

Multiple internal sorting determinants can contribute to the trafficking of cruciferin to protein storage vacuoles

Dwayne D. Hegedus · Cathy Coutu ·
Myrtle Harrington · Brad Hope · Kelsey Gerbrandt ·
Ivo Nikolov

Received: 27 November 2014 / Accepted: 13 February 2015 / Published online: 22 February 2015
© Springer Science+Business Media Dordrecht 2015

Abstract Trafficking of seed storage proteins to protein storage vacuoles is mediated by carboxy terminal and internal sorting determinants (ISDs). Protein modelling was used to identify candidate ISDs residing near surface-exposed regions in *Arabidopsis thaliana* cruciferin A (AtCruA). These were verified by AtCruA fusion to yellow fluorescent protein (YFP) and expression in developing embryos of *A. thaliana*. As the presence of endogenous cruciferin was found to mask the effects of weaker ISDs, experiments were conducted in a line that was devoid of cruciferin. In total, nine ISDs were discovered and a core determinant defined using a series of alanine scanning and deletion mutant variants. Coupling of functional data from AtCruA ISD-YFP fusions with statistical analysis of the physiochemical properties of analogous regions from several 11/12S globulins revealed that cruciferin ISDs likely adhere to the following rules: (1) ISDs are adjacent to or within hydrophilic, surface-exposed regions that serve to present them on the protein's surface; (2) ISDs generally have a hydrophobic character; (3) ISDs tend to have Leu or Ile residues at their core; (4) ISDs are approximately eight amino acids long with the physiochemical consensus [hydrophobic][preferably charged][small or hydrophobic,

but not tiny][IL][polar, preferably charged][small, but not charged][hydrophobic, not charged, preferably not polar][hydrophobic, not tiny, preferably not polar]. Microscopic evidence is also presented for the presence of an interconnected protein storage vacuolar network in embryo cells, rather than discreet, individual vacuoles.

Keywords Seed storage protein · Cruciferin · Sorting determinants · Protein storage vacuole

Introduction

Plant cells possess two functionally distinct types of vacuoles, the most studied being the lytic vacuole (LV) and the protein storage vacuole (PSV) (Frigerio et al. 2008; Zouhar and Rojo 2009), that appear sequentially during embryo development (Hoh et al. 1995). LVs are the site of degradative processes and contain a myriad of hydrolases. Many of these enzymes are directed to the LV via clathrin-coated vesicles in the post-Golgi network (Paris and Neuhaus 2002) by BP-80 like vacuolar sorting receptors (VSR) that interact with sequence-specific signals, generally an NPRI motif near the amino terminus of the protein (Matsuoka and Neuhaus 1999). The mechanisms underlying trafficking of seed storage proteins (SSP) to the PSV are less clear and at least three types of intermediate vesicles or compartments have been implicated (Jolliffe et al. 2005; Robinson et al. 2005; Vitale and Hinz 2005). Precursor-accumulating (PAC) vesicles are thought to capture seed protein aggregates that accumulate in the endoplasmic reticulum (ER) and shuttle them directly to the PSV (Hara-Nishimura et al. 1998). These may be functionally analogous to the protein bodies that bud from the ER during the trafficking of prolamin in cereals (Herman and Schmidt 2004). SSP aggregates

Electronic supplementary material The online version of this article (doi:10.1007/s11103-015-0297-y) contains supplementary material, which is available to authorized users.

D. D. Hegedus (✉) · C. Coutu · M. Harrington · B. Hope ·
K. Gerbrandt · I. Nikolov
Agriculture and Agri-Food Canada, 107 Science Place,
Saskatoon, SK S7N 0X2, Canada
e-mail: Dwayne.Hegedus@agr.gc.ca

D. D. Hegedus · B. Hope
Department of Food and Bioproduct Sciences, University
of Saskatchewan, Saskatoon, SK, Canada

also emerge from the periphery of the Golgi cisternae as electron dense vesicles (DV) (Chrispeels 1983; Hohl et al. 1996; Hinz et al. 1999) and may proceed directly to the PSV or via a prevacuolar compartment (PVC), the multi-vesicular body (MVB) (Robinson et al. 1998; Jiang et al. 2000), that normally shuttles proteins to the LV (Hanton et al. 2007). Recent studies support a more unified theory where MVB are involved in trafficking of SSP during the early stages of embryogenesis, while DV are the predominant transport vehicle during the later stages (Wang et al. 2012). This is in keeping with earlier observations that the composition of SSP within the ER dictates the transport pathway to the PSV (Mori et al. 2004), since the types and amounts of SSP that accumulate in the embryo change as it matures.

Two types of VSRs have been implicated in the transport of SSP to PSVs; BP-80-like receptors and receptor-like membrane RING-H2 domain (RMR) receptors. The two main integral membrane proteins associated with pumpkin seed PAC vesicles, PV72 and PV82, are BP-80 orthologs and interact with SSPs (Shimada et al. 1997; 2002). The *Arabidopsis thaliana* VSRs, VSR1 thru VSR7, are also BP-80 orthologs. AtVSR1 is found mostly in the trans-Golgi system and MVB, and is recycled back to the trans-Golgi from the MVB (Hinz et al. 2007, Hanton and Brandizzi 2006). Mutation of *AtVSR1* causes partial mis-sorting of 2S and 12S SSP (Shimada et al. 2003a) and reporters carrying their sorting determinants (Craddock et al. 2008) to the apoplast. Pyramiding of mutations in *AtVSR1+3* or *AtVSR1+4* caused missorting of nearly all 2S and 12S SSP, respectively (Zouhar et al. 2010). In *A. thaliana* plants, DV, MVB and other pre-vacuolar compartments that are distinct from the lytic pre-vacuolar compartments contain RMR receptors (Hinz et al. 2007; Shen et al. 2011), which bind carboxy terminal vesicle sorting determinants (ctSDs), but these are not recycled back to the Golgi (Jiang et al. 2000; Park et al. 2007).

Identification of SSP sorting determinants, let alone their cognate receptors, has been difficult and the major classes of SSP may take different routes to the PSV. Vacuolar targeting was determined by amino acids within the internal processed peptide of the *Ricinus communis* (caster bean) 2S albumin precursor (Brown et al. 2003) and the analogous proricin linker (Frigerio et al. 2001). Subsequently, the associated sorting determinants were found to be recognized by BP-80 like receptors (Jolliffe et al. 2004). This supported earlier reports showing that PV72 and PV82 interact with the internal propeptide and carboxy terminal peptide of pumpkin 2S albumin (Shimada et al. 1997; 2002). In other studies, AtVSR1 was shown to bind to the carboxy terminal peptide of the 12S globulin, cruciferin, in a calcium-dependent manner, similar to the interaction of BP-80 like receptors with NPIR motifs (Shimada et al.

2003a). Similarly, *A. thaliana* RMR1 (AtRMR1) interacted with the carboxy terminal peptide of the *Phaseolus vulgaris* 7S globulin phaseolin (Park et al. 2005) and chitinase (Park et al. 2007). 7S and 12S globulins are the major class of SSP in many dicots and both contain β -barrel cupin-like domains that are structurally similar, the main difference being that 7S globulins form trimers while 12S globulins form hexamers at the quaternary level (Shewry et al. 1995). Four carboxy terminal amino acids (AFVY) comprise a major sorting determinant for phaseolin and deletion or mutation of these residues leads to secretion (Frigerio et al. 1998). The 7S SSP from *Glycine max*, conglycinin, contains a similar carboxy terminal sorting determinant (AFY), but also a second internal sorting determinant (ISD) a few residues upstream (Nishizawa et al. 2003, 2004, 2006). CtSDs and ISDs have also been found in the 11S globulins glycinin from *G. max* (Maruyama et al. 2006) and amaranthin from *Amaranthus hypochondriacus* (Petruccioli et al. 2007). In these proteins, the ISD resides adjacent to a loop or disordered region (DR) on the surface of the protein with an Ile residue at its core.

Our goal was to identify sorting determinants that direct the 12S globulin, cruciferin, to the PSV. Here we used protein modelling to identify candidate ISDs near *A. thaliana* cruciferin (AtCruA) surface-exposed DRs. The determinants were verified by fusion to yellow fluorescent protein (YFP) and their sorting patterns determined in cotyledons of an *A. thaliana* line that was devoid of endogenous cruciferin. This strategy allowed identification of weaker ISDs that could not be evaluated when conducting experiments in the wild type line. The core determinant was defined using a systematic series of alanine scanning and deletion mutant variants that were also fused to YFP. In addition, three dimensional imaging of confocal microscopy data revealed that SSP reside within an interconnected protein storage vacuolar network, rather than discreet, individual vacuoles.

Results

Homology modeling of AtCruA and identification of regions containing internal sorting determinants

In silico modeling of the monomeric form of AtCruA revealed five surface-exposed regions, four of which were spatially analogous to the DRs found in the soybean 11S proglycinin A1aB1b trimer (Adachi et al. 2001) (Fig. 1). Previous studies indicated that these regions may either contain ISDs or help to display them on the surface of the protein (Maruyama et al. 2006; Petruccioli et al. 2007). Extending these studies, we appended the AtCruA DR, as well as surface-exposed amino acids adjacent to them,

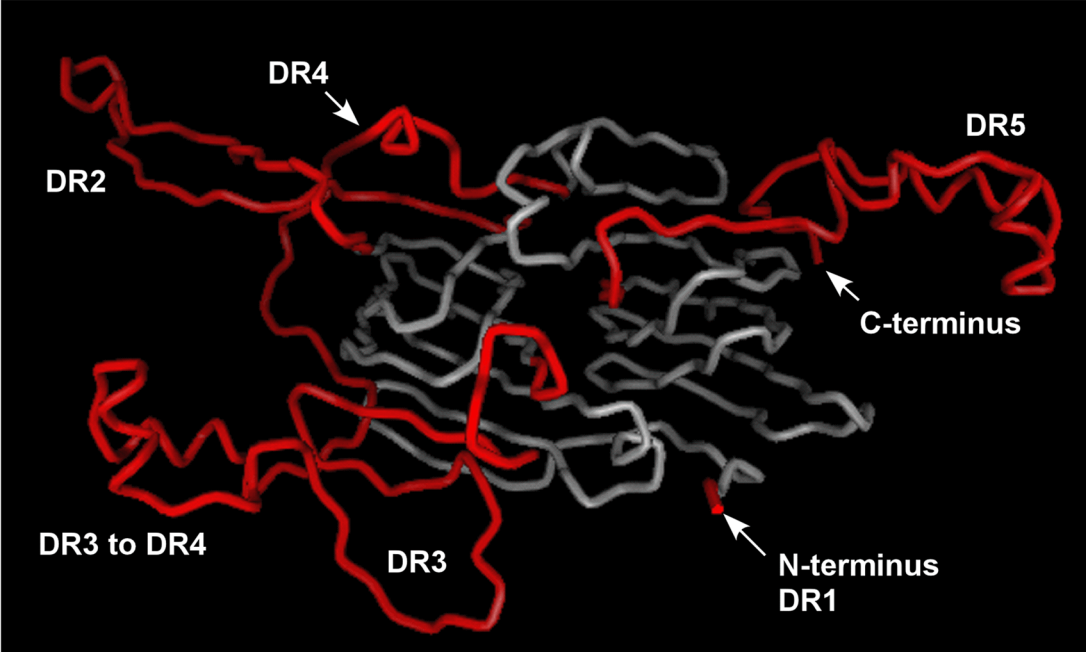
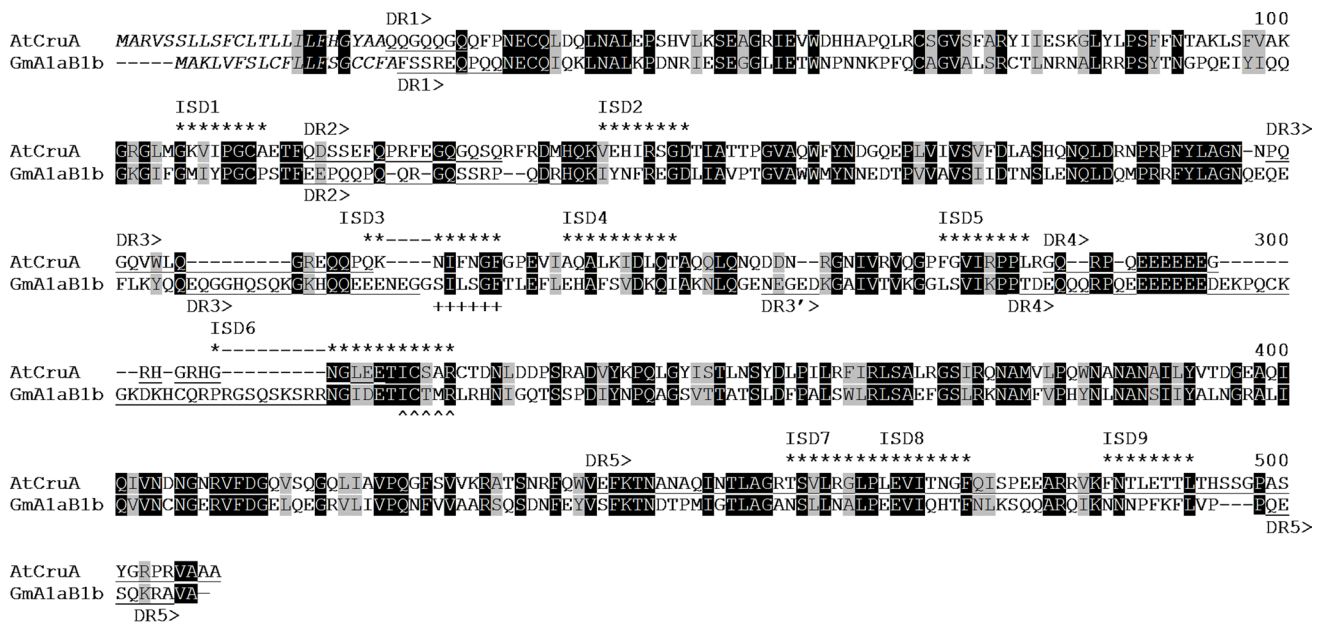


Fig. 1 Alignment of *A. thaliana* CruA (AtCruA) with *G. max* A1aB1b (GmA1aB1b) showing the location of disordered regions (DR1-5 and *underlined*) and AtCruA internal sorting determinants (ISD and denoted with *asterisk*). *Bottom panel* shows the three

dimensional organization of the polypeptide chains from a single AtCruA subunit. The core cupin domain β -barrel is shown in *grey* and the surface exposed/disordered regions in *red*. The DR3–DR4 region interacts with the DR5 region in the Cru trimer

to YFP and expressed the fusion proteins in *A. thaliana* plants (Table 1). Initially, a single construct containing a region with a known ISD (the AtCruA DR4 + C' region analogous to that containing the *G. max* glycinin ISD) was expressed in wild type *A. thaliana* plants, as well as in a plant line devoid of cruciferin. The rationale for this was based on the notion that competition with endogenous cruciferin for receptor molecules makes identification

of weaker individual sorting determinants difficult in a wild type background (Maruyama et al. 2006). Indeed, we observed strong PSV localization with the known ISD linked to YFP in the cruciferin-deficient mutant, but only limited or no PSV localization in the wild type line (Fig. 2). As in earlier studies (Maruyama et al. 2006; Petruccioli et al. 2007), we appended either four or six glycine residues to the putative ISDs, since this places

Table 1 Localization of YFP linked to putative cruciferin internal vacuolar sorting determinants

Region	Sequence ^{a,b}	YFP localization ^c			ISD designation
		EM	PSV	Sec	
DR1 + C'	SP-YFP-AQQGQQGQQFPNECQLDQLNALEPSHVLKSE-GGGG	+++		+	
DR2 + N'	SP-YFP-GKVIPGCAETFDQSSEFQPRFEGQGQSQ-GGGG		+	+++	
	SP-YFP-GKVIPGCA-GGGGGG	+	+	+++	ISD1
DR2 + C'	SP-YFP-QDSSEFQPRFEGQGQSQRFRDMHQKVEHIRSGD-GGGG		++	++	
	SP-YFP-VEHIRSGD-GGGGGG			++++	ISD2
DR3 + N'	SP-YFP-DLASHQNQLDRNPRPFYLAGNNPQGQVWLOGREQQPQ-GGGG		+	+++	
DR3 + C'	SP-YFP-PQGQVWLOGREQQPQKNIFNGF-GGGG		++	++	
	SP-YFP-QKNIFNGF-GGGGGG	+++		+	ISD3
DR3-DR4	SP-YFP-KNIFNGFGPEVIAQALKIDLQTAQQLQNQDDNRGN-GGGG	++++			
	SP-YFP-AQALKIDL-GGGGGG	++		++	
	SP-YFP-ALKIDLQT-GGGGGG	++++			ISD4
DR4 + N'	SP-YFP-IVRVQGPFVIRPPLRGQRPOEEEEEGRHGRHGNGLLEE-GGGG	++++			
	SP-YFP-FGVIRPPL-GGGGGG	++		++	ISD5
DR4 + C'	SP-YFP-GQRPOEEEEEGRHGRHGNGLLEETICSARCTDNL-GGGG		+++	+	
	SP-YFP-GNGLEETI-GGGGGG	+	+	+++	
	SP-YFP-EETICSAR-GGGGGG		+	+++	ISD6
DR4 + C'CruB	SP-YFP-RGEGGQOPHEIANGLEETLCTMRCTENL-GGGG		++	++	
	SP-YFP-EETLCTMR-GGGGGG		++	++	
DR5 + N'	SP-YFP-VEFKTNANAQINTLAGRTSVLRGLPLEVITNGFQISPEEARVKFNTLETT-GGGG		++++		
	SP-YFP-TSVLRGLP-GGGGGG		++++		ISD7
	SP-YFP-LEVITNGF-GGGGGG		++	++	ISD8
	SP-YFP-FNTLETTL-GGGGGG		++	++	ISD9

^a Disordered regions are underlined. Putative internal sorting determinants are in italics

^b DR1 to DR5 = disordered regions; SP = signal peptide; YFP = yellow fluorescent protein

^c Proportion of YFP sorted to the endomembrane system (EM), to the protein storage vacuole (PSV) or secreted (Sec) to the apoplast based on fluorescence intensity. All (++++), most (+++), partial (++) , slight amount (+)

the determinants in an internal context and blocks a weak cryptic, PSV sorting determinant present near the carboxyl terminus of YFP (Tables 1, 2).

DR1 resides at the amino terminus of the protein and is situated at the base of the β -barrel that defines the cupin domain. YFP fused to DR1 along with the amino acids carboxyl terminal to it was neither secreted nor localized to the PSV. Rather, the majority of the protein remained within the cell in what appears to be the endomembrane system (Fig. 3). YFP fused to DR2 along with the amino terminal amino acids was mostly secreted, though a small amount was observed in the PSV. YFP fused to DR2 with the carboxyl terminal amino acids was also secreted, but a higher proportion was found in the PSV suggesting that this region contains a weak ISD. Similar to DR2, YFP fused to DR3 and the amino terminal amino acids was mostly secreted, while YFP fused to DR3 and the carboxyl terminal amino acids was secreted and localized to the PSV in approximately similar proportions. The region between

DR3 and DR4 is an α -helical region that interacts with the α -helical region upstream of DR5 to form the trimer. This was examined separately as an ISD was found previously in this area; that being the NIFGRF motif in the *A. hypochondriacus* 11S globulin amaranthin (Petruccioli et al. 2007). Indeed, YFP fused to the DR3-DR4 region was strongly localized to the PSV. DR4 forms an extended loop that laps onto the β -barrel of the adjacent protomer to further stabilize the trimer. YFP fused to DR4 and either the amino or carboxyl terminal amino acids was strongly localized to the PSV. The region carboxyl terminal to DR4 contains the amino acids ICSAR at the same location as the ICTMR ISD of the *G. max* A1aB1b glycinin (Maruyama et al. 2006). DR4 and carboxyl terminal amino acids from *A. thaliana* CruB (AtCruB), which contains LCTMR at the analogous location, was more effective in directing YFP to the PSV. DR5 comprises the carboxyl terminal amino acids and was shown to contain both a carboxyl terminal sorting determinant (ctSD) and an ISD (Nishizawa et al.

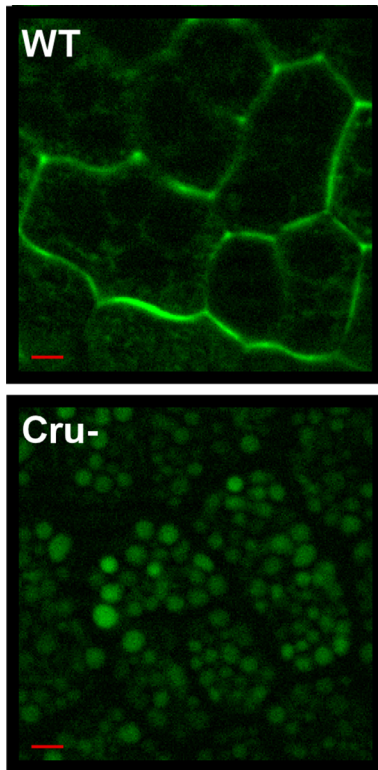


Fig. 2 Epifluorescence microscopic images of embryos cells from wild type (WT) and cruciferin-deficient (Cru-) *A. thaliana* lines expressing YFP-linked to a region containing a CruA internal sorting determinant (AtCruA DR4 + C'). YFP was both secreted and localized in the PSV in the WT line, while YFP was localized only in the PSV in the Cru- line. The red bar in the lower left of each panel is 5 μ m

2006). We examined the area upstream of this which forms an α -helical region that, as noted above, interacts with the DR3 to DR4 α -helical region of an adjacent protomer in the trimer. DR5 is at the end of AtCruA and was shown to contain a putative ctSD (Shimada et al. 2003a); therefore, this region was excluded from the construct. However, when fused to the region upstream of DR5, YFP was strongly localized to the PSV indicating that a strong ISD was also present.

Identification of individual internal sorting determinants

The identification of ISDs and an associated physiochemical property ISD consensus was conducted using a combination of empirical and informatics approaches and is described in detail below. In brief, regions adjacent to the DRs in AtCruA were tested for their ability to localize YFP to the PSV (above). 13 amino acid blocks centered on Ile or Leu were then compiled from regions that exhibited some ability to localize YFP to the PSV, as well as from the analogous regions in the other *A. thaliana* cruciferins.

A consensus for each position was developed based on over-representation of amino acid physiochemical properties relative to random surface exposed regions. Application of this physiochemical consensus back to AtCruA yielded a list of putative ISDs that were appended to YFP and tested in planta. A refined physiochemical consensus was then determined by comparing only functional AtCruA determinants. Mutation analysis was used to better define the length of the ISD and amino acids critical to their function. Finally, hidden Markov models (HMM) were used to identify an ISD consensus for each individual determinant by comparison to analogous regions in other 11S globulins.

Nine of the ten regions tested in the first generation constructs showed some level of PSV localization. In earlier studies, substitution of the Ile or Leu residue was shown to inactivate the vacuolar targeting ISDs in glycinin (Maruyama et al. 2006), conglycinin (Nishizawa et al. 2006), ricin (Frigerio et al. 2001), *R. communis* 2S albumin (Brown et al. 2003) and *Ipomoea batatas* (sweet potato) sporamin (Matsuoka and Nakamura 1999). In accordance, a second set of constructs was designed to identify individual ISDs within the regions flanking the DRs with either Ile or Leu at the core of the determinant (Table 1). In total, 26 Ile/Leu residues were present in the regions that showed some PSV targeting in the first generation constructs. A tiered screening process was used to select regions more likely to be ISDs. Initially, three regions in which the Ile/Leu was within three amino acids of the end of the tested regions were excluded since the ISD is likely to be incomplete. Subsequently, the AtCruA regions (Ile/Leu plus 6 flanking amino acids) were aligned with the homologous regions of AtCruB, AtCruB' and AtCruC. It is worth noting that all of the 23 remaining AtCruA Ile/Leu regions were conserved in at least one other AtCru member. The list was further refined by removing three putative ISDs obtained from very weak PSV targeting regions (DR2 + N' and DR3 + N'), leaving 19 putative AtCruA ISDs and 52 putative ISD sequences from all four cruciferins (Supplemental Table 2). The frequency of individual amino acids at each position was calculated, but no consensus was obvious. Subsequently, the representation of amino acids with or without specific chemical properties (Livingstone and Barton 1993) was determined for each position. Two methods were then used to identify a physiochemical consensus. Firstly, the most common physiochemical properties at each position (over 50 %) were determined (Supplemental Table 3.1) yielding the following consensus: [small or polar or hydrophobic][small or hydrophobic][polar or hydrophobic][small or hydrophobic][polar][small or polar or hydrophobic][IL][polar][small or polar or hydrophobic][small or hydrophobic][polar or hydrophobic][small or polar or hydrophobic][small or polar or hydrophobic], which can also be written as [any][ACDFGHKLMNPSTVWY][ACDEF

Table 2 Localization of YFP linked to alanine scanning and deletion mutants of a cruciferin internal vacuolar sorting determinant

Construct	Sequence	Localization ^a	
		PSV	Secreted
ISD3	SP-YFP- Q K N I F N G F -GGGGGG	+++	+
A1	SP-YFP- Q K A I F N G F -GGGGGG	+	+++
A2	SP-YFP- Q K N A F N G F -GGGGGG	++	++
A3	SP-YFP- Q K N I A N G F -GGGGGG	++	++
A4	SP-YFP- Q K N I F A G F -GGGGGG	+++	+
A5	SP-YFP- Q K N I F N A F -GGGGGG	++++	-
D1	SP-YFP- Q K N I F N G - -GGGGGG	+	+++
D2	SP-YFP- Q K N I F N - - -GGGGGG	+	+++
D3	SP-YFP- Q K N I F - - - -GGGGGG	+	+++
D4	SP-YFP- Q K N I - - - - -GGGGGG	-	++++
D5 ^b	SP-YFP- k K N I F N G F -GGGGGG	+	+++
D6 ^b	SP-YFP- y k N I F N G F -GGGGGG	+	+++
D7 ^b	SP-YFP- l y k I F N G F -GGGGGG	+++	+

^a Proportion of YFP sorted to the protein storage vacuole (PSV) or secreted to the apoplast based on fluorescence intensity. All (++++), most (+++), partial (++), slight amount (+)

^b The terminal amino acids of YFP are Leu (l), Tyr (y) and Lys (k) which occupy the deleted positions

GHIKLMNQRSTVWY][ACDFGHIKLMNPSTVWY][CDEHKNQRSTVWY][any][IL][CDEHKNQRSTVWY][any][ACDFGHIKLMNPSTVWY)][ACDEFGHIKLMNQRSTVWY][any][any] or [any][not EQR][not P] [not EQR][not AFGILMPV][any][IL][not AFGILMPV][any][notEQR][not P][any][any].

Secondly, recognizing that some physiochemical properties are more common in surface exposed regions, an average and standard deviation for each property in the regions tested was determined (Supplemental Table 3.2) by comparison to a reference set comprising 51 eighteen amino acid blocks from surface exposed regions in the four *A. thaliana* cruciferins (Supplemental Tables 3.3 and 3.4). This resulted in a similar, but more refined consensus that included fewer amino acids than the first as follows: [any][hydrophobic, not charged][any][hydrophobic, not polar][not hydrophobic][hydrophobic][IL][polar, not hydrophobic][small, not charged][any][hydrophobic][tiny][any], which can also be written [any][ACFGILMTVWY][any][AFGILMV][DENPQRS][ACFGHIKLMTVWY][IL][DENQRS][ACGNPSTV][any][ACFGHIKLMTVWY][ACGS][any] or [any][not DEHKNPQRS][any][not CDEHKNPQRSTVWY][not ACFGHIKLMTVWY][not DENPQRS][IL][not ACFGHIKLMPTVWY][notDEFHIKLMQRWY][any][not DENPQRS][not DEFHIKLMNPQRTVWY][any].

Several factors were considered when determining the length of the putative ISDs to test in the second generation constructs. Firstly, the ISDs in amaranthin (NIFRGF) and glycinin (ICTMR) include four amino acids downstream of the Ile residue, the last of which coincided with the highly significant (1 % confidence level) amino acid in the consensus at Position 11 (Supplemental Table 3.2). Position 12 was also highly significant, but with a requirement for a tiny amino acid that is provided by the first Gly in the hexaglycine appendage. Secondly, the amaranthin and glycinin ISDs included either no or a single amino acid upstream of the Ile/Leu; however, amino acids with significantly over-represented physiochemical properties were found one, two, three and five residues upstream of the Ile/Leu at consensus Positions 6, 4 and 2, respectively. Since the penultimate amino acid in YFP conforms to the consensus for Position 2, the second generation constructs consisted of eight amino acids (three upstream and four downstream of the Ile/Leu). When the two related eight amino acid consensus were applied to the regions flanking the DRs in AtCruA, 15 of the 22 potential ISDs were identified by one or both methods as having less than three mismatches (Supplemental Table 4). Eleven ISDs were selected for further analysis, including the regions analogous to the amaranthin and glycinin ISDs. In addition, the putative ISD

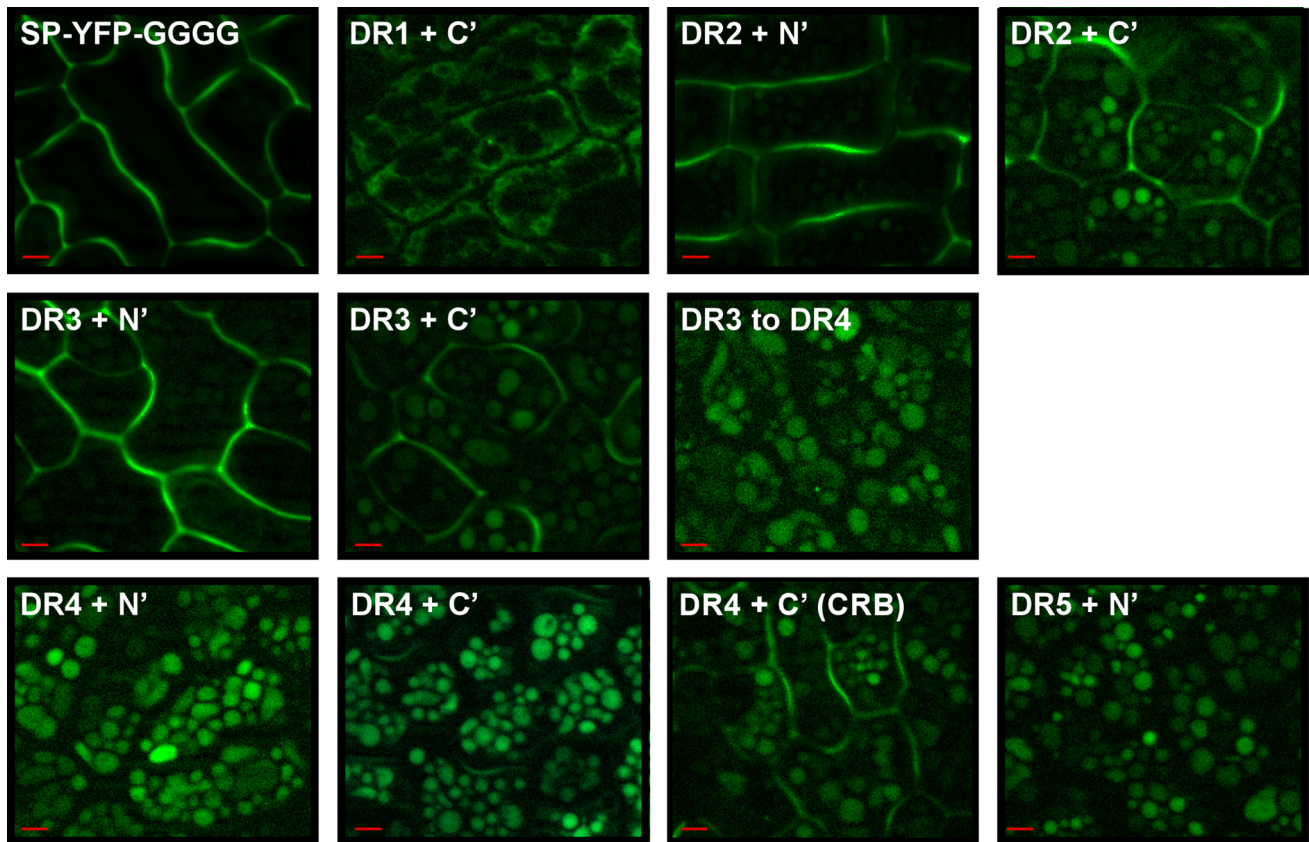


Fig. 3 Epifluorescence microscopic images of embryos cells from a cruciferin-deficient (Cru-) *A. thaliana* line expressing YFP alone (YFP-GGGG) or YFP fused to cruciferin disordered regions and adja-

cent putative internal sorting determinants. The amino acid sequences of the regions tested are listed in Table 1. The red bar in the lower left of each panel is 5 μm

from AtCruB (EETLCTMR) was included since it was more similar to the glycinin ISD (ICTMR) than that found in AtCruA (EETICSAR).

When these eight amino acid sequences were fused to YFP, 11 of 12 showed some PSV-localization activity (Table 1). The original YFP fusion to DR2 indicated that weak ISDs may reside in the amino and carboxyl terminal regions. Fusion of the amino terminal GKVIPGCA (ISD1) or carboxyl terminal VEHIRSGD (ISD2) sequences to YFP directed only a small amount or none of the fusion protein to the PSV, respectively. The DR3-DR4 region contained three putative individual ISDs, including QKNIFNGF (ISD3) which is in an analogous position to the NIFGRF determinant from amaranthin. The AtCruA QKNIFNGF sequence was a relatively strong sorting determinant and directed most of the YFP to the PSV, as did ALKIDLQT (ISD4). The AQALKIDL sequence which comprises part of the ISD4 determinant was weaker and directed only some of the YFP to the PSV. In the original construct, DR4 and the amino terminal region strongly directed YFP to the PSV. The FGVIRPPL (ISD5) sequence from this region also functioned as an ISD, though more weakly in

this context. The DR4 and carboxyl terminal region was a relatively strong ISD. This region contains the sequence ICSAR in AtCruA and LCTMR in AtCruB which are similar and in analogous locations to the glycinin A1aB1b sorting determinant ICTMR. The YFP fusion containing the CruA EETICSAR sequence was mostly secreted; however, the fusion protein containing the CruB EETLCTMR, which is more similar to the glycinin determinant, was partially directed to the PSV. The region amino terminal to DR5 strongly localized YFP to the PSV and three individual sorting determinants were found in this region. TSVLRGLP (ISD7) was a strong sorting determinant, while LEVITNGF (ISD8) and FNTLETTL (ISD9) were weaker, but did direct some YFP to the PSV.

Mutation analysis of an internal sorting determinant

To further define the minimal requirements for an ISD, a series of alanine scanning and deletion mutants were generated based on the AtCruA QKNIFNGF sequence (denoted as ISD3) (Table 2). The analogous region in amaranthin (GNIFRGF) is also a known ISD (Petruccioli et al. 2007).

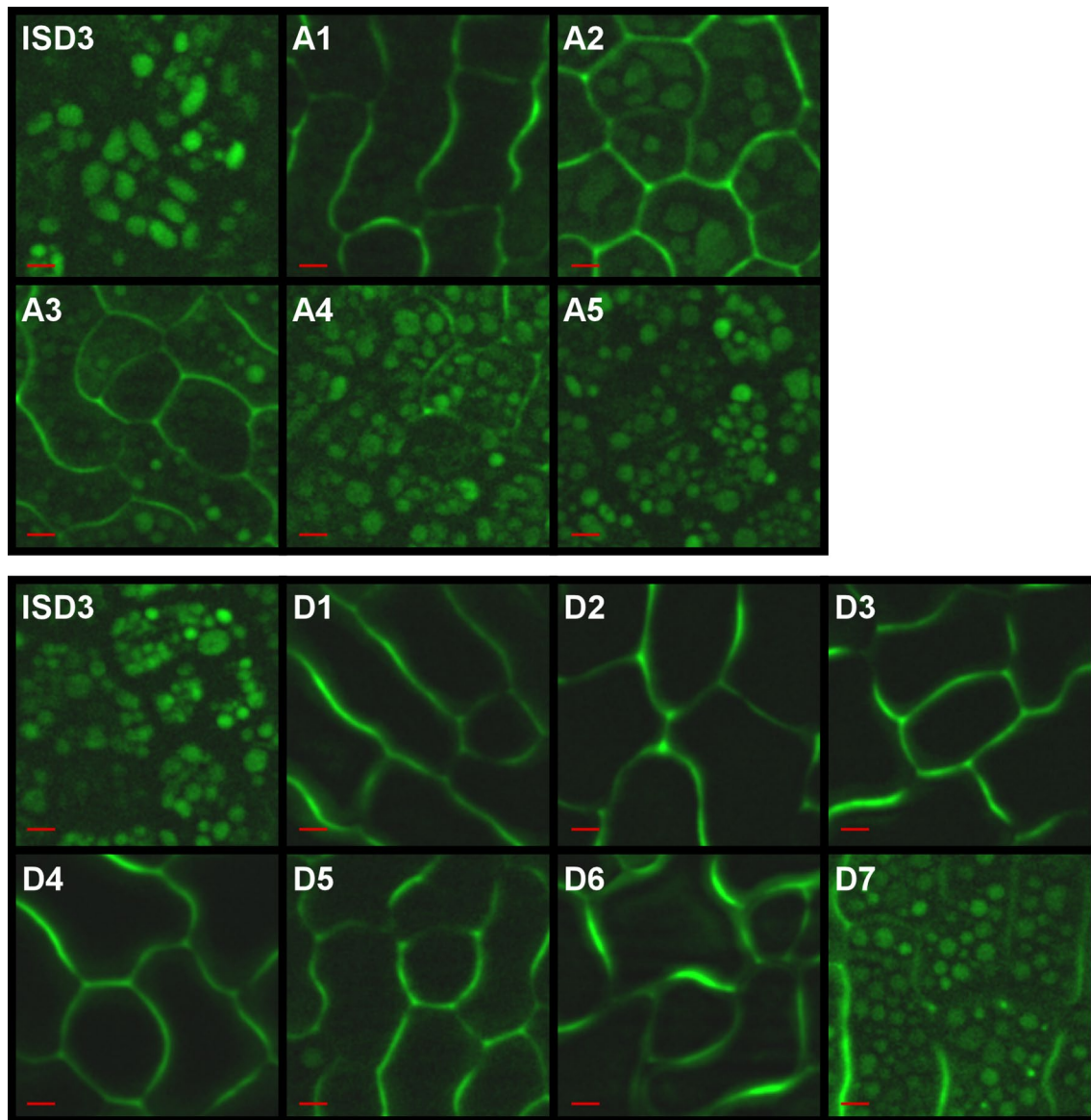


Fig. 4 Epifluorescence microscopic images of embryos cells from a cruciferin-deficient (Cru-) *A. thaliana* line expressing YFP fused to AtCruA internal sorting determinant 3 (ISD3) or to alanine scan-

ning (A1-A5) or deletion (D1-D7) mutant derivatives. The amino acid sequences of the ISD3 mutants and their localization patterns are listed in Table 2. The red bar in the lower left of each panel is 5 μm

The QKNIFNGF sequence was a moderate to strong ISD and directed about half of the YFP to the PSV, with the remainder secreted to the apoplast (Fig. 4). Mutational analysis of an ISD with intermediate strength allowed identification of changes that either reduced or improved sorting efficiency. Replacement of the Asn by Ala greatly increased the proportion of YFP that was secreted, though a small amount was still directed to the PSV. Replacement of the Ile, which was presumed to be at the core of the motif, or the Phe residue immediately adjacent to it with Ala, reduced, but did not abolish sorting of YFP to the PSV. Replacement of the second Asn had no effect on sorting of

YFP to the PSV. Interestingly, replacement of the glycine with alanine, both of which have similar physiochemical properties, slightly increased ISD function resulting in exclusive localization of the YFP to the PSV. Conversely, deletion of Gln or Gln and the adjacent Lys residue from the amino terminus of the ISD greatly decreased the proportion of YFP sorted to the PSV. Similarly, deletion of one or more of the carboxy terminal amino acids severely affected ISD function, suggesting that the minimal functional length of this particular ISD is at least eight residues. Interestingly, deletion of the Gln, Lys and Asn residues restored ISD sorting capacity. This might be due to amino

Table 3 Physiochemical properties of amino acids in functional *A. thaliana* cruciferin internal vacuolar sorting determinants^a

Category	Members	Frequency of physiochemical properties (%) at individual ISD positions relative to surface exposed regions on cruciferin ^a							
		Position 1 ^b	Position 2	Position 3	Position 4	Position 5	Position 6	Position 7	Position 8
Tiny	ACGS	14	28	0 [^]	0	14	14	28	0 [^]
Small	ACDGNPSTV	28 [^]	42	85 ^{**}	0	42	85 ^{**}	57	28
Positive	HKR	0 [^]	14	14	0	28 ^{**}	0 [^]	0 [^]	14
Negative	DE	14	28 ^{**}	0	0	28 ^{**}	0	0	0
Charged	DEHKR	14	42 ^{**}	14	0	57 ^{**}	0 [^]	0 [^]	14
Polar	CDEHKNRSTWY	42	71	57	0	85 ^{**}	57	28 [^]	28 [^]
Aliphatic	ILV	14	14	42 ^{**}	100	0 [^]	14	14	28
Aromatic	FHWY	28 ^{**}	0	0	0	14	0	0	28 ^{**}
Hydrophobic	ACFGHIKLMTVWY	71 ^{**}	42	85 ^{**}	100	42	57	71 ^{**}	71 ^{**}
Consensus property	Hydrophobic	Preferably charged	Small or hydrophobic, but not tiny, preferably V or T	IL	Polar, preferably charged	Small, not charged	Hydrophobic, not charged, preferably not polar	Hydrophobic, not tiny, preferably not polar	
Amino acids with consensus property	ACFGHIKLMT-DEHKR-VWY	DFHIKLMNPT-VWY	IL	CDEHKNRSTWY	ACGNPSTV	AFGILMV	FILMV		
Amino acids found in AtCruA ISDs ^b									
ISD3	Q	K	N	I	F	N	G	F	
ISD4	A	L	K	I	D	L	Q	T	
ISD5	F	G	V	I	R	P	P	L	
ISD6 (CruB)	E	E	T	L	C	T	M	R	
ISD7	T	S	V	L	R	G	L	P	
ISD8	L	E	V	I	T	N	G	F	
ISD9	F	N	T	L	E	T	T	L	

^a Functional ISDs examined were: ISD3, ISD4, ISD5, ISD6 (CruB), ISD7, ISD8 and ISD9. Physiochemical properties over-represented at the 1 % (**), confidence interval, or under-represented at the 1 % (^) and 5 % (^) confidence interval relative to amino acids from surface-exposed regions of AtCruA, AtCruB, and AtCruC (Supplemental Table 3.3 and 3.4)

^b Amino acids represented by 1-letter symbol in bold conform to the consensus

acids at the carboxyl terminus of YFP (LYK) that are now in these positions, the first and third of which conform to the consensus.

Determination of an internal sorting determinant consensus

Attempts were made to identify an ISD consensus among 7 ISDs that directed at least some YFP to the PSV. The secondary structure surrounding the core I/L residue in the *A. thaliana* cruciferin ISDs was determined by Predict Protein, as well as by homology modeling using either the *G. max* proglycinin (PDB:2d5f) and *B. napus* cruciferin (PDB:3kg1A) as templates. No β -sheet structure was detected in any of the ISDs and no apparent consensus was found as to the location or length of α -helical regions (Supplemental Table 5). The physiochemical properties of the most functional *A. thaliana* ISDs were also compared (Table 3) which yielded the following consensus:

[hydrophobic][preferably charged][small or hydrophobic, but not tiny][IL][polar, preferably charged][small, but not charged][hydrophobic, not charged, preferably not polar][hydrophobic, not tiny, preferably not polar], which can also be written as [ACFGHIKLMTVWY][DEHKR][DFHIKLMNPTVWY][IL][CDEKHNRSSTWY][ACGNPSTV][AFGILMV][FILMV]. Of note, was the preponderance of hydrophobic amino acids adjacent to charged and/or polar amino acids. Hydrophobic amino acids are normally in reduced abundance on the surface of proteins and this may impart the ISD with a unique a physiochemical signature.

An attempt was also made to define an ISD consensus using hidden Markov models (HMM). Initially, seven ISDs from AtCruA (ISD3, ISD4, ISD5, ISD6, ISD7, ISD8 and ISD9) were aligned and used to generate a HMM. ISD1 and ISD2 were excluded since they are weak or non-functional. ISD6 was also a weak determinant, but was included

since it is analogous to the amaranthin ISD. When the TAIR database was searched with this model, no results were returned. To increase the power of the HMM, the analogous AtCruB and AtCruC ISD regions were included in the analysis. A search of the TAIR database with this HMM returned all three cruciferins with 5 of 7 (ISD3, 5, 6, 8 and 9) input ISDs (Supplemental Table 6). When the search was extended to the *B. rapa* and *B. oleracea* databases, cruciferin was identified with the same 5 of 7 input ISDs. When the search was further extended to all plant sequences in the NCBI refseq database, no hits were returned (Supplemental Table 6). This was not unexpected since the HMM was based on short peptide sequences which have an increased likelihood of being present by chance within a larger database causing the e-value to drop below the significance threshold.

When analogous ISD regions from a variety of dicots were aligned it was noted that there was more conservation of ISDs within a group than between ISD groups (Supplemental Table 7). Analogous regions corresponding to seven functional *A. thaliana* ISDs (ISD1, ISD3, ISD4, ISD5, ISD6, ISD7 and ISD8) were aligned and used as input to create HMMs for each ISD. ISD9 was excluded since it was not conserved outside the Brassicaceae. When the individual ISD alignments were used as HMM input and the resulting models used to search the TAIR, *B. rapa* or *B. oleracea* databases, ISD1, ISD3, ISD6 and ISD8 returned cruciferin or cruciferin-like proteins, while ISD4, ISD5 and ISD7 did not (Supplemental Tables 8.1–8.7). In some cases, proteins that were unlikely to be targeted to the PSV were returned, but all of these were predicted to contain multiples of a single ISD which inflated the scores. In total, three *A. thaliana*, four *B. rapa* and five *B. oleraceae* cruciferins were identified (Supplemental Table 8.8). A search of the plant delimited NCBI refseq database with these HMM models failed to return any cupin domain proteins. As before, this is likely due to the short length of the input sequences relative to the large size of the database searched.

Three dimensional visualization of the protein storage vacuole system

Two dimensional images of embryo cells accumulating fluorescent protein fusions in this and other studies show what appear to be numerous, discreet PSVs of varying sizes, shapes and fluorescent intensities (Figs. 2, 3, 4). During the course of the localization studies above, it was noted that when YFP was appended to a strong ISD and expressed in the cruciferin-deficient plant line, the YFP signal co-localized completely with PSV autofluorescence. A similar observation was made previously with red fluorescent proteins attached to ricin or phaseolin PSV sorting

determinants (Hunter et al. 2007). When we examined Z-stacks of the PSV fluorescence, connections between the PSVs appeared to be present. To explore this further, high resolution Z-stacks of embryo cells in seeds of a cruciferin-deficient line expressing SP-YFP-DR5 + N' were combined to form a three dimensional image of the PSVs. The PSV autofluorescence and YFP signals again co-localized; however, PSV autofluorescence was preferred for imaging since it was observed with a shorter 2-photon excitation wavelength (720 nm instead of 920 nm) resulting in better resolution, and had a better signal to noise ratio that yielded a cleaner image. Two-photon confocal microscopy was used to minimize bleaching in the large Z-stack. This analysis revealed that the PSVs are not discreet, individual, spherical entities, but rather are part of an interconnected network (Fig. 5; Supplemental Movie 1). Irregular shaped nodes that appear as distinct vacuoles when observed on a single plane by conventional microscopy were shown to be fused to other nodes or connected by short passages. For two cells, the ratio of cell (2525 and 2656 μm^3) to PSV (1103 and 1215 μm^3) volume was determined which indicated that the total volume of the cell occupied by the PSV network ranged from 43.9 to 45.7 %.

Discussion

Nine weak to strong ISDs adjacent to DRs were identified in *A. thaliana* CruA. Earlier studies revealed that the carboxy terminal peptide of CruA was recognized by the vacuolar sorting receptor VSR1 and this was presumed to be sufficient to direct it to the PSV (Shimada et al. 2003a). Subsequently, other 11S/12S globulins were found to possess ISDs that reside adjacent to a loop (DR) on the surface of the protein with an Ile residue at their core (Maruyama et al. 2006; Petruccioli et al. 2007). The genome of *A. thaliana* cv. Columbia contains four genes that encode subunits of the 12S globulin cruciferin. Three of these genes, At5g44120 (CruA), At1g03880 (CruB) and At4g28520 (CruC), are highly transcribed and mass spectroscopy revealed that they contribute to hexamer formation in the following order CruC > CruA > CruB (Wan et al. 2007). In silico models for the homohexameric forms of each have been generated (Withana-Gamage et al. 2011) and our examination of the CruA structure revealed five surface exposed regions that were spatially analogous to the DRs found in the *G. max* proglycinin A1aB1b trimer (Adachi et al. 2001) and the fully processed A3B4 hexamer (Adachi et al. 2003). The location of the DRs is conserved among plant seed 11/12S globulins and, although their lengths vary, they generally have a highly hydrophilic character (Tandang-Silvas et al. 2010). Situating hydrophobic ISDs immediately adjacent to these DRs ensures that they are

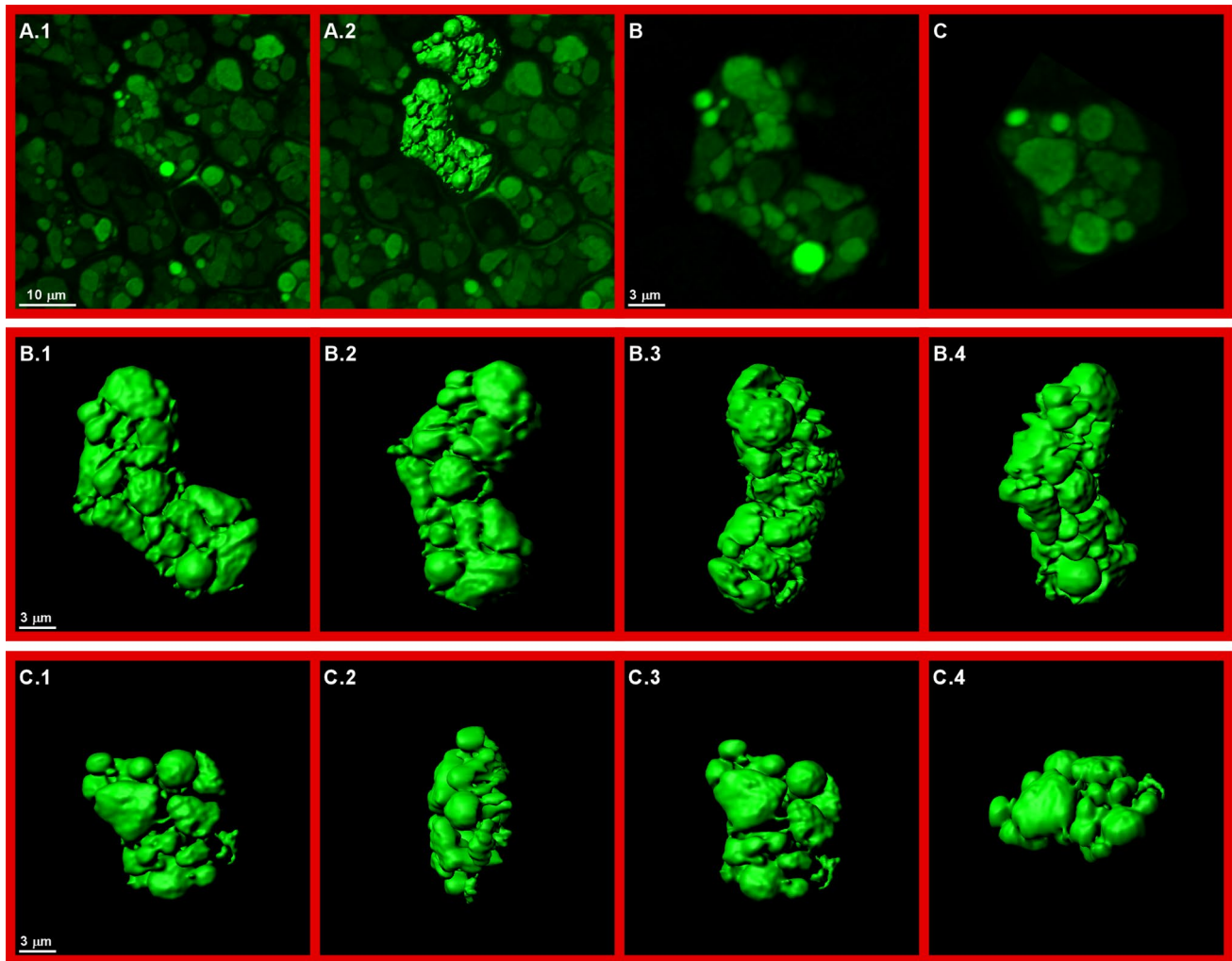


Fig. 5 Confocal microscopic images of embryo cells from a cruciferin-deficient (Cru-) *A. thaliana* line expressing YFP fused to ISD4. *Panel A* Two dimensional field of view (A.1) and three dimensional rendering of combined Z stacks from two individual cells (A.2).

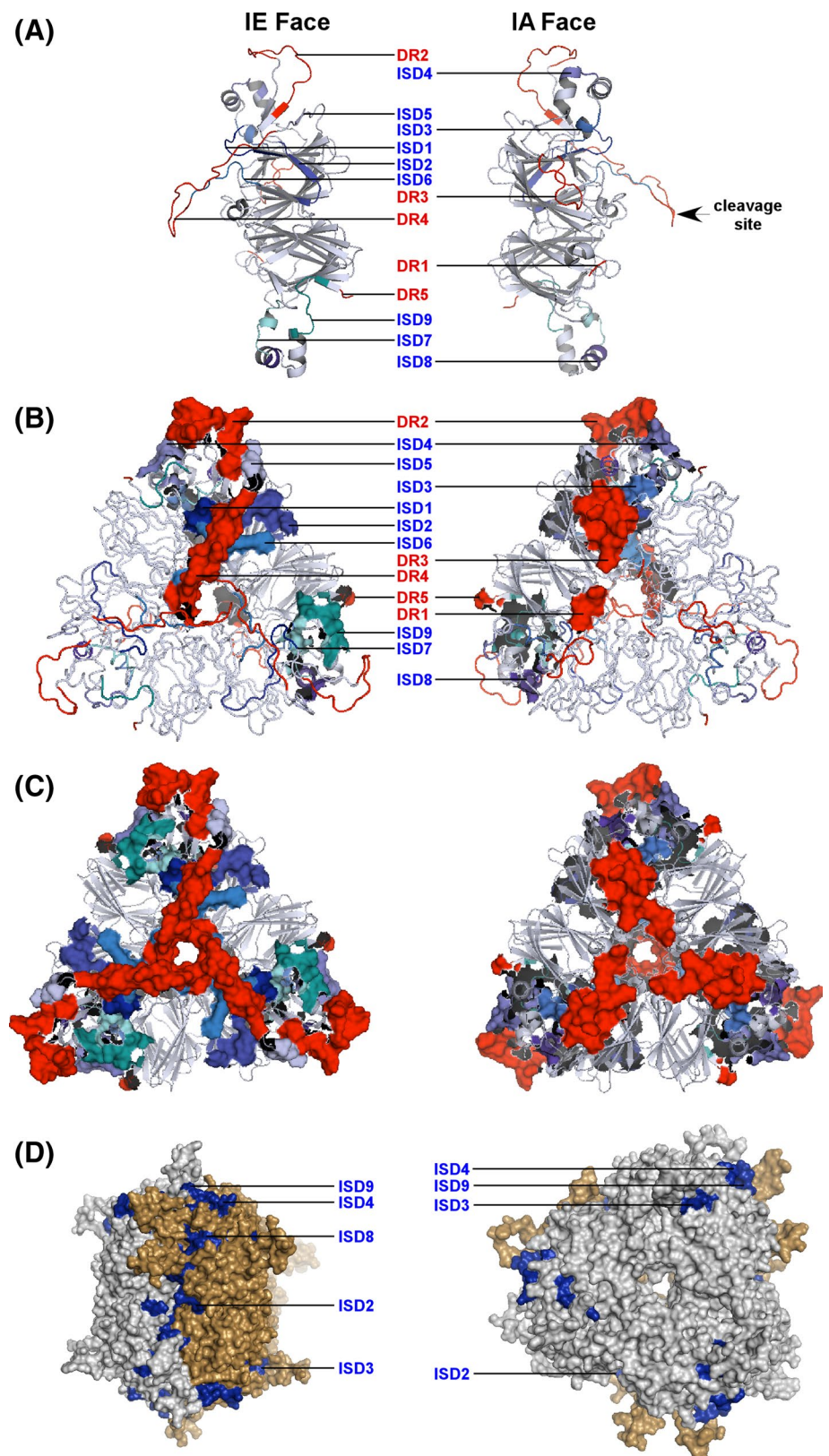
Panels B.1–4 and C.1.4: Three dimensional rendering of same cells presented from different viewing angles to show the protein storage vacuole network. See also Supplemental Movie 1

displayed on the surface of the molecule and are accessible to sorting receptors. The presence of multiple ISDs in addition to a carboxy terminal sorting determinant indicates that trafficking of cruciferin to the PSV is a cooperative process. This may be necessary given the large volume of SSPs being trafficked through ER and associated intermediate vesicles/bodies at the time of seed filling.

Three dimensional modeling of AtCruA revealed some additional interesting features about the location of the ISDs. Of the nine ISDs, seven (ISD1, ISD2, ISD4, ISD5, ISD6, ISD7 and ISD9) are exposed on the interchain disulfide-containing (IE) face of the trimer, while three (ISD3, ISD4 and ISD8) are exposed on the intrachain disulfide-containing (IA) face (Fig. 6a–c; Supplemental Figure 1). Hexamer formation occurs through interaction of the IE faces, but only after proteolytic processing of

the propeptides that comprise the trimer into the large and small globulin subunits. This is accompanied by the movement of a mobile region (DR4) away from the IE face to allow trimer–trimer interaction (Adachi et al. 2003). Trimer assembly occurs in the ER (Chrispeels et al. 1982) where they are then directed into DVs that bud from the Golgi periphery (Robinson et al. 2005; Vitale and Hinz 2005). The enzymes responsible for the proteolytic processing of the propeptides, namely β -vacuolar processing enzyme and aspartic protease A1 (Jung et al. 1998; Shimada et al. 2003b), are packaged by the ER/Golgi apparatus into separate vesicles that eventually fuse with the DVs to form MVBs en route to the PSV (Otegui et al. 2006). As such, trimer processing and hexamer formation likely occur in the MVB or after deposition into the PSV. In the trimer, the IE and IA faces are exposed and available to interact

Fig. 6 Homology model of AtCruA showing disordered regions (DR, shades of red) and internal sorting determinants (ISD, shades of blue). Darker shades are nearer to the top surface in each field of view. **a** Single AtCruA subunit. **b** AtCruA homotrimer with disordered regions and internal sorting determinants depicted in space fill mode for one subunit. **c** AtCruA homotrimer with disordered regions and internal sorting determinants shown in all subunits. The Interchain disulfide-containing (IE) face is shown on the left and the Intra-chain disulfide-containing IA face on the right in all panels. The site of protease cleavage that releases the α and β cruciferin chains is indicated by an arrowhead. **d** Location of ISDs on AtCruA homohexamer as viewed from the side and top of the molecule. Figures showing the location of individual ISDs can be found in Supplemental Figure 1



with receptors, be they VSR or RMR receptors. Many of the ISDs residing on the IE face would be completely or partially masked and once the hexamer has formed. This suggests that once SSPs exit the Golgi and enter the DVs, sorting determinant-receptor interactions may be less crucial for sorting than during ER and Golgi transit and that those determinants that remain exposed are sufficient to direct the protein to the PSV. Indeed, sorting determinants have been shown to act cooperatively (Holkeri and Vitale, 2001) in this regard. Alternatively, once cruciferin exits the ER or Golgi it may follow a path committed to deposition within the PSV.

As with the NPIR motif near the amino terminus of proteins trafficked to LVs (Matsuoka and Neuhaus, 1999), the Ile in the ISDs is believed to be critical for sorting to the PSV as substitution of this residue inactivated the determinant in glycinin (Maruyama et al., 2006), conglycinin (Nishizawa et al., 2006), ricin (Frigerio et al., 2001) and sporamin (Matsuoka and Neuhaus 1999). Leu is also critical for the *R. communis* 2S albumin sorting determinant as substitution of this amino acid within the LRMP sequence severely compromised vacuolar targeting (Brown et al. 2003). Our study revealed that Leu may also reside at the core of the determinant as several cruciferin ISDs contained a Leu in the central position, namely ISD6 (EETLCTMR) in AtCruB, and ISD7 (TSVLRGLP) and ISD9 (FNTLETTL) in AtCruA. However, comparison of the regions analogous to the nine AtCruA ISDs in other 11S globulins revealed that not only long chain aliphatic amino acids, but other hydrophobic amino acids such as Val, Met and Phe might also substitute in this position (Supplemental Table 7), though these putative ISDs have not been tested. Furthermore, our mutation analysis showed that Ala, a moderately hydrophobic aliphatic amino acid, could also substitute in this position, at least in ISD3. These comparisons and observations indicate that the core of the ISD can accommodate a much wider range of hydrophobic amino acids than previously thought possible. Deletion analysis based on ISD3 confirmed the statistical analysis which predicted that the minimal length of a fully functional ISD was approximately eight residues. Three of these were amino terminal to the core hydrophobic residue and four were carboxy terminal. Deletion of even one of the four carboxyl terminal residues profoundly affected ISD function.

Our coupled functional and statistical analysis based on AtCruA ISD regions and analogous regions in other 11S globulins indicated that more than one type of ISD may be present. Unlike the conserved NPIR motif (Matsuoka and Neuhaus 1999), a consensus has not yet emerged for either for ctSDs or ISDs that direct proteins to the PSV. The carboxy terminal tetrapeptide (AFVY) of the 7S globulin phaseolin (Frigerio et al. 1998) or the similar tripeptide (AFY) from conglycinin (Nishizawa et al. 2006) are sufficient for

sorting to the PSV. In 11S globulins, the carboxy terminal 10 amino acids (PQESQKRAVA) from glycinin (Maruyama et al. 2006) and the pentapeptide (KISIA) from amaranthin (Petruccioli et al. 2007) were also functional PSV sorting determinants. The carboxy terminal region of AtCruA (SYGRPRVAAA) has been implicated as a sorting determinant (Shimada et al. 2003a) and bears some resemblance to that of AtCruB (SPMSYGRPRA), but not AtCruC (QLIEEIVEA). Identifying a consensus for ISDs has been equally difficult. A sequence near the carboxy terminus of conglycinin with SIL at its core (Nishizawa et al. 2006), as well as the ICTMR sequence in glycinin (Maruyama et al. 2006) and the GNIFRGG in amaranthin (Petruccioli et al. 2007) which are near DRs, were declared to be ISDs. A region (LLIRP) within the internal propeptide of ricin also functions as a vacuolar sorting determinant in vegetative tissue (Frigerio et al. 2001). Based on the current and previous studies, cruciferin ISDs likely adhere to the following rules: (1) ISDs are adjacent to or within hydrophilic, surface exposed regions that serve to present them on the protein's surface; (2) ISDs generally have a hydrophobic character; (3) ISDs tend to have Leu or Ile residues at their core; (4) ISDs are approximately eight amino acids long with the physiochemical consensus [hydrophobic] [preferably charged][small or hydrophobic, but not tiny] [IL][polar, preferably charged][small, but not charged] [hydrophobic, not charged, preferably not polar][hydrophobic, not tiny, preferably not polar].

When viewed on a single plane by two-dimensional microscopy, the PSVs appear as the characteristic, discreet vacuoles. Three dimensional imaging of PSVs revealed them to be part of a semi-continuous system that might be better described as a protein storage vacuolar network. The PSV is unique to plants and therefore little can be inferred about its structure and biogenesis from studies in other eukaryotes. It is known to be a compound organelle consisting of a matrix containing mostly SSPs, a crystalloid comprising a membrane-derived lattice, and a globoid containing phytic acid crystals and hydrolytic enzymes (Jiang et al. 2000, 2001). Few studies have addressed PSV ontogeny. An early study by Craig (1986) revealed a fused network of cisternae within pea cotyledon cells that were often associated with electron-dense protein deposits. Subsequent work with developing cotyledons showed that PSVs derive from tubular, cisternal membranes that enlarge as they accumulate SSPs, eventually engulfing resident LVs which may form globoids (Hoh et al. 1995; Jiang et al. 2001). This ontogeny is consistent with our observation of an extended PSV network in mature cotyledons which may be a remnant of the cisternal membrane complex from which it originated. There appears to be a close association between protein trafficking and PSV formation/maturation. Ectopic expression of SSP in vegetative tissues (Hayashi

et al. 1999) and protoplasts induces the formation of structures capable of encapsulating them; however, these structures may be more similar to PAC vesicles than terminal PSVs. In *A. thaliana* plants, mutation of the *KAM2/GRV2* gene, which encodes a DnaJ domain-containing Receptor-Mediated Endocytosis-8 protein involved in the endocytotic pathway from the plasma membrane to lysosomes, leads to SSP secretion and the appearance of malformed PSVs (Tamura et al. 2007). In this mutant, the PSVs are larger and fused resulting in less segmentation within the network in comparison with the wild type. Further evidence is provided by the *A. thaliana vamp727/syp22* double mutant that produces a defective SNARE complex involved in fusion of the PVC and vacuole membranes. This mutant also mis-sorts SSPs and produces smaller, more fragmented PSVs than the wild type (Ebine et al. 2008). Taken together, these studies indicate that PSVs originate from pre-existing cis-ternal membranes (or at least ones that formed during early embryo development) which then expand to form a PSV network by fusing with PVCs and acquiring their cargo of SSPs.

In conclusion, this study has demonstrated that multiple sorting determinants are available to traffic cruciferin to the PSV. This may be necessary to ensure proper localization during the period of time when most of the seed's resources are being directed toward storage compound deposition. Furthermore, the PSV system was shown to be a highly complex, integrated structure that may originate by expansion of a pre-existing endosomal membrane network.

Materials and methods

Plasmid construction

Constructs to identify and characterize ISDs comprised the AtCruA signal peptide (MARVSSLLSFCLTLLILFHGYAA) fused to the N-terminus of enhanced yellow fluorescent protein (eYFP, Clontech Laboratories Inc., Mountain View, CA, USA) followed by putative sorting determinants from AtCruA and in one case AtCruB. Four or six glycine residues were appended to the C-terminus to mask any cryptic carboxy terminal protein sorting determinants as per Maruyama et al. (2006) and Petruccioli et al. (2007).

The constructs were generated using SOEing PCR (Horton et al. 1990). The sequences of all primers are provided in Supplemental Table 1. The first generation constructs used to identify putative ISDs used oligonucleotides of 121–226 bp in length that were cloned into pUC57 after synthesis (BIO BASIC Inc., Markham, ON, Canada). The oligonucleotide encoding the signal peptide also encoded the first seven amino acids of eYFP. The other oligonucleotides encoded the last seven amino acids of eYFP,

followed by the AtCruA region to be tested and four glycine residues. Each oligonucleotide was flanked by *BfuAI* restriction sites. PCR primers were generated by digestion with *BfuAI* which cuts precisely four nucleotides 3' (sense strand) and eight nucleotides 3' (anti-sense strand) from the recognition site leaving termini that were complementary to eYFP. The oligonucleotide encoding the AtCruA signal peptide also contained a *KpnI* restriction site internal to the *BfuAI*, while the oligonucleotides encoding the sorting determinants contained a *SacI* restriction site internal to the *BfuAI* to allow for subsequent subcloning of the constructs into a plant transformation vector. The first SOEing PCR reaction fused the signal peptide to eYFP, while the second reaction incorporated the ISD. The second set of constructs comprised the AtCruA signal peptide fused to the amino terminus of eYFP, followed by eight amino acids encoding a putative ISD fused to six glycine residues. Constructs were generated by amplifying the SP-YFP using the forward primer Pnap_CRA_F-Kpn-2 and construct-specific reverse primers which encoded the ISD to be tested and glycine residues followed by a *SacI* restriction site. Four to six bp extensions were included adjacent to the termini of the *SacI* and *KpnI* restriction sites to allow efficient digestion of the PCR product. The third set of constructs comprised the AtCruA signal peptide fused to the amino terminus of eYFP, followed by a modified ISD based on QKNIFNGF fused to six glycine residues. Two types of modifications were tested: (1) A series of alanine scanning mutants, and (2) a series of serial deletions of the QKNIFNGF sequence. Constructs were generated using the same methods as in the second set above.

For all constructs, the resultant PCR products were digested with *KpnI* and *SacI* and cloned into pMDC32-N in *Escherichia coli* DH10B (Invitrogen). pMDC32-N is a derivative of pMDC32 (Curtis and Grossniklaus 2003) in which the CaMV 35S promoter was replaced with the seed-specific *napin* promoter (1934 bp upstream of the start codon of At4g27160). Napin and cruciferin are the main SSP in *A. thaliana* and the genes encoding them have very similar expression patterns according to the ABRC microarray database (Zimmermann et al., 2004). The plasmids were then introduced into *Agrobacterium tumefaciens* GV3101 pMP90 for plant transformation.

Plant material and transformation

Arabidopsis thaliana ecotype Columbia and a cruciferin deficient line were propagated under controlled conditions (16-h ~800 W m² light at 21 °C and 8-h dark at 16 °C). To generate the cruciferin deficient plant line the following T-DNA insertion lines obtained from either the Arabidopsis Biological Resources Centre (www.abrc.osu.edu) or GABI-Kat (www.gabi-kat.de) were used: *cruACRUBCRUC*

(SALK 002668, T-DNA inserted in the second exon of At5g44120), *CRUAcruBCRUC* (SALK 045987, T-DNA inserted in the fourth exon of At1g03880) and *CRU-ACRUBcruC* (GK-283D09, T-DNA inserted in the first exon of At4g28520). A homozygous triple-knockout line (*cruAcruBcruC* or *CRU-*) was obtained by several rounds of conventional crossing and selection with appropriate PCR primers (Withana-Gamage et al. 2013).

Arabidopsis thaliana plants were transformed using the floral dip method (Clough and Bent 1998) modified to include a 30 s exposure to vacuum (25 mmHg) immediately after dipping. This was found to reduce the tissue damage observed when the vacuum was applied during the dipping procedure (Bechtold et al. 1993) and improved transformation efficiency. Plants transformed with the neomycin phosphotransferase (NptII) selection marker were identified as follows. Seeds were sterilized (incubated with shaking in 95 % ethanol for 5 min, 30 % bleach + 0.05 % Tween 20 for 15 min, and rinsed 5 times with sterile water), and then placed at 4 °C for 2–5 days in sterile 0.1 % agar. Sterile seeds were germinated and grown on ½ MS agar (0.22 % Murashige and Skoog basal salts, 0.10 % MES, 1 % sucrose, 0.7 % agar, pH 5.7), with 300 µg/ml Timentin and 50 µg/ml kanamycin for 7–20 days.

Epi-fluorescence microscopy

Epi-fluorescence microscopy was used to evaluate the localization of the various cruciferin-YFP fusion constructs. Mature *A. thaliana* seeds were soaked in ½ MS for 2–3 h. The embryos were removed from the seed coat by gently pressing them between a slide and a cover slip in ½ MS and transferred to a drop of ½ MS on a new slide for live cell imaging. Embryos were viewed on a Zeiss Imager.Z1 compound microscope outfitted with an Apotome. A 40X apochromat water-immersion lens (NA 1.2) was used in conjunction with the YFP filter (excitation wavelengths 490–510 nm, emission wavelengths collected 520–550 nm) and a DIC lens (for bright field images). Autofluorescence was negligible as it only became visible with exposure times over 3 s (generally 4–5 s) and most images of YFP were taken with exposure times of 1.5 s or less. Control images of the non-transformed line were taken over a range of settings; however, at the exposure times used to visualize YFP in PSV these were dark fields. Exposure times for every sample were noted since the levels of expression varied between independent events/lines per construct and poorly expressing lines were excluded from the analysis to avoid long exposure times. Images were viewed and exported in TIFF format using Axiovision 4.8.2. For each construct, five cotyledons from each of three independent transformed lines were viewed. If a consensus phenotype was not obvious, two additional independent lines were assessed.

Confocal microscopy

Confocal microscopy was used to visualize PSVs in three dimensions. Mature *A. thaliana* seeds containing the SP-YFP-DR5 + N' construct which is completely localized to the PSV were soaked in ½ MS for 2–3 h. The embryos were removed from the seed coat by gently pressing them between a slide and a cover slip in ½ MS and transferred to a drop of ½ MS on a new slide for live cell imaging. Embryos were viewed on a Zeiss LSM 710 compound microscope outfitted with a tunable Sapphire laser (set to 720 nm) and an Argon laser (458, 488, and 514 nm laser lines). The YFP signal was detected using 514 nm excitation and a 518–621 nm emission filter. PSV autofluorescence was detected using 2-photon imaging with 720 nm excitation, a long pass 480 nm filter and a non-descanned detector. In the cruciferin-deficient line, the SP-YFP-DR5+ N' signal co-localizes with PSV autofluorescence. At 720 nm excitation, only PSV autofluorescence was detected, while the YFP was not excited, allowing exclusive imaging of PSVs. A 40X apochromat water-immersion lens (NA 1.1) was used to obtain a Z-stack of 45 high quality images (pixel size 0.159 × 0.159 × 0.3 µm, averaging 4 scans per line to reduce noise). Images were deconvolved using AutoQuant X2 (blind PSF, 10 cycles, medium noise). At 720 nm excitation with a lens numerical aperture (NA) of 1.1 and a condenser NA of 0.55, the theoretical resolution was 0.5 µm (Inoue 2006). Imaris 7.2.3 software was used to create 3D images from the Z-stack data.

Protein homology modeling

The theoretical structure of AtCruA was predicted using the SWISS-MODEL server (Arnold et al. 2006; swissmodel.expasy.org). The template identification tool was used to identify the best template for modeling AtCruA. The settings used for the InterPro Domain Scan were: HMMPfam, HMMTigr, ProfileScan, SuperFamily, BlastProDom. The settings for the Gapped Blast Query were: E-value cut off (0.000001), Matrix [Blosom62, 11(G), 1E] and Alignments (50). Chain A from the crystal structure of procruciferin, 11S globulin from *Brassica napus* (PDB 3KGL) with a resolution of 2.98 angstroms (Tandang-Silvas et al. 2010) had the best alignment and was used as the template for homology modeling (Zdobnov and Apweiler 2001). The quaternary structure of AtCruA was assumed by the SWISS-MODEL server to be identical. The constructed model was structurally aligned to the template using SWISS-PDB Viewer (www.expasy.org/spdbv). Untemplated regions in the model, Arg125-Ser132 and Arg263-Glu286, were modeled using the ModLoop server (<http://modbase.compbio.ucsf.edu/modloop>). Energy minimization was performed with the Gromos96 implementation in SWISS-PDB. The

energy calculated for the AtCruA model was $-23,971$ kJ/mol. The trimer was assembled using the Matchmaker function in Chimera with the following default settings: Alignment algorithm (Needleman-Wunsch); Matrix (Blosum62, gap extension penalty 1, iterate by pruning long atom pairs until no pair exceeds); 2.0 angstroms. Imaging of the protein was done in Chimera (www.cgl.ucsf.edu/chimera).

Determination of ISD consensus

The physiochemical property of amino acid R-groups was assessed as per the scheme of Livingstone and Barton (1993). Secondary structure was determined using Predict Protein (www.predictprotein.org), as well as by homology modeling using either the *G. max* proglycinin (PDB:2d5f) or *Brassica napus* cruciferin (PDB:3kglA) as templates. Hidden Markov models (HMM) were developed using HMMER (Eddy, 1998) and searched against the TAIR (TAIR10_pep_20101214.fasta, downloaded June 14/2011 from ftp://ftp.arabidopsis.org/home/tair/Proteins/TAIR10_protein_lists/), *Brassica rapa* (Brapa_197_July_2014_annotated.fa, downloaded July 14/2014 from <http://www.plantgdb.org/XGDB/phplib/download.php?GDB=Br>), *Brassica oleracea* (B.oleracea_v1.0.scaffolds_proteins.txt, downloaded July 15/14 from <http://www.ocri-genomics.org/bolbase/login.htm>) and NCBI refseq protein (downloaded July 18, 2013 from <ftp://ftp.ncbi.nih.gov/blast/db/>) databases. The 11/12S globulins used were: *A. hypochondriacus* amaranthin (Genbank X82121); *A. thaliana* CruA (At5g44120), CruB (At1g03880), CruB' (At1g3890) and CruC (At4g28520); *Arachis hypogaea* Ara H 3 (PDB 3C3 V); *B. napus* cruciferin (PDB 3KGL); *Cucurbita maxima* 11S globulin (PDB 2E9Q); *G. max* glycinin A1aB1b (PDB 1FXZ) and A3B3 (PDB 2DH5); *Pisum sativum* legumin (PDB 3KSC).

Acknowledgments This work was supported with funding from the Agriculture and Agri-Food Canada Canadian Crop Genomics Initiative.

References

Adachi M, Takenaka Y, Gidamis AB, Mikami B, Utsumi S (2001) Crystal structure of soybean proglycinin A1aB1b homotrimer. *J Mol Biol* 305:291–305

Adachi M, Kanamori J, Masuda T, Yagasaki K, Kitamura K, Mikami B, Utsumi S (2003) Crystal structure of soybean 11S globulin: glycinin A3B4 homo-hexamers. *Proc Natl Acad Sci USA* 100:7395–7400

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201

Bechtold N, Ellis J, Pelletier F (1993) *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C R Acad Sci Paris Life Sci* 316:1194–1199

Brown JC, Jolliffe NA, Frigerio L, Roberts LM (2003) Sequence-specific, Golgi-dependent vacuolar targeting of castor bean 2S albumin. *Plant J* 36:711–719

Chrispeels MJ (1983) The Golgi apparatus mediates the transport of phytoheagglutinin to the protein bodies in bean cotyledons. *Planta* 158:140–151

Chrispeels MJ, Higgins TJ, Spencer D (1982) Assembly of storage protein oligomers in the endoplasmic reticulum and processing of the polypeptides in the protein bodies of developing pea cotyledons. *J Cell Biol* 93:306–313

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735–743

Craddock CP, Hunter PR, Szakacs E, Hinz G, Robinson DG, Frigerio L (2008) Lack of a vacuolar sorting receptor leads to non-specific missorting of soluble vacuolar proteins in *Arabidopsis* seeds. *Traffic* 9:408–416

Craig S (1986) Fixation of a vacuole-associated network of channels in protein-storing pea cotyledon cells. *Protoplasma* 135:67–70

Curtis MD, Grossniklaus U (2003) A Gateway cloning vector set for high-throughput functional analysis of genes in *planta*. *Plant Physiol* 133:462–469

Ebine K, Okatani Y, Uemura T, Goh T, Shoda K, Niihama M, Morita MT, Spitzer C, Otegui MS, Nakano A, Ueda T (2008) A SNARE complex unique to seed plants is required for protein storage vacuole biogenesis and seed development of *Arabidopsis thaliana*. *Plant Cell* 20:3006–3021

Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763

Frigerio L, de Virgilio M, Prada A, Faoro F, Vitale A (1998) Sorting of phaseolin to the vacuole is saturable and requires a short C-terminal peptide. *Plant Cell* 10:1031–1042

Frigerio L, Jolliffe NA, Di Cola A, Felipe DH, Paris N, Neuhaus JM, Lord JM, Ceriotti A, Roberts LM (2001) The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting. *Plant Physiol* 126:167–175

Frigerio L, Hinz G, Robinson DG (2008) Multiple vacuoles in plant cells: rule or exception. *Traffic* 9:1564–1570

Hanton SL, Brandizzi F (2006) Post-Golgi protein traffic in the plant secretory pathway. *Microsc Res Tech* 69:152–159

Hanton SL, Matheson LA, Chatre L, Rossi M, Brandizzi F (2007) Post-Golgi protein traffic in the plant secretory pathway. *Plant Cell Rep* 26:1431–1438

Hara-Nishimura I, Shimada T, Hatano K, Takeuchi Y, Nishimura M (1998) Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles. *Plant Cell* 10:825–836

Hayashi M, Toriyama K, Kondo M, Hara-Nishimura I, Nishimura M (1999) Accumulation of a fusion protein containing 2S albumin induces novel vesicles in vegetative cells of *Arabidopsis*. *Plant Cell Physiol* 40:263–272

Herman E, Schmidt M (2004) Endoplasmic reticulum to vacuole trafficking of endoplasmic reticulum bodies provides an alternate pathway for protein transfer to vacuoles. *Plant Physiol* 136:3440–3446

Hinz G, Hillmer S, Bäumer M, Hohl I (1999) Vacuolar storage proteins and the putative vacuolar sorting receptor BP-80 exit the Golgi apparatus of developing pea cotyledons in different transport vesicles. *Plant Cell* 11:1509–1524

Hinz G, Colanesi S, Hillmer S, Rogers JC, Robinson DG (2007) Localization of vacuolar transport receptors and cargo proteins in the Golgi apparatus of developing *Arabidopsis* embryos. *Traffic* 8:1452–1464

Hoh B, Hinz G, Jeong BK, Robinson DG (1995) Protein storage vacuoles form *de novo* during pea cotyledon development. *J Cell Sci* 108:299–310

- Hohl I, Robinson DG, Chrispeels MJ, Hinz G (1996) Transport of storage proteins to the vacuole is mediated by vesicles without a clathrin coat. *J Cell Sci* 109:2539–2550
- Holkeri H, Vitale A (2001) Vacuolar sorting determinants within a plant storage protein trimer act cooperatively. *Traffic* 2:737–741
- Horton RM, Cai ZL, Ho SN, Pease LR (1990) Gene splicing by overlap extension: Tailor-made genes using the polymerase chain reaction. *Biotechniques* 8:528–535
- Hunter PR, Craddock CP, Di Benedetto S, Roberts LM, Frigerio L (2007) Fluorescent reporter proteins for the tonoplast and the vacuolar lumen identify a single vacuolar compartment in Arabidopsis cells. *Plant Physiol* 145:1371–1382
- Inoue S (2006) Foundations of confocal scanned imaging in light microscopy. In: JB Pawley (ed) *Handbook of biological confocal microscopy*, 3rd Edn. Springer Science and Business Media, LLC.
- Jiang L, Phillips TE, Rogers SW, Rogers JC (2000) Biogenesis of the protein storage vacuole crystalloid. *J Cell Biol* 150:755–769
- Jiang L, Phillips TE, Hamm CA, Drozdowicz YM, Rea PA, Maeshima M, Rogers SW, Rogers JC (2001) The protein storage vacuole: a unique compound organelle. *J Cell Biol* 155:991–1002
- Jolliffe NA, Brown JC, Neumann U, Vicré M, Bachi A, Hawes C, Ceriotti A, Roberts LM, Frigerio L (2004) Transport of ricin and 2S albumin precursors to the storage vacuoles of *Ricinus communis* endosperm involves the Golgi and VSR-like receptors. *Plant J* 39:821–833
- Jolliffe NA, Craddock CP, Frigerio L (2005) Pathways for protein transport to seed storage vacuoles. *Biochem Soc Trans* 33:1016–1018
- Jung R, Scott MP, Nam YW, Beaman TW, Bassüner R, Saalbach I, Müntz K, Nielsen NC (1998) The role of proteolysis in the processing and assembly of 11S seed globulins. *Plant Cell* 10:343–357
- Livingstone C, Barton G (1993) Protein sequence alignments: a strategy for hierarchical analysis of residue conservation. *Comp Appl Biosci* 9:745–756
- Maruyama N, Mun LC, Tatsuhara M, Sawada M, Ishimoto M, Utsumi S (2006) Multiple vacuolar sorting determinants exist in soybean 11S globulin. *Plant Cell* 18:1253–1273
- Matsuoka K, Nakamura K (1999) Large alkyl side-chains of isoleucine and leucine in the NPIRL region constitute the core of the vacuolar sorting determinant of sporamin precursor. *Plant Mol Biol* 41:825–835
- Matsuoka K, Neuhaus J-M (1999) *Cis*-elements of protein transport to the plant vacuoles. *J Exp Biol* 50:165–174
- Mori T, Maruyama N, Nishizawa K, Higasa T, Yagasaki K, Ishimoto M, Utsumi S (2004) The composition of newly synthesized proteins in the endoplasmic reticulum determines the transport pathways of soybean seed storage proteins. *Plant J* 40:238–249
- Nishizawa K, Maruyama N, Satoh R, Fuchikami Y, Higasa T, Utsumi S (2003) A C-terminal sequence of soybean beta-conglycinin alpha' subunit acts as a vacuolar sorting determinant in seed cells. *Plant J* 34:647–659
- Nishizawa K, Maruyama N, Satoh R, Higasa T, Utsumi S (2004) A vacuolar sorting determinant of soybean beta-conglycinin beta subunit resides in a C-terminal sequence. *Plant Sci* 167:937–947
- Nishizawa K, Maruyama N, Utsumi S (2006) The C-terminal region of alpha' subunit of soybean beta-conglycinin contains two types of vacuolar sorting determinants. *Plant Mol Biol* 62:111–125
- Otegui MS, Herder R, Schulze J, Jung R, Staehelin LA (2006) The proteolytic processing of seed storage proteins in Arabidopsis embryo cells starts in the multivesicular bodies. *Plant Cell* 18:2567–2581
- Paris N, Neuhaus J-M (2002) BP-80 as a vacuolar sorting receptor. *Plant Mol Biol* 50:903–914
- Park M, Lee D, Lee G-J, Hwang I (2005) AtRMR1 functions as a cargo receptor for protein trafficking to the protein storage vacuole. *J Cell Biol* 170:757–767
- Park JH, Oufattole M, Rogers JC (2007) Golgi-mediated vacuolar sorting in plant cells: RMR proteins are sorting receptors for the protein aggregation/membrane internalization pathway. *Plant Sci* 172:728–745
- Petrucelli S, Molina MI, Lareu FJ, Circosta A (2007) Two short sequences from amaranth 11S globulin are sufficient to target green fluorescent protein and beta-glucuronidase to vacuoles in Arabidopsis cells. *Plant Physiol Biochem* 45:400–409
- Robinson DG, Bäumer M, Hinz G, Hohl I (1998) Vesicle transfer of storage proteins to the vacuole: the role of the Golgi apparatus and multivesicular bodies 152:659–667
- Robinson DG, Oliviusson P, Hinz G (2005) Protein sorting to the storage vacuoles of plants: a critical appraisal. *Traffic* 6:615–625
- Shen Y, Wang J, Ding Y, Lo SW, Gouzerh G, Neuhaus J-M, Jiang L (2011) The rice RMR1 associates with a distinct prevacuolar compartment for the protein storage vacuole pathway. *Mol Plant* 4:854–868
- Shewry PR, Napier JA, Tatham AS (1995) Seed storage proteins: structures and biosynthesis. *Plant Cell* 7:945–956
- Shimada T, Kuroyanagi M, Nishimura M, Hara-Nishimura I (1997) A pumpkin 72-kDa membrane protein of precursor-accumulating vesicles has characteristics of a vacuolar sorting receptor. *Plant Cell Physiol* 38:1414–1420
- Shimada T, Watanabe E, Tamura K, Hayashi Y, Nishimura M, Hara-Nishimura I (2002) A vacuolar sorting receptor PV72 on the membrane of vesicles that accumulate precursors of seed storage proteins (PAC vesicles). *Plant Cell Physiol* 43:1086–1095
- Shimada T, Fuji K, Tamura K, Kondo M, Nishimura M, Hara-Nishimura I (2003a) Vacuolar sorting receptor for seed storage proteins in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 100:16095–16100
- Shimada T, Yamada K, Kataoka M, Nakaune S, Koumoto Y, Kuroyanagi M, Tabata S, Kato T, Shinozaki K, Seki M, Kobayashi M, Kondo M, Nishimura M, Hara-Nishimura I (2003b) Vacuolar processing enzymes are essential for proper processing of seed storage proteins in *Arabidopsis thaliana*. *J Biol Chem* 278:32292–32299
- Tamura K, Takahashi H, Kunieda T, Fuji K, Shimada T, Hara-Nishimura I (2007) Arabidopsis KAM2/GRV2 is required for proper endosome formation and functions in vacuolar sorting and determination of the embryo growth axis. *Plant Cell* 19:320–332
- Tandang-Silvas MR, Fukuda T, Fukuda C, Prak K, Cabanos C, Kimura A, Itoh T, Mikami B, Utsumi S, Maruyama N (2010) Conservation and divergence on plant seed 11S globulins based on crystal structures. *Biochim Biophys Acta* 1804:1432–1442
- Vitale A, Hinz G (2005) Sorting of proteins to storage vacuoles: how many mechanisms? *Trends Plant Sci* 10:316–323
- Wan L, Ross ARS, Yang J, Hegedus DD, Kermod AR (2007) Phosphorylation of 12S globulin cruciferin in *Arabidopsis thaliana* seeds of wild type and *abi-1* mutant. *Biochem J* 404:247–256
- Wang J, Tse YC, Hinz G, Robinson DG, Jiang L (2012) Storage globulins pass through the Golgi apparatus and multivesicular bodies in the absence of dense vesicle formation during early stages of cotyledon development in mung bean. *J Exp Bot* 63:1367–1380
- Withana-Gamage TS, Hegedus DD, Qiu X, Wanasundara J (2011) *In silico* homology modeling to predict protein functional properties of cruciferin. *J Agric Food Chem* 59:12925–12938
- Withana-Gamage TS, Hegedus DD, Qui X, Yu P, May T, Lydiat D, Wanasundara JPD (2013) Characterization of *Arabidopsis thaliana* lines with altered seed storage protein profiles using synchrotron powered FTIR. *J Agric Food Chem* 61:901–912

- Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847–848
- Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol* 136:2621–2632
- Zouhar J, Rojo E (2009) Plant vacuoles: where did they come from and where are they heading? *Curr Opin Plant Biol* 12:677–684
- Zouhar J, Munoz A, Rojo E (2010) Functional specialization within the vacuolar sorting receptor family: VSR1, VSR3 and VSR4 sort vacuolar storage cargo in seeds and vegetative tissues. *Plant J* 64:577–588