Algal Research 29 (2018) 92–103

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

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

Mehmet Tatli[a], Mayumi Ishihara[b], Christian Heiss[b], Daniel R. Browne[a], Lawrence J. Dangott[a,c], Stanislav Vitha[d], Parastoo Azadi[b], Timothy P. Devarenne[a,*]

[a] Department of Biochemistry & Biophysics, Texas A&M University, College Station, TX 77843, United States
[b] Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, United States
[c] Protein Chemistry Lab, Department of Biochemistry & Biophysics, Texas A&M University, College Station, TX 77843, United States
[d] 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 (T.P. Devarenne).

https://doi.org/10.1016/j.algal.2017.11.018
Received 1 August 2017; Received in revised form 20 October 2017; Accepted 18 November 2017

(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 botryococcenes 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 μm long polysaccharide fibers that completely circumscribe the colony [12,33–35]. These fibers consist mainly of galactose and arabinose with uncommonly abundant (1 → 2), (1 → 3) and (1 → 2), (1 → 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 μl of extraction buffer (1 mM NaCl, 10 mM Tris, pH 6.8, 1 × protease inhibitors) and 200 μl of 2 × SDS-PAGE sample buffer to the shell material, and placing the sample at 95 °C for 5 min. A 4–15% gradient SDS-PAGE gel was used to separate 20 μl 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 ~1 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 °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 °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 μl of PNGase A (Calbiochem) in 36 μl of ^{18}O-labled water ($H_2{}^{18}O$) and 2 μl of 0.5 M citrate phosphate buffer pH 5.0. The sample was dried and resuspended in nanopure dH_2O with protease(s) to remove any possible C-terminal incorporation of ^{18}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)