FEBS Letters 581 (2007) 1811–1818

The C-terminal sequence from common bugle leaf galactan:galactan galactosyltransferase is a non-sequence-specific vacuolar sorting determinant

Esther M. Tapernoux-Lüthi, Thomas Schneider, Felix Keller*

Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, CH-8008 Zurich, Switzerland

Received 9 February 2007; revised 23 March 2007; accepted 23 March 2007

Available online 4 April 2007

Edited by Ulf-Ingo Flügge

Abstract **The *Ajuga reptans* L. galactan:galactan galactosyltransferase (ArGGT) is a vacuolar enzyme that synthesizes long-chain raffinose family oligosaccharides (RFOs), the major storage carbohydrates of this plant. ArGGT is structurally and functionally related to acid plant α-galactosidases (α-Gals) of the glycosylhydrolase family 27, present in the apoplast or the vacuole. Sequence comparison of acid α-Gals with ArGGT revealed that they all contain an N-terminal signal sequence and a highly similar core sequence. Additionally, ArGGT and some acid α-Gals contain C-terminal extensions with low sequence similarities to each other. Here, we show that the C-terminal pentapeptide, SLQMS, is a non-sequence-specific vacuolar sorting determinant. Analogously, we demonstrate that the C-terminal extensions of selected acid α-Gals from Arabidopsis, barley, and rice, are also non-sequence-specific vacuolar sorting determinants, suggesting the presence of at least one vacuolar form of acid α-Gal in every plant species.**

Keywords: Protein targeting; Vacuolar sorting determinant; Vacuole; Tonoplast; α-Galactosidase

1. Introduction

Plant vacuoles may occupy up to 90% of a cell volume and fulfil many vital functions such as storage of water, ions, pigments, or toxic compounds, regulation of turgor pressure and synthesis, storage, and mobilization of metabolites. In plant cells, at least two types of vacuoles exist, the lytic vacuoles and the protein storage vacuoles (PSVs), which might coexist or fuse during development [1,2]. We have had a long-standing interest in a special type of vacuole which both synthesize, store, and hydrolyze water-soluble carbohydrates. Prominent examples of such vacuoles include fructan vacuoles and raffinose family oligosaccharide (RFO) vacuoles which we have shown to contain the respective carbohydrates as well as their anabolic and catabolic enzymes [3–5].

Recently, we characterized the vacuolar long-chain RFO synthesizing enzyme galactan:galactan galactosyltransferase (GGT) from common bugle (*Ajuga reptans*) both biochemically and molecularly [6–8]. ArGGT is a glycosylated protein that catalyzes the chain elongation of RFOs at low pH by transferring a terminal α-galactosyl residue from one RFO molecule to another one, producing RFOs with a degree of polymerization between 5 and 15 [6–8]. So far, GGTs have only been found in plants of the Lamiaceae family [9,10]. The location of ArGGT and its products, the long-chain RFOs, in the acid central vacuole was demonstrated by comparing protoplasts and vacuoles from mesophyll cells of cold-induced leaves [5,11]. It was also demonstrated that ArGGT is a soluble enzyme residing in the cell sap by comparing vacuoles with tonoplast vesicles and cell sap fractions isolated from them [11].

Vacuoles are part of the endomembrane system including the endoplasmic reticulum (ER), the Golgi apparatus, and the plasma membrane as main components. Polypeptides destined for the vacuole or the cell surface carry an N-terminal signal sequence that targets the nascent polypeptide to the lumen of the ER. This signal sequence is removed co-translationally [12]. The default pathway of polypeptides that have entered the ER is the secretion to the cell surface [13]. Retention in the endomembrane system or sorting to the vacuole requires additional information represented by short peptide sequences called ER retention signal and vacuolar sorting determinant (VSD), respectively (for review see [14–17]).

VSDs are thought to interact with protein receptors, which most probably initiate vesicle formation and transport to the vacuole. Sorting of polypeptides to vacuoles often requires transport through the Golgi apparatus; vesicle formation then occurs in the Golgi stacks after processing, e.g. modification by glycosylation [15–19]. Three different types of VSDs have been identified so far and are part of or constitute sequence-specific, usually N-terminal VSDs (ssVSDs), non-sequence-specific C-terminal VSDs (ctVSDs), or physical structure-internal VSDs (psVSDs) [14,15].

ctVSDs have been identified in propeptides of several vacuolar proteins [15]. They are very variable in length and do not possess a conserved sequence motif. Their common feature is a hydrophobic amino acid patch within the propeptide [14,15,17].

*Corresponding author. Fax: +41 44 634 82 04.
E-mail address: fkel@botinst.uzh.ch (F. Keller).

Abbreviations: CLSM, confocal laser scanning microscopy; ctVSD, C-terminal VSD; ER, endoplasmic reticulum; α-Gal, α-galactosidase; GFP, green fluorescent protein; GGT, galactan:galactan galactosyltransferase; PSV, protein storage vacuole; psVSD, physical structure VSD; RFO, raffinose family oligosaccharide; sp, secretory pathway signal peptide; ssVSD, sequence-specific VSD; VSD, vacuolar sorting determinant

0014-5793/$32.00
doi:10.1016/j.febslet.2007.03.068

ArGGT is structurally and functionally related to plant α-galactosidases (α-Gals) of family 27 of glycosylhydrolases (CAZY, http://afmb.cnrs-mrs.fr/CAZY/; [8]. Plant acid α-Gals occur either in the apoplast [20,21] or the vacuole [4]. Comparison of acid α-Gal and ArGGT sequences revealed that both types of premature proteins contain an N-terminal signal sequence and a highly similar core sequence. Additionally, ArGGT and some acid α-Gals also contain a C-terminal extension with low sequence similarities to each other. These findings suggested that the C-terminal extension could be the ctVSD of ArGGT but also of some acid α-Gals. In this study, we demonstrate by means of green fluorescent protein (GFP) fusion with candidate peptides and subsequent exploiting transient expression in *Nicotians plumbaginifolia* mesophyll protoplasts that these C-terminal oligopeptides are non-sequence-specific determinants for protein sorting to the plant vacuole.

2. Materials and methods

2.1. Plant material

Plants of *N. plumbaginifolia* were grown in axenic conditions on half strength MS medium [22], 3% (w/v) Suc at 25 °C with a 16/8 h light/dark program.

2.2. Recombinant DNA

Constructs used in this work are shown in Fig. 1 (not to scale). All protein-coding sequences were fused downstream of the CaMV 35S promoter in the expression vector pART7 [23]. All constructs were inserted into the expression vector as *KpnI/XbaI* fragments. The ArGGT cDNA (ArGGT preprotein, corresponding to GenBank Accession No. AY386246) has been described previously [8]. Chimeric GFP-ArGGT constructs were prepared by the overlapping PCR technique [24] using GFP (GFP_6; [25]) as a reporter protein for intracellular ArGGT peptide trafficking. To improve primer-binding at the GFP 5′-end, the GC/AT ratio of the codons for Asp, Glu, Leu, and Lys was, preserving the encoded amino acids, altered from 35% to 55% using the primer 5′- GCTCTAGACTACTTGTACAGCTCGTCCATGCCATGTGTAATCCC-3′ (restriction site underlined).

In a first step, the sequences to be fused were prepared: (1) The N-terminal signal peptide of ArGGT was identified by SignalP ([26]) and amplified with the primers, 5′-GGGGTACCGATGGAGGCATCAGTGTTCTTCAC-3′ and 5′-*CCTTTACTCA*TTGCGAGGAGATTGCGTCTGTGATG-3′, the latter of which carrying an overlapping region with the 5′-end of GFP (overlapping regions are in italics). The signal peptide amplified corresponded to the amino acids 1–33 of the cDNA derived preprotein. (2) GFP was amplified with the primers, 5′-*CGCAATCTCCTCGC*AATGAGTAAAGGAGGAGAAC-3′ and 5′-*CCATTTCCGGGGC*TTGTACAGCTCGTCCATGCC-3′. (3) The ArGGT core protein including the C-terminus was amplified with the primers, 5′-*GGACGAGCTGTACAAG*CTCCTCGGAAATGGCCTCG-3′ and 5′-GCTCTAGAGTCACGACATTTGGAGTGACTTC-3′, and corresponded to amino acids 29–404 of the cDNA derived preprotein according to the sequence published [8].

In a second PCR step, the signal peptide of ArGGT was fused with the modified GFP by amplification with the primer, 5′-*GGGGTACCG*ATGGAGGCATCAGTGTTCTTCAC-3′, and either 5′-GCTCTAGACTTGTACAGCTCGTCCATGC-3′ for the construct spGFP or 5′-CCATTTCCGAGGCTTGTACAGCTCGTCCATGCC-3′ for all other spGFP-fusion constructs (spGFP-intermediary product).

In a third PCR step, the ArGGT core protein including the C-terminus was fused with the spGFP-intermediary product. SpGFP-ArGGT was obtained using the primers, 5′-GGGGTACCGATGGAGGCATCAGTGTTCTTCAC-3′ and 5′-GCTCTAGAGTCACGACATTTGGAGTGACTTC-3′. Analogously, spGFP-ArGGTΔSLQMS was obtained using the primers, 5′-GGGGTACCGATGGAGGCATCAGTGTTCTTCAC-3′ and 5′-GCTCTAGATCACTTCTTGTCCGCGGGTGTG-3′, resulting in a C-terminally ArGGT core protein truncated by five residues.

The five amino acids, SLQMS of spGFP-SLQMS, were added to spGFP using the primers, 5′-GGGGTACCGATGGAGGCATCAGTGTTCTTCAC-3′ and 5′-GCTCTAGA**TCACGACATTTGGAGTGA**-CTTGTACAGCTCGTCCATGC-3′ (bold letters indicate the corresponding nucleotides). The 10 amino acids, PADKKSLQMS of spGFP-PADKKSLQMS, were added to spGFP using 5′-GGGGTACCGATGGAGGCATCAGTGTTCTTCAC-3′ and 5′-GCTCTAGA**TCACGACATTTGGAGTGACTTCTTGTCCGCGGG**-CTTGTACAGCTCGTCCATGC-3′.

The three termini of Arabidopsis putative acid α-Gal – PQTVSHSDV (Accession Nos. AAL67017, and At3g56310), *Hordeum vulgare* leaf acid α-Gal fragment – PLKSQLD (Accession No. CAA74161; [27]), and *Oryza sativa* putative acid α-Gal – PLISSRAN (Accession No. AAP54408) were added to spGFP likewise with the reverse primers 5′-GCTCTAGA**TCAGTTGGCCCGGGACGAGATGAGGGG**-CTTGTACAGCTCGTCCATGC-3′ to produce spGFP-At αGal-IV/9, 5′-GCTCTAGA**TTAATCCAGCTGAGACTTGAGAGG**-CTTGTACAGCTCGTCCATGC-3′ to produce spGFP-Hv αGal-I/7, and 5′-GCTCTAGA**CTATACATCAGAGTGTGATACTGTCTGGGG**CTTGTACAGCTCGTCCATGC-3′ to produce spGFP-

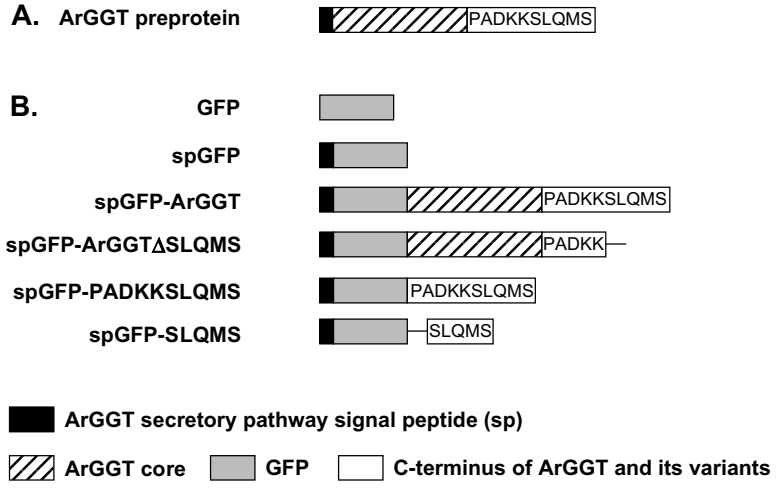

Fig. 1. Schematic diagram of the cDNA-derived ArGGT preprotein and its derivatives used in this study: (A), The ArGGT preprotein corresponding to the functionally expressed ArGGT-1 described previously [8]. (B) GFP reporter gene and ArGGT-GFP chimeric constructs expressed in tobacco mesophyll protoplasts for intracellular localization studies. All coding sequence constructs were under the control of the CaMV 35S promoter.