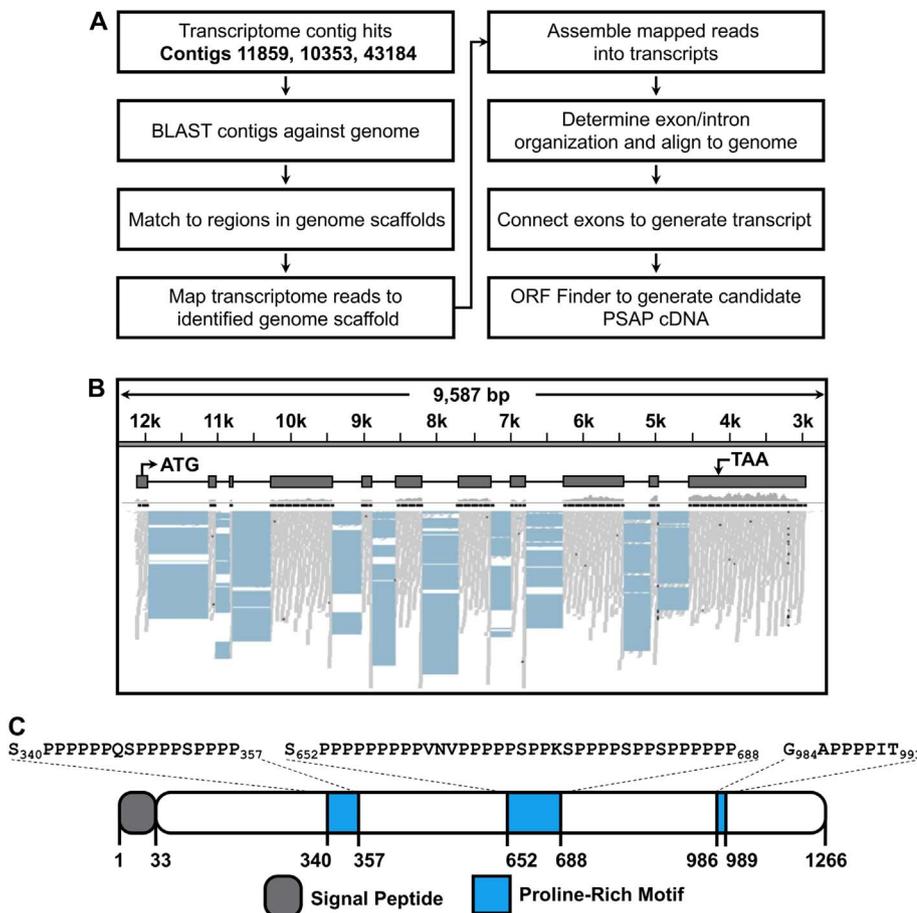
## 3. Results

### 3.1. Identification of cDNA and protein sequences for polysaccharide associated protein

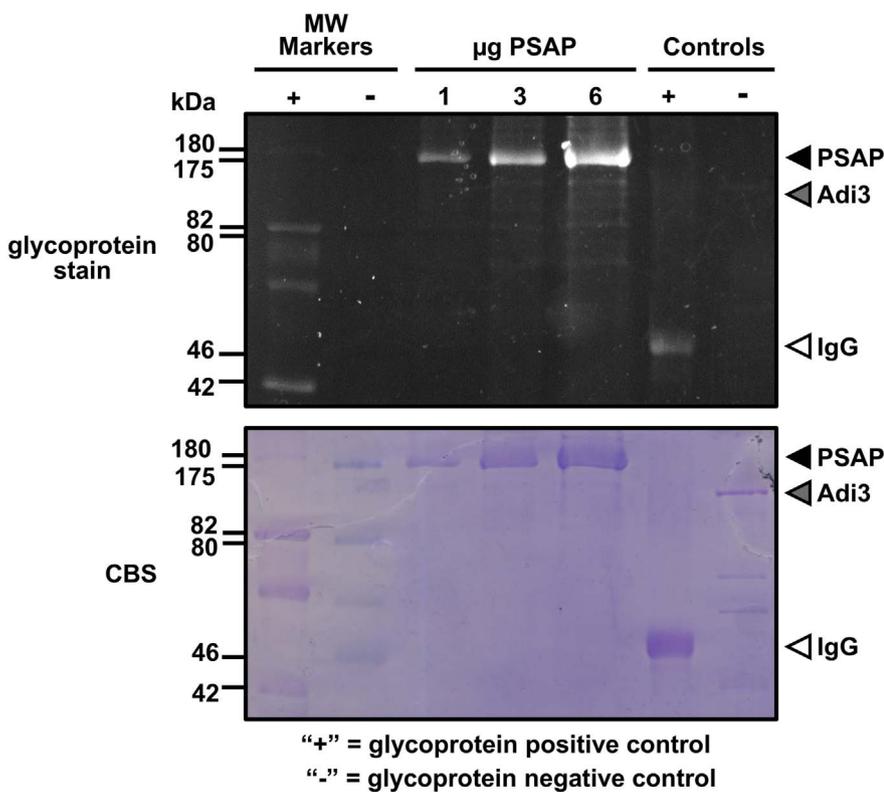
Analysis of the *B. braunii* race B shell material in our previous study showed a single protein species associated with the shells [12]. This was confirmed in our current study (Supplementary Fig. 1A) and we termed this protein polysaccharide associated protein (PSAP). The molecular mass of PSAP was previously estimated at ~150 kDa [12], and a more detailed analysis estimated PSAP molecular mass at 162.2 kDa (Supplementary Fig. 1B).

In order to identify the PSAP protein and clone the associated cDNA, the PSAP protein from SDS-PAGE (Supplementary Fig. 1A) was excised from the gel, digested with trypsin, the generated peptides identified by LC-MS/MS, and the resulting peptides BLASTed against a computational translation of a *B. braunii* race B transcriptome [37] (Supplementary Fig. 2A). This process identified three contigs containing matches to PSAP peptides (Supplementary Figs. 2B and S3). A combination of these contigs did not yield a full length open reading frame, and thus further analysis was done to identify the full length PSAP sequence.

The full length PSAP cDNA and protein sequence were identified as laid out in Fig. 1A. Briefly, the three identified contigs were BLASTed against the *B. braunii* race B genome [38] to match the contigs to a



**Fig. 1.** Identification of the PSAP cDNA and protein sequences using bioinformatics tools. (A) Bioinformatics steps followed to identify the full length PSAP cDNA. (B) Alignment of PSAP transcripts to genomic scaffold 141 to identify exon/intron junctions for assembling the PSAP cDNA. Solid gray line with kb markers indicates location of PSAP gene sequence within scaffold 141. Below the scaffold line the PSAP exons are indicated by gray boxes and introns by black lines. The translation start and stop locations are indicated. Transcriptome sequence reads aligned to the scaffold are shown as gray lines under the PSAP exons. Spaces where no transcriptome reads matched the PSAP sequence are shown as blue boxes. (C) Conserved domains found in the PSAP protein and the amino acid sequence of the Pro-rich motifs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 2.** In-gel glycoprotein staining of PSAP. Three amounts of the PSAP protein plus positive and negative controls were separated by SDS-PAGE, treated with the glycoprotein stain, and visualized by UV light. The gel was stained with Coomassie blue (CBS) following the glycoprotein stain for visualization of all proteins in the assay. Positive controls include glycosylated molecular weight markers and a mouse antibody (IgG). Negative controls include non-glycosylated molecular weight markers and the tomato Adi3 protein. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 1**  
Linkage analysis of the N-linked glycans on PSAP.

| Peak # | RT (min) | Peak area   | Assignment         | Peak area %        |
|--------|----------|-------------|--------------------|--------------------|
| 1      | 7.255    | 895,846     | T-Araf             | 0.6%               |
| 2      | 7.513    | 1,571,949   | T-Arap             | 1.1%               |
| 3      | 7.725    | 6,361,609   | T-Fuc              | 4.6%               |
| 4      | 8.24     | 369,399     | 4-Fuc              | 0.3%               |
| 5      | 8.509    | 85,705,175  | T-Man <sup>a</sup> | 59.3% <sup>a</sup> |
|        |          |             | T-Glc <sup>a</sup> | 2.8% <sup>a</sup>  |
| 6      | 8.646    | 764,836     | T-Gal              | 0.6%               |
| 7      | 9.224    | 11,076,348  | 2-Man              | 8.0%               |
| 8      | 9.321    | 3,721,995   | 3-Man              | 2.7%               |
| 9      | 9.521    | 1,811,817   | 6-Man              | 1.3%               |
| 10     | 10.568   | 9,919,887   | 3,6-Man            | 7.2%               |
| 11     | 12.291   | 14,740,972  | 4-GlcNAc           | 10.7%              |
| 12     | 13.309   | 1,001,179   | 3,4-GlcNAc         | 0.7%               |
| Total  |          | 137,941,012 |                    | 100.0%             |

Araf: arabinofuranose; Fuc: fucose; Arap: arabinopyranose; Man: mannose; Glc: glucose; Gal: galactose; GlcNAc: N-acetylglucosamine; T: Terminal; 2-: 2-linked; 3,6-: 3 and 6 linked (branched); 3,4-: 3 and 4 linked (branched).

<sup>a</sup> These sugars co-eluted on a DB-1 column but were well-separated on an SPB-column. The peak area % of terminal mannose and terminal glucose in the co-eluted peak was calculated based on peak area % of the corresponding peaks observed in SBP-column separation.

9.5 kb region of a single 401 kb genome scaffold (Fig. 1B). The race B transcriptome was then aligned to the identified scaffold to map potential PSAP transcriptome reads to this scaffold (Fig. 1B). The mapped transcriptome reads identified 11 exons and 10 introns spanning approximately 9 kb of genomic sequence (Fig. 1B). The exons were assembled into a 4928 bp cDNA sequence containing a PSAP open reading frame of 3798 bp (Supplementary Fig. 4A) that encodes a protein of 1266 amino acids (Supplementary Fig. 4B).

Analysis of the full length PSAP protein and cDNA sequence by BLAST against the entire GenBank database showed no identity to any proteins in the database. However, a domain search revealed an N-terminal 33 amino acid secretion signal peptide and three Proline (Pro)-rich motifs (Fig. 1C).

### 3.2. PSAP is a glycoprotein

The predicted molecular mass of PSAP is 132.8 kDa based on the open reading frame (Supplementary Fig. 4B) and the estimated PSAP molecular mass from shell material is 162.2 kDa (Supplementary Fig. 1B). This discrepancy in size suggests PSAP posttranslational modification such as glycosylation since many plant extracellular proteins are glycosylated [53]. Thus, glycosylation of PSAP was analyzed using an in-gel glycoprotein stain based on the periodic acid-Schiff (PAS) method. In this assay, periodic acid converts sugar vicinal diols to aldehydes, which are bound by the Schiff reagent for detection under UV light [54]. Three different amounts of PSAP plus positive and negative controls were separated by SDS-PAGE and the gel stained for glycoproteins. Positive controls included glycosylated molecular weight markers and a mouse antibody (IgG). Negative controls included non-glycosylated molecular weight markers and the tomato AGC protein kinase Adi3 as expressed and purified from *E. coli* [55]. In the glycoprotein staining assay all positive and negative controls showed the correct staining pattern, and the PSAP protein showed staining in a concentration dependent manner (Fig. 2). This suggests PSAP is glycosylated and there may be ~29.4 kDa of glycosylation on PSAP to reach the observed 162.2 kDa size.

### 3.3. Identification of PSAP N-linked glycans

Since the glycan staining assay confirmed PSAP as a glycoprotein (Fig. 2), the identity of the PSAP glycans was investigated. First, the possibility of N-linked glycosylation on Asn residues was analyzed. A

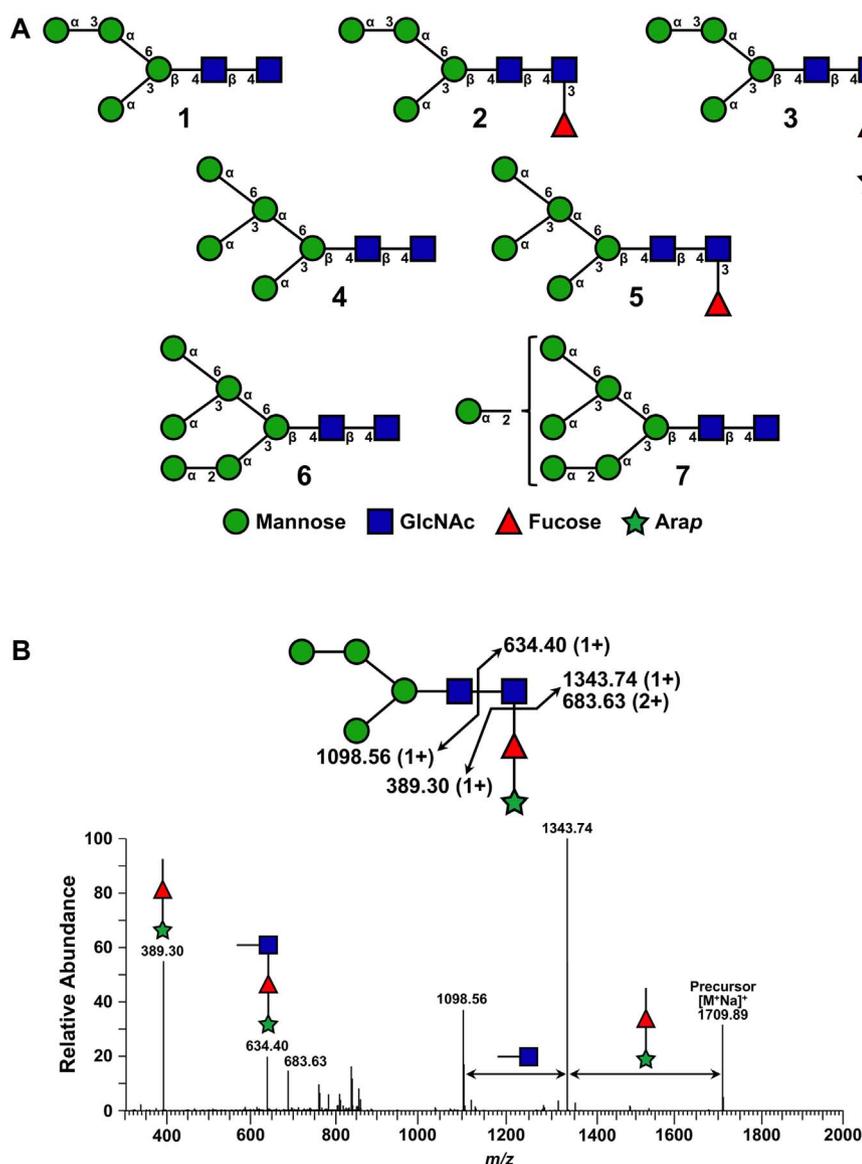
computational analysis showed PSAP contains 96 Asn residues and 14 of these 96 (Supplementary Fig. 5) are found within the known NxS/T N-glycosylation consensus sequence [56]. Which of the 14 Asn residues contained N-linked glycans was then analyzed experimentally. Briefly, PSAP was in-gel digested with chymotrypsin followed by enzymatic N-glycan release using PNGase A in an <sup>18</sup>O-labeled dH<sub>2</sub>O buffer, which will release the N-linked glycans, convert the glycan linked Asn to an <sup>18</sup>O-labeled Asp, and introduce a 3 Da mass shift at the glycosylation site. The resulting peptides were analyzed by LC-MS to identify <sup>18</sup>O-labeled peptides for N-linked glycan site mapping. This analysis identified 25 peptides covering 40% of the PSAP protein sequence including 12 peptides containing 13 of the 14 potential N-linked glycan sites (Supplementary Fig. 5). A peptide containing potential N-linked glycan site N1178 was not identified in the analysis. Within the 12 identified peptides containing potential N-linked glycan sites, residues N54, N718, N728, N767, N897, N930, N1006, N1052, N1064, and N1155 were found to be N-linked glycan sites (Supplementary Fig. 5). Possible N-linked glycan sites N615, N620, and N1212 were found to not be glycosylated. Overall, 10 of the possible 14 N-linked glycan sites were confirmed to harbor glycans.

Next, structural details of the PSAP N-linked glycans were determined by releasing the glycans with PNGase A, purifying the free glycans [48], and subjecting the glycans to in-depth structural characterization using several MS techniques. Conversion of the glycans to partially methylated alditol acetates (PMAAs) followed by a GC-MS glycosyl linkage analysis showed the N-linked glycans contained mainly terminal, 2-linked, 3-linked, 6-linked, and 3,6-linked mannose (Table 1). The next most abundant sugar was 4-linked N-acetylglucosamine (GlcNAc) followed by terminal fucose and terminal arabinopyranose (Table 1).

In order to determine the size and glycosyl sequence of the PSAP N-linked glycans, the released and purified N-glycans were permethylated and profiled directly by NSI-MSn (Supplementary Fig. 6). The analysis revealed seven different glycan structures (Fig. 3A). The main N-glycan components are high-mannose type structures, Man<sub>4,7</sub>GlcNAc<sub>2</sub>, that contain a base of (1 → 4)-connected GlcNAc with branches of mannose connected by (1 → 3)- and (1 → 6)-linkages (Fig. 3A). Interestingly, three of the glycans were seen to contain a fucose (glycans 2, 3, and 5 in Fig. 3A) and one unique glycan contained an additional arabinopyranose (glycan 3 in Fig. 3A). Tandem MS analysis was performed to identify the fucose and arabinose positions on the glycan ion at *m/z* = 1709.85 corresponding to Man<sub>4</sub>GlcNAc<sub>2</sub>Fuc<sub>1</sub>Ara<sub>1</sub> (Supplementary Fig. 6). In the fragmentation analysis of this ion (Fig. 3B), the presence of fragment ions at *m/z* = 389.3 (AraFuc) and *m/z* = 634.4 (AraFucGlcNAc) indicates the AraFuc disaccharide is attached at the reducing-end GlcNAc. This MS/MS data together with the glycosyl linkage analysis indicated the fucose is linked to the terminal GlcNAc at the 3 position (glycans 2, 3, and 5 in Fig. 3A), and the arabinopyranose is linked to the fucose at the 4 position (glycan 3 in Fig. 3A). The total molecular mass of the released glycans is predicted to be between 10–16 kDa.

### 3.4. Identification of PSAP hydroxyproline O-linked glycans

PSAP was also analyzed for mucin-type Ser or Thr O-linked glycosylation by chymotryptic digestion followed by LC-MS analysis of the resulting peptides, but this analysis did not identify any mucin-type O-linked glycans. Next, since PSAP contains three Pro-rich domains, which commonly have the Pro hydroxylated to hydroxyproline (Hyp) and glycosylated in plant extracellular matrix proteins [15], PSAP was analyzed for Hyp O-linked glycans. For this analysis, PSAP was reduced, alkylated, and digested with pronase, which should digest PSAP into single amino acids or linkages of a few amino acids. However, the PSAP Pro-rich regions will remain nearly intact since pronase is not active toward Pro. The resulting digestion mixture was separated and analyzed by LC-MS (Fig. 4A, B). The analysis identified one Pro-rich



**Fig. 3.** Structures of the *N*-linked PSAP glycans. (A) Structures of the seven *N*-linked glycans found on PSAP as determined by the MS data shown in Supplementary Fig. 6. (B) Tandem MS analysis of the unique fucose and arabinose containing *N*-glycan ion detected at  $m/z = 1709.85$  shown in Supplementary Fig. 6. Key ions for identifying the location of arabinose and fucose are shown diagrammatically above the MS spectrum. Presence of a C-type fragment ion at  $m/z = 389.30$ , corresponding to an AraFuc disaccharide unit, and a Y-type fragment ion at  $m/z = 634.40$ , corresponding to an AraFucGlcNAc trisaccharide unit, indicated the AraFuc disaccharide is attached to the *N*-glycan chitobiose core.

peptide containing four consecutive Pro residues at positions 986–989 (Fig. 4B), all of which are hydroxylated to Hyp and carry arabinogalactan-type oligosaccharides consisting of 1 6-deoxyhexose, 4 hexoses, and 7 pentoses (Fig. 4A, B). A linkage or composition analysis was not performed for the *O*-glycans, but the MS data (Fig. 4A, B) indicates a pentose is attached to the peptide backbone and the 6-deoxyhexose is terminally linked. This observation is consistent with the Hyp *O*-glycans found on plant arabinogalactan proteins [57]. Based on this analysis, the derived *O*-linked glycans on PSAP are shown in Fig. 4C. It should be noted, at this time it is not clear which glycan is attached to each Hyp. The total molecular mass of the Hyp *O*-linked glycans is predicted to be between 1–2 kDa.

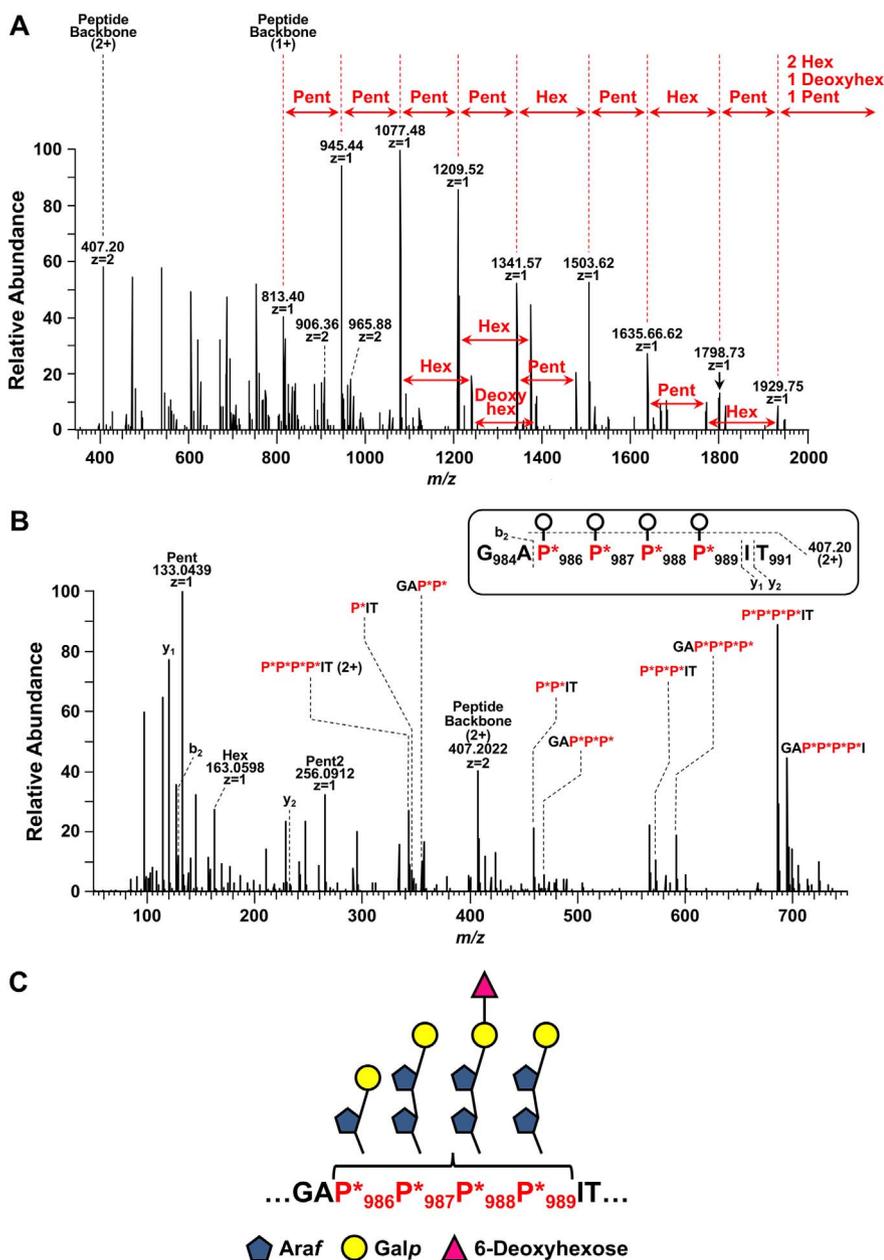
### 3.5. Estimation of PSAP molecular mass based on glycosylations

Taking into account the molecular mass of the naked PSAP protein and the glycosylations found on PSAP, the glycosylated molecular mass of PSAP is calculated to be between 143.8 and 150.8 kDa. This takes into account 132.8 kDa from PSAP, 10–16 kDa from *N*-linked glycans and 1–2 kDa from Hyp *O*-linked glycans. This value is below the calculated value of 162.2 kDa for PSAP based on SDS-PAGE migration (Supplementary Fig. 1B), suggesting there may be additional glycosylations or other modifications on PSAP to account for the

~11.4–18.4 kDa needed to reach the observed mass. It should be noted that if an SDS-PAGE gel is overloaded with PSAP a ladder of higher molecular mass PSAP can be seen above the primary PSAP band. This can be seen in Fig. 2 and may represent PSAP with different degrees of glycosylation. Moreover, the difference between the calculated and observed PSAP mass on SDS-PAGE could also be attributed to the low mobility and higher apparent molecular weight of Pro-rich proteins on SDS-PAGE [58,59].

### 3.6. PSAP contains a functional secretion signal peptide

Next we undertook a set of experiments to confirm that PSAP is indeed localized outside of *B. braunii* cells in the ECM and with the shell material. First, subcellular localization prediction tools [44,45] generated a 100% probability that PSAP is localized to the secretion pathway via the predicted signal peptide. Next, the functionality of the PSAP signal peptide as a secretion signal was tested through an N-terminal translational fusion to GFP (Supplementary Fig. 7A). This fusion was cloned into a galactose inducible yeast expression vector, expressed for 8 h in the presence of galactose followed by collection of the cells and media separately. Analysis of cellular and media proteins for the presence of GFP by  $\alpha$ -GFP western blot showed a small amount of GFP was found in the media without the PSAP signal peptide, while a large



**Fig. 4.** Analysis of PSAP Hydroxyproline O-linked glycans by LC-MS. PSAP was reduced, alkylated, digested with pronase, and analyzed by nano  $C_{18}$ -LC-MS. A glycopeptide ion, corresponding to  $G_{984}AP^*P^*P^*P^*IT_{991}$  from PSAP carrying 1 deoxyhexose, 4 hexoses, and 7 pentoses eluted at 16.81 min of the LC-MS run. (A) The MS/MS-CID spectrum of the glycopeptide at  $m/z = 1266.99$  ( $z = 1266.4869$ ) showing a series of neutral losses of glycan moieties in the order of deoxyhexose, hexose, and then pentose, suggesting the pentose is attached to a hydroxyproline, whereas the hexose and deoxyhexose are on the non-reducing terminal end. (B) MS/MS-HCD spectrum (low mass region) of the same glycopeptide ion. The raw mass region of the HCD data was analyzed to determine amino acid content and sequence. The series of internal fragment ions from the peptide backbone allowed determination of the peptide sequence as  $G_{984}AP^*P^*P^*P^*IT_{991}$  from PSAP. All Pro in the peptide were hydroxylated (hydroxyproline, Hyp). (C) Hydroxyproline O-linked glycan structures found on PSAP on Hyp at positions 986–989 based on MS data shown in (A) and (B). Hex, hexose; Pent, pentose; Deoxyhex, deoxyhexose; P\*, hydroxyproline.

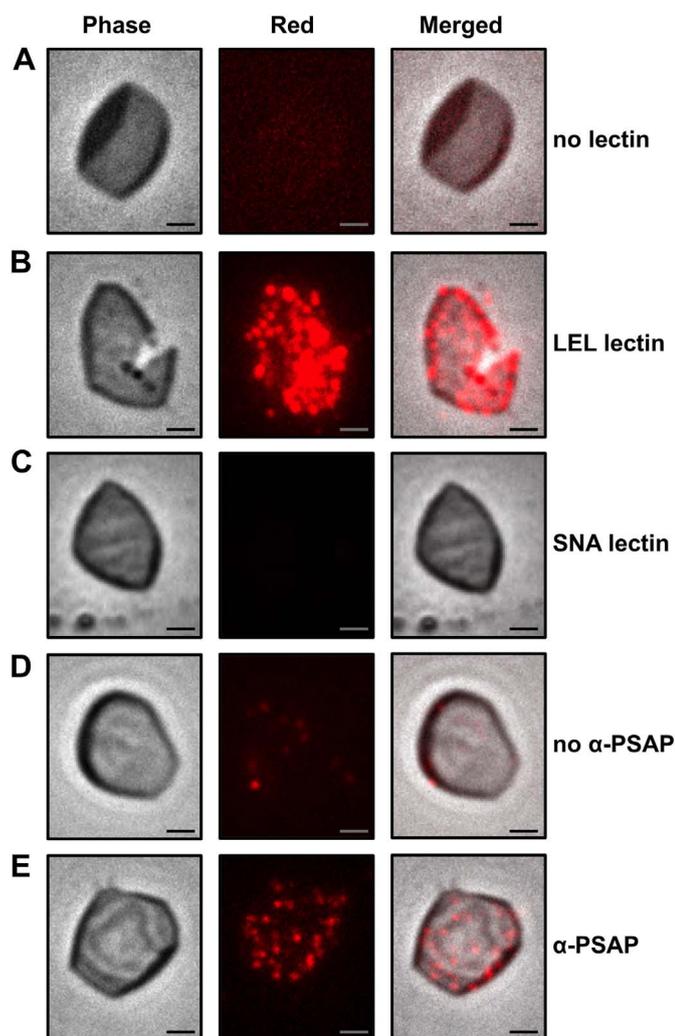
amount of GFP was found in the media when expressed with the PSAP signal peptide (Supplementary Fig. 7B). These studies suggest the PSAP signal peptide is a functional secretion signal for exporting PSAP outside of *B. braunii* cells into the ECM.

### 3.7. Localization of PSAP to shell material

The second set of experiments performed to confirm PSAP localization to the ECM shell material were based on glycan detection with lectins coupled to fluorescence microscopy. Additionally, an  $\alpha$ -PSAP peptide antibody was developed and also used to confirm PSAP localization to shell material. An initial analysis of shell material by electron microscopy showed many small dots on the shell surface (Supplementary Fig. 8) that resembles the granules seen in our previous studies [12]. We speculated that these small dots are the PSAP protein in granules.

We have shown that PSAP is a mannose and GlcNAc containing glycoprotein (Fig. 3A). The *Lycopersicon esculentum* lectin (LEL) specifically binds to GlcNAc [60] and the *Hippeastrum* hybrid lectin (HHL)

binds specifically to  $\alpha$ -1  $\rightarrow$  3 and  $\alpha$ -1  $\rightarrow$  6 linked mannose structures [61]. All of these types of sugars are found in the PSAP glycans (Fig. 3A). Both lectins were capable of detecting PSAP protein in shell material as well as a total protein extract by western blot using biotin labeled lectins and streptavidin-HRP (Supplementary Fig. 9A, B). Next, the lectins were used in fluorescence microscopy. Shell material was incubated separately with biotin labeled LEL or HHL, followed by incubation with streptavidin-Texas red and analysis by fluorescence microscopy. Both lectins were seen to bind to the shell material with the LEL lectin labeling small dots on the shells similar to the granules (Fig. 5A–B and Supplementary Figs. 9E, 10A–C). The *Sambucus nigra* lectin (SNA), which is specific for sialic acid was used as a negative control (Fig. 5C and Supplementary Figs. 9C, 10D). A similar procedure was followed with the exception of following lectin binding with streptavidin-HRP incubation and detection with diaminobenzidine (DAB), which produces a dark staining that can be seen by light microscopy. The results show that the lectins labeled the shell material in a lectin dependent manner (Supplementary Fig. 11). These studies suggest the lectins are specifically binding to PSAP in the shell material



**Fig. 5.** Detection of PSAP in shell material by fluorescence microscopy using PSAP glycan specific lectins and the  $\alpha$ -PSAP antibody ( $\alpha$ -PSAP). (A) Negative control. Shell material was incubated without a lectin followed by incubation with streptavidin-Texas red (1:1000) and analyzed by fluorescence microscopy. (B) PSAP detection in shell material using the LEL lectin. Shell material was incubated with LEL (1:250) followed by incubation with streptavidin-Texas red (1:1000) and fluorescence microscopy. (C) Lectin negative control. Shell material was incubated with the SNA lectin (1:500) followed by incubation with streptavidin-Texas red (1:1000) and fluorescence microscopy. (D) Negative control for  $\alpha$ -PSAP antibody detection of PSAP. Shell material was incubated without  $\alpha$ -PSAP followed by incubation with  $\alpha$ -rabbit-Texas red (1:1000) and fluorescence microscopy. (E) PSAP detection in shell material using the  $\alpha$ -PSAP antibody. Shell material was incubated with  $\alpha$ -PSAP (1:100) followed by incubation with  $\alpha$ -rabbit-Texas red (1:1000) and fluorescence microscopy. Scale bars, 2  $\mu$ m. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

since PSAP is the only glycosylated protein associated with the shell material. Additionally, it is unlikely the lectins are binding to the polysaccharide fibers associated with the shell material since these polysaccharides do not contain the mannose or GlcNAc sugars identified by the lectins [12].

An  $\alpha$ -PSAP antibody was also developed for use in PSAP location studies. A 16 amino acid peptide located at positions 854–869 (Supplementary Fig. 5) was used for antibody development. This region was chosen to avoid any of the identified PSAP glycosylations. However, it should be noted the peptide used contained an Ala-Pro-Pro sequence. These Pro could potentially be hydroxylated and glycosylated in vivo, which would likely interfere with antibody detection. This does not seem to be the case since the  $\alpha$ -PSAP antibody appeared to be very specific, detecting PSAP from shell material and detecting only the

PSAP protein from a total protein extract (Supplementary Fig. 9D). Incubation of the  $\alpha$ -PSAP antibody with shell material followed by incubation with an  $\alpha$ -rabbit-Texas red conjugated secondary antibody and fluorescence microscopy showed labeling of PSAP in small dots on the shell material reminiscent of the protein granules (Fig. 5D, E and Supplementary Fig. 12). The  $\alpha$ -PSAP antibody was also used in immunogold labeling studies on thin slices of *B. braunii* colonies in order to localize PSAP on the surface of the retaining wall. However, gold labeling was not seen, possibly due to the  $\alpha$ -PSAP antibody antigen not being surface exposed under the conditions used for these studies. In the future, whole-cell fluorescence microscopy with the PSAP antibody should be attempted to confirm PSAP localization within the colony ECM. Taken together, these studies suggest the PSAP protein is localized to the shell material and correlates to the protein granules seen in our previous study [12].

#### 4. Discussion

While the makeup of and potential biosynthesis mechanisms for the cross-linked hydrocarbon network portion of the *B. braunii* ECM has been well studied [29–32], and the ECM polysaccharides have been studied in terms of ultrastructure, composition, and production levels [12,27,33,34,36,62–65], proteins have not previously been identified as part of the *B. braunii* ECM. Here we have identified the unique glycosylated *B. braunii* ECM protein PSAP that is associated with the polysaccharide fibers in the ECM, and may have a role in the biosynthesis of these ECM polysaccharide fibers for this alga.

##### 4.1. PSAP is a unique hydroxyproline-rich glycoprotein (HRGP)

We have shown here the *B. braunii* PSAP protein is *N*-glycosylated at ten Asn residues and *O*-glycosylated at four consecutive Hyp residues. In addition to this Hyp-rich region, PSAP contains two additional Pro-rich regions, and thus PSAP is classified as an HRGP. Interestingly, the overall PSAP sequence does not match any other known sequences in the NCBI database at both the DNA and protein levels. Thus, PSAP appears to be a new and unique HRGP. Within the other *B. braunii* races, a PSAP homologue appears to be present in both the A and L races based on screening of unpublished transcriptome sequences generated in our lab. However, a full length PSAP-like transcript is not present in these transcriptomes, and the generation of A and L race genome sequences will be required to positively identify a PSAP homologue from these races.

HRGPs are found within the cell walls and ECMs of photosynthetic organisms ranging from green algae to land plants with a wide variety of roles including cell expansion and responses to pathogens [66,67]. HRGPs are classified into three main categories based on the types of Hyp-rich motifs found in the proteins; arabinogalactan proteins (AGPs) with [Ala/Ser/Thr]-Hyp-[Ala/Ser/Thr]-Hyp repeats, extensins (EXTs) with Ser-Hyp<sub>3-6</sub> repeats, and Pro-rich proteins (PRPs) with Pro-Hyp-Val-Tyr repeats [14,15,66,67]. In all cases the Hyp residues are *O*-glycosylated to varying levels with arabinose and galactose in the non-contiguous Hyp repeats of AGPs consisting of up to 99% of the protein molecular mass [15,67], or with one to four arabinose units in the contiguous Hyp repeats of EXTs constituting 35–65% of the molecular mass [15,68]. The Hyp residues in PRPs are minimally (3%–10%) glycosylated with arabinose or are not glycosylated at all [66].

While PSAP does not fit well into any of these specific HRGP categories, PSAP appears to be most related to EXTs. For example, while the first two PSAP Pro-rich motifs, if the Pro are converted to Hyp, contain the typical Ser-Hyp<sub>3-6</sub> sequence found in EXTs, PSAP also contains shorter Ser-Hyp<sub>2</sub> and longer and Ser-Hyp<sub>9</sub> sequences (Fig. 1C). However, none of the Pro in these domains were found to be hydroxylated or glycosylated. The Hyp region that was found to be glycosylated, Ala-Hyp-Hyp-Hyp (Fig. 4C), while rare in land plants and can be found in both AGPs and EXTs [66], does not match the typical

EXT glycosylated SerHyp<sub>3–6</sub> sequence. Moreover, the O-linked glycans found on this PSAP Hyp region are different from typical EXT Hyp-linked glycans, which contain only AraF [15,69]. Additionally, PSAP contains ~18% glycosylations by molecular mass, which is closer to that of PRPs.

This data raises the question whether the Pro residues in the first two PSAP Pro-rich motifs are actually hydroxylated to Hyp, which in HRGPs typically occurs posttranslationally, followed by glycosylation [15]. Since no glycosylations were found in these Pro rich regions, it is possible they were not converted to Hyp. No peptides covering these regions were found in the original peptide mass fingerprinting to identify the PSAP cDNA (Supplementary Fig. 2B), so it is not known at this time if they are converted to Hyp [70]. Amino acid analysis could be done to determine the amount of Hyp in these Pro-rich regions.

Another difference between PSAP and typical HRGPs is the high degree of N-glycosylation found in PSAP. Generally, plant HRGPs do not contain N-linked glycans [57], although some AGPs have been shown to have N-glycans [71]. On the other hand, PSAP appears to have the majority of its glycosylations at N-linked sites, which is not typical for HRGPs.

The PSAP Hyp O-linked glycans show some similarity to those in HRGPs from other green algae. For example, *Chlamydomonas reinhardtii* HRGPs have been shown to contain two arabinoses with a terminal galactose [72] like that found on PSAP (Fig. 4C). However, in *C. reinhardtii* the galactose exists in the furanose configuration, while in PSAP the galactose is in the pyranose configuration (Fig. 4C). PSAP is further differentiated from the *C. reinhardtii* O-glycans by containing a 6-deoxyhexose (Fig. 4C). Short chain O-linked di- and tri-glycosylations containing arabinose and galactose have also been found on *V. carteri* HRGP ECM proteins with additional unique glycosylations containing two arabinose units connected by a phosphodiester bond [13,73–75].

#### 4.2. PSAP has unique N-linked glycans

Some of the N-glycan structures found in PSAP are unusual. One rare feature is the presence of an internal fucose residue, i.e. a fucose that is further glycosylated. Usually, fucose is found as a terminal decoration of the chitobiose core of N-glycans [76]. In mammalian glycoproteins, fucose is attached to the 6-position of the inner GlcNAc [77], while in plants and insects fucose is normally attached to the 3-position of the inner GlcNAc [78]. In fruit fly, fucose can also be attached to both the 3- and the 6-position of the inner GlcNAc [48]. Internal fucose has so far only been found in invertebrates, such as *C. elegans* [79], *planaria* [80], and squid [81]. In these cases, the fucose is substituted at O-4 with a galactose residue. We have found that O-4 of the core fucose in PSAP N-glycans is instead substituted by the pentose arabinose in the pyranose form (glycan 3 in Fig. 3A). While xylose, another pentose sugar, is common in plant N-glycans [57,78], arabinose has only been found in a few plant N-glycans [82], including carrot [83] and tomato [84]. A recent *B. braunii* N-glycoproteomic analysis [85] identified similar N-glycans to that presented here, but with the addition of methylated hexoses. Also, that study did not identify any PSAP peptides. In summary, while there are few reports presenting N-glycans with internal fucose or N-glycans with arabinose, to our knowledge, this is the first time an N-glycan with an arabinose-fucose disaccharide attached to the chitobiose core has been identified.

#### 4.3. Potential functional roles for PSAP

PSAP function is likely to be quite different from EXT proteins, which are involved in cell wall formation and expansion [14], since the PSAP protein is not associated with the cell wall, but rather the ECM. Thus, PSAP may be involved in the formation of the retaining wall or the polysaccharide fibers that extend from the retaining wall. When these studies began, we hypothesized PSAP was the enzyme responsible for forming the polysaccharide fibers by polymerizing sugar monomers

or oligomers. This is supported by our finding that a paste-like substance, presumably sugar monomers/oligomers, were found to enter the cell side of the retaining wall near the PSAP protein granules and exit the protein granule layer as the polysaccharide fibers [12]. It is also possible PSAP does not have enzymatic activity and functions as an anchor for the polysaccharide fibers. Given that PSAP is the only protein associated with the shell material it may serve both roles. Many attempts to obtain purified PSAP protein for enzymatic studies were not successful. Purification of PSAP from shell material without a denaturing environment (i.e. no SDS) or attempts at expressing PSAP in heterologous hosts such as *E. coli*, *Saccharomyces cerevisiae*, *Nicotiana benthamiana*, and *Chlamydomonas reinhardtii* were unsuccessful.

This suggests PSAP may contain intrinsically disordered regions possibly due to the Hyp-rich motif and the Pro-rich regions, which is another characteristic of HRGPs [66]. The PSAP disordered regions may be leading to difficulties in heterologous expression. Indeed, an analysis of PSAP for disordered regions found that the Hyp-rich motif and the two Pro-rich regions are predicted to be highly disordered (Supplementary Fig. 13).

Thus, identifying a specific function for PSAP will require refinements in purifying PSAP from *B. braunii* shell material or in recombinant expression. Understanding PSAP function through knock-down, knockout, or overexpression in *B. braunii* is not possible since *B. braunii* transformation is not possible at this time.

#### Author contributions

M.T., M.I., D.R.B., S.V., and L.J.D. performed the experiments. M.T., M.I., C.H., S.V., L.J.D., P.A., and T.P.D. planned the experiments and analyzed the data. M.T., M.I., C.H., P.A., and T.P.D. wrote the paper. All authors approved final version of the manuscript.

#### Acknowledgments

This work was supported by the National Science Foundation-EFRI-PSBR [grant number 1240478 to T.P.D.], the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, U.S. Department of Energy [grant number DE-SC0015662 to P.A.], and the National Institutes of Health [grant numbers 1S10OD018530 and P41GM103490-14 to P.A.]. We thank Lanying Zeng and Jimmy Trinh, Texas A&M University, for use of and instruction on the Nikon Eclipse epifluorescence microscope.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.algal.2017.11.018>.

#### References

- [1] G. Michel, T. Tonon, D. Scornet, J.M. Cock, B. Kloareg, The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes, *New Phytol.* 188 (2010) 82–97.
- [2] C. Brownlee, Role of the extracellular matrix in cell-cell signalling: paracrine paradigms, *Curr. Opin. Plant Biol.* 5 (2002) 396–401.
- [3] D.S. Domozych, M. Ciancia, J.U. Fangel, M.D. Mikkelsen, P. Ulvskov, W.G. Willats, The cell walls of green algae: a journey through evolution and diversity, *Front. Plant Sci.* 3 (2012) 82.
- [4] D.S. Domozych, C.E. Domozych, Multicellularity in green algae: upsizing in a walled complex, *Front. Plant Sci.* 5 (2014) 649.
- [5] B. Zhong, L. Liu, Z. Yan, D. Penny, Origin of land plants using the multispecies coalescent model, *Trends Plant Sci.* 18 (2013) 492–495.
- [6] B. Zhong, Z. Xi, V.V. Goremykin, R. Fong, P.A. McLenachan, P.M. Novis, C.C. Davis, D. Penny, Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes, *Mol. Biol. Evol.* 31 (2014) 177–183.
- [7] A.W. Coleman, A comparative analysis of the Volvocaceae (Chlorophyta), *J. Phycol.* 48 (2012) 491–513.
- [8] P. Metzger, C. Largeau, *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids, *Appl. Microbiol. Biotechnol.* 66 (2005) 486–496.

- [9] M. Sumper, A. Hallmann, Biochemistry of the extracellular matrix of *Volvox*, *Int. Rev. Cytol.* 180 (1998) 51–85.
- [10] G. Matt, J. Umen, *Volvox*: a simple algal model for embryogenesis, morphogenesis and cellular differentiation, *Dev. Biol.* 419 (2016) 99–113.
- [11] J.G. Umen, Green algae and the origins of multicellularity in the plant kingdom, *Cold Spring Harb. Perspect. Biol.* 6 (2014) a016170.
- [12] T.L. Weiss, R. Roth, C. Goodson, S. Vitha, I. Black, P. Azadi, J. Rusch, A. Holzenburg, T.P. Devarenne, U. Goodenough, Colony organization in the green alga *Botryococcus braunii* (Race B) is specified by a complex extracellular matrix, *Eukaryot. Cell* 11 (2012) 1424–1440.
- [13] F. Ender, K. Godl, S. Wenzl, M. Sumper, Evidence for autocatalytic cross-linking of hydroxyproline-rich glycoproteins during extracellular matrix assembly in *Volvox*, *Plant Cell* 14 (2002) 1147–1160.
- [14] D.T. Lampport, M.J. Kieliszewski, Y. Chen, M.C. Cannon, Role of the extensin superfamily in primary cell wall architecture, *Plant Physiol.* 156 (2011) 11–19.
- [15] M. Hijazi, S.M. Velasquez, E. Jamet, J.M. Estevez, C. Albenne, An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture, *Front. Plant Sci.* 5 (2014) 395.
- [16] M.J. Kieliszewski, D.T. Lampport, Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny, *Plant J.* 5 (1994) 157–172.
- [17] M.J. Kieliszewski, E. Shpak, Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins, *Cell. Mol. Life Sci.* 58 (2001) 1386–1398.
- [18] A. Banerjee, R. Sharma, Y. Chisti, U.C. Banerjee, *Botryococcus braunii*: a renewable source of hydrocarbons and other chemicals, *Crit. Rev. Biotechnol.* 22 (2002) 245–279.
- [19] P. Metzger, C. Berkaloff, E. Casadevall, A. Coute, Alkadiene-producing and botryococcene-producing races of wild strains of *Botryococcus braunii*, *Phytochemistry* 24 (1985) 2305–2312.
- [20] P. Metzger, J. Templier, C. Largeau, E. Casadevall, An *n*-alkatriene and some *n*-alkadienes from the A race of the green alga *Botryococcus braunii*, *Phytochemistry* 25 (1986) 1869–1872.
- [21] J. Templier, C. Largeau, E. Casadevall, Hydrocarbon formation in the green-alga *Botryococcus braunii*. 4. Mechanism of non-isoprenoid hydrocarbon biosynthesis in *Botryococcus braunii*, *Phytochemistry* 23 (1984) 1017–1028.
- [22] J. Templier, C. Largeau, E. Casadevall, Biosynthesis of *n*-Alkatrienes in *Botryococcus braunii*, *Phytochemistry* 30 (1991) 2209–2215.
- [23] P. Metzger, E. Casadevall, M.J. Pouet, Y. Pouet, Structures of some botryococcenes: branched hydrocarbons from the B-race of the green alga *Botryococcus braunii*, *Phytochemistry* 24 (1985) 2995–3002.
- [24] P. Metzger, M. David, E. Casadevall, Biosynthesis of triterpenoid hydrocarbons in the B-race of the green alga *Botryococcus braunii*. Sites of production and nature of the methylating agent, *Phytochemistry* 26 (1987) 129–134.
- [25] P. Metzger, E. Casadevall, A. Couté, Botryococcene distribution in strains of the green alga *Botryococcus braunii*, *Phytochemistry* 27 (1988) 1383–1388.
- [26] P. Metzger, E. Casadevall, Lycopadiene, a tetraterpenoid hydrocarbon from new strains of the green alga *Botryococcus braunii*, *Tetrahedron Lett.* 28 (1987) 3931–3934.
- [27] P. Metzger, B. Allard, E. Casadevall, C. Berkaloff, A. Coute, Structure and chemistry of a new chemical race of *Botryococcus braunii* (Chlorophyceae) that produces lycopadiene a tetraterpenoid hydrocarbon, *J. Phycol.* 26 (1990) 258–266.
- [28] H.R. Thapa, M.T. Naik, S. Okada, K. Takada, I. Molnar, Y. Xu, T.P. Devarenne, A squalene synthase-like enzyme initiates production of tetraterpenoid hydrocarbons in *Botryococcus braunii* race L, *Nat. Commun.* 7 (2016) 11198.
- [29] P. Metzger, M.N. Rager, C. Largeau, Polyacetals based on polymethylsqualene diols, precursors of algaenan in *Botryococcus braunii* race B, *Org. Geochem.* 38 (2007) 566–581.
- [30] O. Bertheas, P. Metzger, C. Largeau, A high molecular weight complex lipid, aliphatic polyaldehyde tetraterpenediol polyacetal from *Botryococcus braunii* (L race), *Phytochemistry* 50 (1999) 85–86.
- [31] P. Metzger, Y. Pouet, R. Bischoff, E. Casadevall, An aliphatic polyaldehyde from *Botryococcus braunii* (A race), *Phytochemistry* 32 (1993) 875–883.
- [32] P. Metzger, M.N. Rager, C. Fosse, Braunicetals: acetals from condensation of macrocyclic aldehydes and terpene diols in *Botryococcus braunii*, *Phytochemistry* 69 (2008) 2380–2386.
- [33] C. Berkaloff, B. Rousseau, A. Coute, E. Casadevall, P. Metzger, C. Chirac, Variability of cell-wall structure and hydrocarbon type in different strains of *Botryococcus braunii*, *J. Phycol.* 20 (1984) 377–389.
- [34] C. Largeau, E. Casadevall, C. Berkaloff, P. Dhamelincourt, Sites of accumulation and composition of hydrocarbons in *Botryococcus braunii*, *Phytochemistry* 19 (1980) 1043–1051.
- [35] F.R. Wolf, E.R. Cox, Ultrastructure of active and resting colonies of *Botryococcus braunii* (Chlorophyceae), *J. Phycol.* 17 (1981) 395–405.
- [36] Y. Uno, I. Nishii, S. Kagiwada, T. Noguchi, Colony sheath formation is accompanied by shell formation and release in the green alga *Botryococcus braunii* (race B), *Algal Res.* 8 (2015) 214–223.
- [37] I. Molnar, D. Lopez, J.H. Wisecaver, T.P. Devarenne, T.L. Weiss, M. Pellegrini, J.D. Hackett, Bio-crude transcriptomics: gene discovery and metabolic network reconstruction for the biosynthesis of the terpene of the hydrocarbon oil-producing green alga, *Botryococcus braunii* race B (Showa), *BMC Genomics* 13 (2012) 576.
- [38] D.R. Browne, J. Jenkins, J. Schmutz, S. Shu, K. Barry, J. Grimwood, J. Chiniqay, A. Sharma, T.D. Niehaus, T.L. Weiss, A.T. Koppisch, D.T. Fox, S. Dhungana, S. Okada, J. Chappell, T.P. Devarenne, Draft nuclear genome sequence of the liquid hydrocarbon-accumulating green microalga *Botryococcus braunii* race B (Showa), *Genome Announc.* 5 (2017) e00215–00217.
- [39] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, *Bioinformatics* 28 (2012) R36.
- [40] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms, *Nat. Biotechnol.* 28 (2010) 511–515.
- [41] M. Pagni, V. Ioannidis, L. Cerutti, M. Zahn-Zabal, C.V. Jongeneel, J. Hau, O. Martin, D. Kuznetsov, L. Falquet, MyHits: improvements to an interactive resource for analyzing protein sequences, *Nucleic Acids Res.* 35 (2007) W433–437.
- [42] J.D. Bendtsen, H. Nielsen, D. Widdick, T. Palmer, S. Brunak, Prediction of twin-arginine signal peptides, *BMC Bioinform.* 6 (2005) 167.
- [43] L. Kall, A. Krogh, E.L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method, *J. Mol. Biol.* 338 (2004) 1027–1036.
- [44] S. Briesemeister, J. Rahnenfuhrer, O. Kohlbacher, Going from where to why—interpretable prediction of protein subcellular localization, *Bioinformatics* 26 (2010) 1232–1238.
- [45] S. Briesemeister, J. Rahnenfuhrer, O. Kohlbacher, YLoc—an interpretable web server for predicting subcellular localization, *Nucleic Acids Res.* 38 (2010) W497–502.
- [46] L. Chill, L. Trinh, P. Azadi, M. Ishihara, R. Sonon, E. Karnaukhova, Y. Ophir, B. Golding, J. Shiloach, Production, purification, and characterization of human  $\alpha$ 1 proteinase inhibitor from *Aspergillus niger*, *Biotechnol. Bioeng.* 102 (2009) 828–844.
- [47] P.M. Angel, J.M. Lim, L. Wells, C. Bergmann, R. Orlando, A potential pitfall in  $^{18}\text{O}$ -based *N*-linked glycosylation site mapping, *Rapid Commun. Mass Spectrom.* 21 (2007) 674–682.
- [48] K. Aoki, M. Perlman, J.M. Lim, R. Cantu, L. Wells, M. Tiemeyer, Dynamic developmental elaboration of *N*-linked glycan complexity in the *Drosophila melanogaster* embryo, *J. Biol. Chem.* 282 (2007) 9127–9142.
- [49] L. Ma, Z. Chen, W. da Huang, G. Kutty, M. Ishihara, H. Wang, A. Abouelleil, L. Bishop, E. Davey, R. Deng, X. Deng, L. Fan, G. Fantoni, M. Fitzgerald, E. Gogineni, J.M. Goldberg, G. Handley, X. Hu, C. Huber, X. Jiao, K. Jones, J.Z. Levin, Y. Liu, P. Macdonald, A. Melnikov, C. Raley, M. Sassi, B.T. Sherman, X. Song, S. Sykes, B. Tran, L. Walsh, Y. Xia, J. Yang, S. Young, Q. Zeng, X. Zheng, R. Stephens, C. Nusbaum, B.W. Birren, P. Azadi, R.A. Lempicki, C.A. Cuomo, J.A. Kovacs, Genome analysis of three *Pneumocystis* species reveals adaptation mechanisms to life exclusively in mammalian hosts, *Nat. Commun.* 7 (2016) 10740.
- [50] K.R. Anumula, P.B. Taylor, A comprehensive procedure for preparation of partially methylated alditol acetates from glycoprotein carbohydrates, *Anal. Biochem.* 203 (1992) 101–108.
- [51] C.M. Woo, A. Felix, W.E. Byrd, D.K. Zuegel, M. Ishihara, P. Azadi, A.T. Iavarone, S.J. Pitteri, C.R. Bertozzi, Development of IsoTaG, a chemical glycoproteomics technique for profiling intact *N*- and *O*-glycopeptides from whole cell proteomes, *J. Proteome Res.* 16 (2017) 1706–1718.
- [52] C.P. Mercado, M.V. Quintero, Y. Li, P. Singh, A.K. Byrd, K. Talabnin, M. Ishihara, P. Azadi, N.J. Rusch, B. Kuberan, L. Maroteaux, F. Kilic, A serotonin-induced *N*-glycan switch regulates platelet aggregation, *Sci. Rep.* 3 (2013) 2795.
- [53] M. Ghahremani, K.A. Stigter, W. Plaxton, Extraction and characterization of extracellular proteins and their post-translational modifications from *Arabidopsis thaliana* suspension cell cultures and seedlings: a critical review, *Proteome* 4 (2016) 25.
- [54] N. Moreno, S. Bougourd, J. Haseloff, J.A. Feijó, Imaging plant cells, in: J.B. Pawley (Ed.), *Handbook of Biological Confocal Microscopy*, Springer Science and Business Media, New York, 2006, pp. 769–787.
- [55] T.P. Devarenne, S.K. Ekengren, K.F. Pedley, G.B. Martin, Adi3 is a Pdk1-interacting AGC kinase that negatively regulates plant cell death, *EMBO J.* 25 (2006) 255–265.
- [56] J.L. Mellquist, L. Kasturi, S.L. Spitalnik, S.H. Shakin-Eshleman, The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of *N*-linked core glycosylation efficiency, *Biochemistry* 37 (1998) 6833–6837.
- [57] E. Nyguema-Ona, M. Vre-Gibouin, M. Gotte, B. Plancot, P. Lerouge, M. Bardor, A. Driouch, Cell wall *O*-glycoproteins and *N*-glycoproteins: aspects of biosynthesis and function, *Front. Plant Sci.* 5 (2014) 499.
- [58] R.G. Creasey, N.H. Voelcker, C.J. Schultz, Investigation of self-assembling proline- and glycine-rich recombinant proteins and peptides inspired by proteins from a symbiotic fungus using atomic force microscopy and circular dichroism spectroscopy, *Biochim. Biophys. Acta* 1824 (2012) 711–722.
- [59] G. Gutierrez-Cruz, A.H. Van Heerden, K. Wang, Modular motif, structural folds and affinity profiles of the PEVK segment of human fetal skeletal muscle titin, *J. Biol. Chem.* 276 (2001) 7442–7449.
- [60] H. Kawashima, S. Sueyoshi, H. Li, K. Yamamoto, T. Osawa, Carbohydrate binding specificities of several poly-*N*-acetylglucosamine-binding lectins, *Glycoconj. J.* 7 (1990) 323–334.
- [61] R.A. Naidu, C.J. Ingle, C.M. Deom, J.L. Sherwood, The two envelope membrane glycoproteins of *Tomato spotted wilt virus* show differences in lectin-binding properties and sensitivities to glycosidases, *Virology* 319 (2004) 107–117.
- [62] B. Allard, E. Casadevall, Carbohydrate-composition and characterization of sugars from the green microalga *Botryococcus-Braunii*, *Phytochemistry* 29 (1990) 1875–1878.
- [63] H.L. Fernandes, M.M. Tome, F.M. Lupi, A.M. Fialho, I. Sá-Correia, J.M. Novais, Biosynthesis of high concentrations of an exopolysaccharide during the cultivation of the microalga *Botryococcus braunii*, *Biotechnol. Lett.* 11 (1989) 433–436.
- [64] F.M. Lupi, H.L. Fernandes, I. Sá-Correia, J.M. Novais, Temperature profiles of cellular growth and exopolysaccharide synthesis by *Botryococcus braunii* Kütz. UC 58, *J. Appl. Phycol.* 3 (1991) 35–42.
- [65] F.M. Lupi, H.L. Fernandes, M.M. Tome, I. Sá-Correia, J.M. Novais, Influence of

- nitrogen source and photoperiod on exopolysaccharide synthesis by the microalga *Botryococcus braunii* UC 58, *Enzym. Microb. Tech.* 16 (1994) 546–550.
- [66] K.L. Johnson, A.M. Cassin, A. Lonsdale, A. Bacic, M.S. Doblin, C.J. Schultz, A motif and amino acid bias bioinformatics pipeline to identify hydroxyproline-rich glycoproteins, *Plant Physiol.* 174 (2017) 886–903.
- [67] K.L. Johnson, A.M. Cassin, A. Lonsdale, G.K. Wong, D.E. Soltis, N.W. Miles, M. Melkonian, B. Melkonian, M.K. Deyholos, J. Leebens-Mack, C.J. Rothfels, D.W. Stevenson, S.W. Graham, X. Wang, S. Wu, J.C. Pires, P.P. Edger, E.J. Carpenter, A. Bacic, M.S. Doblin, C.J. Schultz, Insights into the evolution of hydroxyproline-rich glycoproteins from 1000 plant transcriptomes, *Plant Physiol.* 174 (2017) 904–921.
- [68] P.B. Kavi Kishor, P. Hima Kumari, M.S. Sunita, N. Sreenivasulu, Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny, *Front. Plant Sci.* 6 (2015) 544.
- [69] M.C. Cannon, K. Terneus, Q. Hall, L. Tan, Y. Wang, B.L. Wegenhart, L. Chen, D.T. Lampion, Y. Chen, M.J. Kieliszewski, Self-assembly of the plant cell wall requires an extensin scaffold, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 2226–2231.
- [70] S.W. Chan, J. Greaves, N.A. Da Silva, S.W. Wang, Assaying proline hydroxylation in recombinant collagen variants by liquid chromatography-mass spectrometry, *BMC Biotechnol.* 12 (2012) 51.
- [71] M. Hijazi, D. Roujol, H. Nguyen-Kim, L. Del Rocio Cisneros Castillo, E. Saland, E. Jamet, C. Albenne, Arabinogalactan protein 31 (AGP31), a putative network-forming protein in *Arabidopsis thaliana* cell walls? *Ann. Bot.* 114 (2014) 1087–1097.
- [72] K. Bollig, M. Lamshoft, K. Schweimer, F.J. Marner, H. Budzikiewicz, S. Waffenschmidt, Structural analysis of linear hydroxyproline-bound O-glycans of *Chlamydomonas reinhardtii*—conservation of the inner core in *Chlamydomonas* and land plants, *Carbohydr. Res.* 342 (2007) 2557–2566.
- [73] K. Godl, A. Hallmann, S. Wenzl, M. Sumper, Differential targeting of closely related ECM glycoproteins: the pterophorin family from *Volvox*, *EMBO J.* 16 (1997) 25–34.
- [74] O. Holst, V. Christoffel, R. Frund, H. Moll, M. Sumper, A phosphodiester bridge between two arabinose residues as a structural element of an extracellular glycoprotein of *Volvox carteri*, *Eur. J. Biochem.* 181 (1989) 345–350.
- [75] H. Ertl, R. Mengele, S. Wenzl, J. Engel, M. Sumper, The extracellular matrix of *Volvox carteri*: molecular structure of the cellular compartment, *J. Cell Biol.* 109 (1989) 3493–3501.
- [76] B. Schiller, A. Hykollari, S. Yan, K. Paschinger, I.B. Wilson, Complicated N-linked glycans in simple organisms, *Biol. Chem.* 393 (2012) 661–673.
- [77] M. Takahashi, Y. Kuroki, K. Ohtsubo, N. Taniguchi, Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins, *Carbohydr. Res.* 344 (2009) 1387–1390.
- [78] R. Strasser, Plant protein glycosylation, *Glycobiology* 26 (2016) 926–939.
- [79] T. Takeuchi, K. Hayama, J. Hirabayashi, K. Kasai, *Caenorhabditis elegans* N-glycans containing a Gal-Fuc disaccharide unit linked to the innermost GlcNAc residue are recognized by *C. elegans* galectin LEC-6, *Glycobiology* 18 (2008) 882–890.
- [80] K. Paschinger, E. Razzazi-Fazeli, K. Furukawa, I.B. Wilson, Presence of galactosylated core fucose on N-glycans in the planaria *Dugesia japonica*, *J. Mass Spectrom.* 46 (2011) 561–567.
- [81] N. Takahashi, K. Masuda, K. Hiraki, K. Yoshihara, H.H. Huang, K.H. Khoo, K. Kato, N-glycan structures of squid rhodopsin, *Eur. J. Biochem.* 270 (2003) 2627–2632.
- [82] J.P. Utille, B. Priem, Synthesis of allyl 2-O-( $\alpha$ -L-arabinofuranosyl)-6-O-( $\alpha$ -D-mannopyranosyl)- $\beta$ -D-mannopyranosid e, a unique plant N-glycan motif containing arabinose, *Carbohydr. Res.* 329 (2000) 431–439.
- [83] A. Sturm, A wound-inducible glycine-rich protein from *Daucus carota* with homology to single-stranded nucleic acid-binding proteins, *Plant Physiol.* 99 (1992) 1689–1692.
- [84] B. Priem, R. Gitti, C.A. Bush, K.C. Gross, Structure of ten free N-glycans in ripening tomato fruit. Arabinose is a constituent of a plant N-glycan, *Plant Physiol.* 102 (1993) 445–458.
- [85] S. Schulze, E. Urzica, M. Reijnders, H. van de Geest, S. Warris, L.V. Bakker, C. Fufezan, V.A.P. Martins Dos Santos, P.J. Schaap, S.A. Peters, M. Hippler, Identification of methylated GnTI-dependent N-glycans in *Botryococcus braunii*, *New Phytol.* (2017), <http://dx.doi.org/10.1111/nph.14713>.