Carbohydrate Research 352 (2012) 70–81

Contents lists available at SciVerse ScienceDirect

Carbohydrate Research

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

An *exo*-β-(1→3)-D-galactanase from *Streptomyces* sp. provides insights into type II arabinogalactan structure

Naomi X.-Y. Ling [a], Joanne Lee [a], Miriam Ellis [a], Ming-Long Liao [a], Shaio-Lim Mau [a], David Guest [c], Peter H. Janssen [d], Pavol Kováč [e], Antony Bacic [a,b], Filomena A. Pettolino [a,*]

[a] Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Victoria 3010, Australia
[b] ARC Centre of Excellence in Plant Cell Walls, School of Botany, University of Melbourne, Victoria 3010, Australia
[c] Faculty of Agriculture, Food and Natural Resources, Biomedical Building C81, The University of Sydney, Eveleigh, NSW 2015, Australia
[d] Grasslands Research Centre, AgResearch Ltd, Tennent Drive, Private Bag 11008, Palmerston North 4442, New Zealand
[e] Laboratory of Bioorganic Chemistry, Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health, Bethesda, MD 20892-0815, USA

ARTICLE INFO

Article history:
Received 21 December 2011
Received in revised form 21 February 2012
Accepted 28 February 2012
Available online 8 March 2012

Keywords:
β-D-Galactanases
Arabinogalactan-protein
Streptomyces sp.
De-arabinosylated gum arabic
CAZy family GH 43

ABSTRACT

An *exo*-β-(1→3)-D-galactanase (SGalase1) that specifically cleaves the β-(1→3)-D-galactan backbone of arabinogalactan-proteins (AGPs) was isolated from culture filtrates of a soil *Streptomyces* sp. Internal peptide sequence information was used to clone and recombinantly express the gene in *E. coli*. The molecular mass of the isolated enzyme was ~45 kDa, similar to the 48.2 kDa mass predicted from the amino acid sequence. The pI, pH and temperature optima for the enzyme were ~7.45, 3.8 and 48 °C, respectively. The native and recombinant enzymes specifically hydrolysed β-(1→3)-D-galacto-oligo- or poly-saccharides from the upstream (non-reducing) end, typical of an *exo*-acting enzyme. A second homologous *Streptomyces* gene (*SGalase2*) was also cloned and expressed. SGalase2 was similar in size (47.9 kDa) and enzyme activity to SGalase1 but differed in its pH optimum (pH 5). Both SGalase1 and SGalase2 are predicted to belong to the CAZy glycosyl hydrolase family GH 43 based on activity, sequence homology and phylogenetic analysis. The K_m and V_{max} of the native *exo*-β-(1→3)-D-galactanase for de-arabinosylated gum arabic (dGA) were 19 mg/ml and 9.7 μmol D-Gal/min/mg protein, respectively. The activity of these enzymes is well suited for the study of type II galactan structures and provides an important tool for the investigation of the biological role of AGPs in plants. De-arabinosylated gum arabic (dGA) was used as a model to investigate the use of these enzymes in defining type II galactan structure. Exhaustive hydrolysis of dGA resulted in a limited number of oligosaccharide products with a trisaccharide of Gal_2GlcA_1 predominating.

1. Introduction

Arabinogalactan-proteins (AGPs) are a class of plant proteoglycans containing 90–99% (w/w) carbohydrate and 1–10% (w/w) protein. The predominant carbohydrate components of AGPs are type II arabinogalactans (AGs)[1] together with some short oligoarabinosides.[2,3] Type II AGs consist of a highly branched and complex framework of a β-(1→3)-D-galactan backbone and β-(1→6)-D-galactopyranosyl (Gal*p*) side chains, which are further substituted with other sugars such as α-L-arabinofuranose/pyranose (Ara*f*/*p*), glucopyranosyluronic acid (Glc*p*A), 4-*O*-methylglucopyranosyluronic acid (4-*O*-MeGlc*p*A), L-rhamnopyranose (L-Rha*p*) or L-fucopyranose (L-Fuc*p*).[4–7] Numerous studies have investigated the structures of the carbohydrate component of AGPs from various plants, for example *Arabidopsis*,[8] gum arabic (GA),[9–11] gum tragacanth,[12] larch,[13] apple,[14] grape,[15] radish[16,17] and ryegrass.[18] Recently, the Hyp-contiguity hypothesis has also contributed to the better understanding of the glycosylation pattern of the protein backbones of AGPs. The Hyp-contiguity hypothesis predicts that non-contiguous Hyp residues would be exclusively substituted with AG polysaccharide chains, while contiguous Hyp residues would be glycosylated with oligoarabinosides.[19–22] This hypothesis was generated by studying heterologously-expressed synthetic AGP protein backbones in tobacco and *Arabidopsis*.[23] Hyp glycosylation may be tissue-specific and vary between species.[22–24]

Defining the precise structure of the carbohydrate moiety of AGPs is essential in furthering our understanding of their biosynthesis, as well as their biological and industrial functions. AGPs are proposed to be responsible for a range of functions in plants.[7] For example, AGPs are proposed to play a role in signalling,[25,26] embryogenesis,[27,28] xylem differentiation,[26,29] as well as cell expansion and extension of apical tip growth in the moss *Physcomitrella patens*.[30] The general structure of type II AG is well known, but its glycosyl

* Corresponding author at present address. CSIRO Plant Industry, Black Mountain Laboratories, Canberra ACT 2601, Australia. Tel.: +61 02 62464052; fax: +61 02 62465000.
E-mail address: Filomena.Pettolino@csiro.au (F.A. Pettolino).

0008-6215/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.carres.2012.02.033

sequences are yet to be elucidated, particularly the distribution and degree of polymerisation (DP) of (1→3)- and (1→6)-linked Gal*p* chains.[31] Chemical methods such as partial acid hydrolysis, alkaline degradation, acetolysis and Smith degradation have been invaluable in the analysis of the type II glycan chains of AGPs.[15,18,32] However, due to the complexity of the carbohydrate moiety, chemical approaches alone are not sufficient.[33] Therefore, specific enzymes, such as α-L-arabinofuranosidases (EC 3.2.1.55),[14,15,34] β-glucuronidases (EC 3.2.1.31),[35] and β-D-galactosidases (EC 3.2.1.23),[34,36] *exo*-β-1,3-D-galactanases (EC 3.2.1.145)[32,37] and *endo*-β-(1→6)-D-galactanases (EC 3.2.1.164)[31,38,39] have been used to complement the chemical methods. Most of these enzymes are not commercially available and as a result, have only been used in a limited number of studies. This is especially true of the galactanases that cleave β-(1→3)- or (1→6)-D-Gal*p* linkages in either an *exo*- or *endo*-manner. Furthermore, there is only one report to date of the isolation of an *endo*-β-(1→3)-D-galactanase.[40] Enzymes may have different modes of action and/or catalytic sites. It is therefore important to isolate a wide spectrum of hydrolytic enzymes that cleave specific glycosyl linkages of the type II AGs. In the current study, an *exo*-β-(1→3)-D-galactanase was isolated from culture filtrates of a *Streptomyces* sp. isolated from soil and the kinetic properties of the enzyme determined. This enzyme, and a second similar enzyme from the same source, were cloned and expressed in *E. coli* and their activity characterised. We also demonstrate how the activity of these enzymes is well suited for the study of type II galactan structure.

2. Results

2.1. Native enzyme purification

Isolate 19 was selected from 26 cultures for its ability to grow on medium supplemented with dGA and secrete protein with β-galactanase activity. The 16S rRNA gene sequence revealed isolate 19 to be a *Streptomyces* sp. related to a clade of species including *S. achromogenes*, *S. acidiscabies*, *S. antibioticus*, *S. bobili*, *S. coelicolor*, *S. cyaneus*, *S. flavovariabilis*, *S. galilaeus*, *S. griseochromogenes*, *S. griseoruber*, *S. longisporus*, *S. mirabilis*, and *S. phaeofaciens* (data not shown).

The culture filtrate (day 3) of *Streptomyces* sp. isolate 19 was used to obtain β-galactanase by precipitating the proteins using ammonium sulfate (crude enzyme preparation) followed by ion exchange and gel filtration chromatography (Table 1). Most of the unwanted β-D-galactosidase activity (~90%) in the crude enzyme preparation was removed by binding to DEAE Trisacryl (M) at pH 8 (data not shown). The remaining DEAE unbound fraction separated into four major protein peaks by cation exchange chromatography (Fig. 1). Most (89%) of the β-galactanase activity eluted as a single peak (EMD Peak 2, Fig. 1) in the range of 0.1–0.15 M NaCl, while most of the α-L-arabinofuranosidase and remaining β-D-galactosidase activities eluted at or above 0.75 M NaCl, well separated from the region associated with β-galactanase activity. EMD Peak 2 was further fractionated by size on a TSK HW 55 column and the β-galactanase activity eluted as a major peak, corresponding to a molecular mass of ~43 kDa (data not shown). The yield and the overall purification factor of the β-galactanase achieved in the current study were 3.3% and ~2-fold, respectively (Table 1). The enriched β-galactanase fraction, HW 55 (major peak), consisted of one major band at ~45 kDa, one minor band at ~66 kDa and three other protein bands between 25 and 40 kDa in very low abundance (Fig. 2, Lane 5).

2.2. Galactanase cloning and heterologous expression

The ~45 kDa protein band was excised from the gel for N-terminal sequencing and internal sequencing post in-gel trypsin digestion. The N-terminal peptide sequence (A S A S F T L G A T Y T D Q N) (Fig. 2 Lane 5) was compared to the NCBI non-redundant protein databases using the BLASTp algorithm. Thorough searches for short, nearly exact matches found no significant matches. However, the first 10 amino acids of the N-terminal peptide sequence were the same as one of five trypsin generated peptides (highlighted in Fig. 3). The de novo amino acid sequences of the remaining four trypsin-generated peptides (also in Fig. 3) from the same protein band had a relatively high similarity to a number of proteins across different regions of the amino acid sequences, with the top hit in the NCBI BLAST search being *exo*-β-(1→3)-D-galactanase sequences from the CAZy database (http://www.cazy.org/CAZY). These amino acid sequences were from *Phanerochaete chrysosporium* galactan (1→3)-β-galactosidase (Pc1,3Gal43A, gi 63108312), *Coprinopsis cinerea* hypothetical protein (CC1G_03773, gi 116510091), *Streptomyces avermitilis exo*-β-(1→3)-D-galactanase (Sal1,3Gal43A, gi 29828651) and *Clostridium thermocellum exo*-β-(1→3)-D-galactanase (Ct1,3Gal43A, gi 67916108). Commonly conserved features of these sequences include regions showing sequence similarity to CAZy glycosyl hydrolase family 43 (GH 43) and ricin carbohydrate-binding domains.

Using degenerate primers based on the N-terminal and internal peptide sequences of the isolated protein, the *SGalase1* gene sequence was amplified and sequenced, along with another putative *Streptomyces* sp. galactanase gene, *SGalase2*. The predicted proteins encoded by the two genes share 85% amino acid sequence similarity (using mature protein sequence only) and a predicted size of ~48 kDa for the mature enzymes (Fig. 3). *SGalase2* has a predicted N-terminal signal peptide and conserved domain searches[41] found that both of the predicted proteins contain a GH 43 catalytic domain at the N-terminal end, and a ricin-type β-trefoil carbohydrate-binding domain at the C-terminal end (Fig. 3). Despite repeated attempts, the very 5′ end of *SGalase 1* proved difficult to obtain, however based on protein sequence analysis, the majority of the enzyme was cloned. Sequences for SGalase 1 and SGalase 2 were deposited in Genbank with accession numbers JQ683399 and JQ683400 respectively.

The amino acid sequences for the catalytic domains of the characterised GH43 enzymes were aligned with those of SGalase1 and SGalase2, and the program MEGA 5 was used to generate a Neighbor-Joining tree (Fig. 4). The tree shows, with high bootstrap support, that the two SGalase enzymes are more similar to each other than to any of the other enzymes, and that they are nested within the other *exo*-β-(1→3)-galactanases.

Table 1
Protein content and enzyme activity of different fractions obtained in the purification of β-galactanase from filtrate of 3 day-old culture of *Streptomyces* sp.

Fraction	Total protein (mg)	Protein yield (%)	Total activity[a] (Units)	Specific activity (Units/mg protein)	Yield factor (%)	Purification (fold)
$(NH_4)_2SO_4$ ppt (Crude enzyme)	323	100	6975	22	100	1
DEAE Trisacryl (Unbound)	111	34	3720	34	53	1.5
Fractogel EMD SO_3^- (EMD Peak 2)	7	2.2	279	40	4	1.8
Fractogel TSK HW 55 (Major peak)	5.86	1.8	233	40	3.3	1.8

[a] Units = One unit (U) of β-galactanase activity is defined as the amount of enzyme that liberates 1 nmol of low molecular weight substrate fragments bound with RB5 dye per minute under the assay conditions described.