Structural insights into how vacuolar sorting receptors recognize the sorting determinants of seed storage proteins

Hsi-En Tsao[a,1], Shu Nga Lui[a,1], Anthony Hiu-Fung Lo[a,1], Shuai Chen[a], Hiu Yan Wong[a], Chi-Kin Wong[a], Liwen Jiang[b,c,d,e], and Kam-Bo Wong[a,2]

[a]Centre for Protein Science and Crystallography, School of Life Sciences, The Chinese University of Hong Kong (CUHK), Shatin, Hong Kong, China; [b]Centre for Cell and Developmental Biology, School of Life Sciences, CUHK, Shatin, Hong Kong, China; [c]State Key Laboratory of Agrobiotechnology, School of Life Sciences, CUHK, Shatin, Hong Kong, China; [d]Institute of Plant Molecular Biology and Agricultural Biotechnology, CUHK, Shatin, Hong Kong, China; and [e]Shenzhen Research Institute, CUHK, Shenzhen 518057, China

Edited by Natasha Raikhel, Department of Botany and Plant Science and Center for Plant Cell Biology, University of California, Riverside, CA; received June 22, 2021; accepted November 17, 2021

In *Arabidopsis*, vacuolar sorting receptor isoform 1 (VSR1) sorts 12S globulins to the protein storage vacuoles during seed development. Vacuolar sorting is mediated by specific protein–protein interactions between VSR1 and the vacuolar sorting determinant located at the C terminus (ctVSD) on the cargo proteins. Here, we determined the crystal structure of the protease-associated domain of VSR1 (VSR1-PA) in complex with the C-terminal pentapeptide ($_{468}$RVAAA$_{472}$) of cruciferin 1, an isoform of 12S globulins. The $_{468}$RVA$_{470}$ motif forms a parallel β-sheet with the switch III residues ($_{127}$TMD$_{129}$) of VSR1-PA, and the $_{471}$AA$_{472}$ motif docks to a cradle formed by the cargo-binding loop ($_{95}$RGDCYF$_{100}$), making a hydrophobic interaction with Tyr99. The C-terminal carboxyl group of the ctVSD is recognized by forming salt bridges with Arg95. The C-terminal sequences of cruciferin 1 and vicilin-like storage protein 22 were sufficient to redirect the secretory red fluorescent protein (spRFP) to the vacuoles in *Arabidopsis* protoplasts. Adding a proline residue to the C terminus of the ctVSD and R95M substitution of VSR1 disrupted receptor–cargo interactions in vitro and led to increased secretion of spRFP in *Arabidopsis* protoplasts. How VSR1-PA recognizes ctVSDs of other storage proteins was modeled. The last three residues of ctVSD prefer hydrophobic residues because they form a hydrophobic cluster with Tyr99 of VSR1-PA. Due to charge–charge interactions, conserved acidic residues, Asp129 and Glu132, around the cargo-binding site should prefer basic residues over acidic ones in the ctVSD. The structural insights gained may be useful in targeting recombinant proteins to the protein storage vacuoles in seeds.

protein–protein interaction | crystal structure | vacuolar sorting determinant | storage protein | receptor-mediated trafficking

During seed development, storage proteins are deposited in a specialized organelle called the protein storage vacuole (PSV) and are mobilized to provide sources of carbon, nitrogen, and sulfur during germination (1). Seed storage proteins are synthesized as secretory proteins that are translocated into the endoplasmic reticulum (ER). How these proteins are transported to the PSV is not fully understood. In the receptor-mediated sorting pathway, storage proteins are sorted to the vacuoles via sequence-specific interactions with transmembrane sorting receptors (2). According to the latest model, sorting receptors could pick up the cargo proteins as early as in the ER and transport them through the Golgi apparatus to the *trans*-Golgi network (TGN), which then matures into the prevacuolar compartment (PVC) and PSV (2). Alternatively, storage proteins are concentrated and aggregated at the periphery of the *cis*-Golgi, where the dense vesicles (DVs) are formed (3, 4). DVs later bud off and fuse with the PVC, which matures into the PSV. Receptor–cargo interaction could play a role in the aggregation of storage proteins. For example, removal of the C-terminal hydrophobic (AFVY) residues of phaseolin, a storage protein of French bean (*Phaseolus vulgaris*), abolished aggregation of phaseolin and missorted it to the extracellular space (5).

There are two families of sorting receptors, namely vacuolar sorting receptors (VSRs) and receptor-homology-transmembrane-RING-H2 (RMR) proteins (6–8). There are seven homologs of VSR and six homologs of RMR in the *Arabidopsis thaliana* genome. Unlike lysosomal sorting in animal cells that recognizes the posttranslational modification of mannose-6-phosphate (9), vacuolar sorting in yeast and plant cells is mediated by specific protein–protein interactions between the sorting receptors and the cargo proteins (6, 10–19). These sorting receptors recognize sequence-specific information, or vacuolar sorting determinants (VSDs), on the cargo proteins (20, 21). There are two types of VSD, the sequence-specific VSD (ssVSD) and the C-terminal VSD (ctVSD) (22–24). VSRs can recognize both ssVSDs and ctVSDs (6, 15, 25), while RMRs can only recognize ctVSDs (26, 27). ssVSD, often found in acidic hydrolases targeting lytic vacuoles, contains an NPIR motif with the consensus sequence of (N/L)-(P/I/L)-(I/P)-(R/N/S) (28). Mutations in the

Significance

Seeds such as rice and soybean are major food staples in the human diet. During seed development, storage proteins are deposited in a specialized organelle called the protein storage vacuole and are mobilized to provide nutrients during germination. Storage proteins are transported as cargoes via specific protein–protein interactions with the vacuolar sorting receptors. Supported by structural and mutagenesis studies, our work provides insights into how the sequence-specific information, or the vacuolar sorting determinant, on the storage proteins is recognized by the vacuolar sorting receptors for their targeting to the vacuoles. Insights gained into the rules of receptor–cargo recognition will be useful in engineering recombinant proteins for biotechnological applications of the protein storage vacuoles in seeds.

Author contributions: H.-E.T., S.N.L., A.H.-F.L., C.-K.W., L.J., and K.-B.W. designed research; H.-E.T., S.N.L., A.H.-F.L., S.C., H.Y.W., and K.-B.W. performed research; H.-E.T., S.N.L., A.H.-F.L., and K.-B.W. analyzed data; and H.-E.T., S.N.L., A.H.-F.L., and K.-B.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]H.-E.T., S.N.L., and A.H.-F.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: kbwong@cuhk.edu.hk.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2111281119/-/DCSupplemental.

Published December 30, 2021.

NPIR motif disrupt receptor–cargo interactions and lead to missorting of cargo proteins (18, 21, 29, 30). Unlike ssVSD that is located at internal sequence positions, ctVSD is only found at the C terminus of cargo proteins. No consensus sequence has been identified for ctVSD, but it is usually rich in hydrophobic residues (20). For example, the AFVY motif at the C terminus of phaseolin was found to be essential for targeting seed proteins to the PSV (31).

VSRs are type I transmembrane proteins that contain a protease-associated (PA) domain, a central domain, and three epidermal growth factor (EGF) repeats in the luminal N-terminal region, followed by a single transmembrane domain (TMD) and a C-terminal cytoplasmic tail (Fig. 1*A*) (6, 12, 32). The PA domain and central domain are involved in sequence-specific interactions with ssVSDs (21, 33). We have previously determined the crystal structure of the PA domain of vacuolar sorting receptor isoform 1 (VSR1-PA) in complex with the ssVSD of barley aleurain (21) and showed that the PA domain is responsible for recognizing the sequences preceding the NPIR motif. Cargo binding induces the C-terminal tail to undergo a swivel motion that could relocate the central domain to cooperate with the PA domain for ssVSD recognition (21). The EGF repeats have unclear functions, but they might regulate the cargo binding by calcium-dependent conformational change of the PA and central domains (33, 34).

The role of VSRs in sorting seed storage proteins has been supported by genetic studies in *Arabidopsis*. In a pioneer study, Shimada and coworkers showed that the *vsr1* knockout mutant missorted the seed storage proteins 12S globulin and 2S albumin to the extracellular space in *Arabidopsis* seeds (15). Zouhar and coworkers further showed that *vsr1vsr3* and *vsr1vsr4* double mutants reduced the amount of the mature form of 12S globulin in the PSV, suggesting that VSR1, VSR3, and VSR4 are the sorting receptors for 12S globulin (35). Moreover, tagging the C-terminal 24 residues of β-conglycinin to the C terminus of a secretory green fluorescent protein (GFP) was sufficient to target the fluorescent protein to the PSV in *Arabidopsis* seeds, while the fluorescent protein was missorted to the extracellular space in the *vsr1* mutant (19). Since VSR1 can bind to the C-terminal sequence of both cruciferin 1 (CRU1), an isoform of 12S globulin, and β-conglycinin (15, 19), it is likely that VSR1 recognizes the sorting determinants in these sequences and sorts them to the PSV in seeds.

The molecular mechanism of how VSRs recognize the sorting determinants of cargo proteins remains elusive. In this study, we report the crystal structure of the PA domain of VSR1 in complex with the C-terminal pentapeptide ($_{468}$RVAAA$_{472}$) of CRU1. Structural insights into receptor–cargo interaction were supported by mutagenesis and functional studies, which showed that a specific recognition between the VSR and ctVSD is essential for vacuolar sorting.

Results

VSR1-PA Interacts with the C-Terminal Sequence of CRU1. It has been previously shown that *Arabidopsis* VSR1 sorts storage proteins in seeds by interacting with the C-terminal sequence of a 12S seed storage protein, CRU1 (15). How VSR1 recognizes the C-terminal sequence of CRU1 is not known. The N-terminal luminal region of VSR1 (VSR1-NT) responsible for cargo binding contains a PA, a central domain, and three EGF repeats (Fig. 1*A*). The PA domain is conserved in both VSRs and RMRs (*SI Appendix*, Fig. S4) and can recognize the ctVSD of soluble cargo proteins (15, 18, 19, 26, 27, 29, 36–38). We therefore hypothesized that the PA domain (residues 20 to 182) of VSR1 (VSR1-PA) is responsible for recognizing the C-terminal sequence of CRU1 and tested this hypothesis by in vitro pull-down assays. Recombinant VSR1-PA purified from *Escherichia coli* was loaded onto *N*-hydroxysuccinimide (NHS)–resins coupled with the C-terminal peptide of CRU1 or to control resins coupled with glycine (Fig. 1*B*). Our results showed that VSR1-PA was associated with the NHS-resins coupled with the C-terminal peptide of CRU1 but not with the control resins (Fig. 1*B*). Our results suggested that VSR1-PA can interact with the C-terminal sequence of CRU1.

Crystallographic Analysis Reveals How the C-Terminal Sequence of CRU1 Is Recognized by VSR1. To understand how VSR1-PA recognizes the C-terminal sequence of CRU1, we have determined the crystal structure of VSR1-PA in complex with the C-terminal pentapeptide ($_{468}$RVAAA$_{472}$) of CRU1. VSR1-PA and the C-terminal pentapeptide of CRU1 were cocrystallized under two conditions: at pH 6.5 and 9.0 (*SI Appendix*, Table S1). Diffraction data were collected for both conditions, and the structures were solved to 2.35 Å at pH 6.5 and 2.45 Å at pH 9.0 (*SI Appendix*, Table S1). The structure determined at pH 6.5 has better electron density for residues in the switch II region. Otherwise, the two structures are essentially identical, with an average Cα displacement value of 0.18 Å (*SI Appendix*, Fig. S1). In the crystal structures, the CRU1 C-terminal pentapeptide is fully accommodated by VSR1-PA, in which the cargo-binding loop and the switch III region constitute the cargo-binding site (Fig. 1*C*). The bound pentapeptide ($_{468}$RVAAA$_{472}$) is well-defined by electron density except for the side chain of the N-terminal arginine residue that was disordered. By forming backbone–backbone hydrogen bonds, $_{468}$RVA$_{470}$ of CRU1 forms a short parallel β-sheet with $_{127}$TMD$_{129}$ in the switch III region of VSR1 (Fig. 1*D*). The last two residues ($_{471}$AA$_{472}$) of CRU1 dock onto a cradle formed by the cargo-binding loop ($_{95}$RGDCYF$_{100}$). The interaction is stabilized by a backbone–backbone hydrogen bond with Phe100. In addition, the C-terminal carboxyl group of CRU1 is recognized by forming salt bridges with the side chain of Arg95 of VSR1 (Fig. 1*D*). The cargo-binding loop contains a $_{95}$RGxCxF$_{100}$ motif that is conserved in both VSRs and RMRs (Fig. 1*F* and *SI Appendix*, Fig. S4), suggesting that RMRs may use the same motif in binding ctVSD.

VSR1-PA Undergoes Major Conformational Changes in Switch I to III Regions upon Binding the C-Terminal Sequence of CRU1. The structure of VSR1-PA in complex with the CRU1 pentapeptide was compared with that of the apo form of VSR1-PA previously determined by us (21). The Cα displacement plot reveals that binding of the C-terminal sequence of CRU1 induces conformational changes in three regions of VSR1-PA, namely switches I, II, and III (*SI Appendix*, Fig. S2). In the apo form of VSR1-PA, the cargo-binding site is occupied by residues in the switch III region, where Glu133 of switch III forms salt bridges with the invariant Arg95 of the cargo-binding loop (Fig. 1*E*). Binding of the CRU1 sequence displaces the switch III residues from the cargo-binding site, causing the unwinding of helix 4 such that residues $_{127}$TMD$_{129}$ can form a parallel β-sheet with the C-terminal sequence of CRU1. Conformational changes are propagated to switch II, in which Asn46 of switch II moves toward and forms a hydrogen bond with Met128 of switch III (Fig. 1*E*). N-terminal residues $_{20}$RFVVE$_{24}$, which are originally located between switches II and III in the apo form, become disordered to make way for switch II and III interactions in the bound form of VSR1-PA. Switch I residues $_{25}$KNN$_{27}$ straighten up to extend the β-1 strand that forms an antiparallel β-sheet with $_{40}$YEC$_{42}$ of β-2. While similar conformational changes in switches I to III were also observed in VSR1-PA upon binding to the ssVSD of aleurain (21), the C-terminal residues (residues 176 to 182) of VSR1-PA, corresponding to the switch IV region in the VSR1-PA–ssVSD structure, was disordered in the VSR1-PA–CRU1 structure.

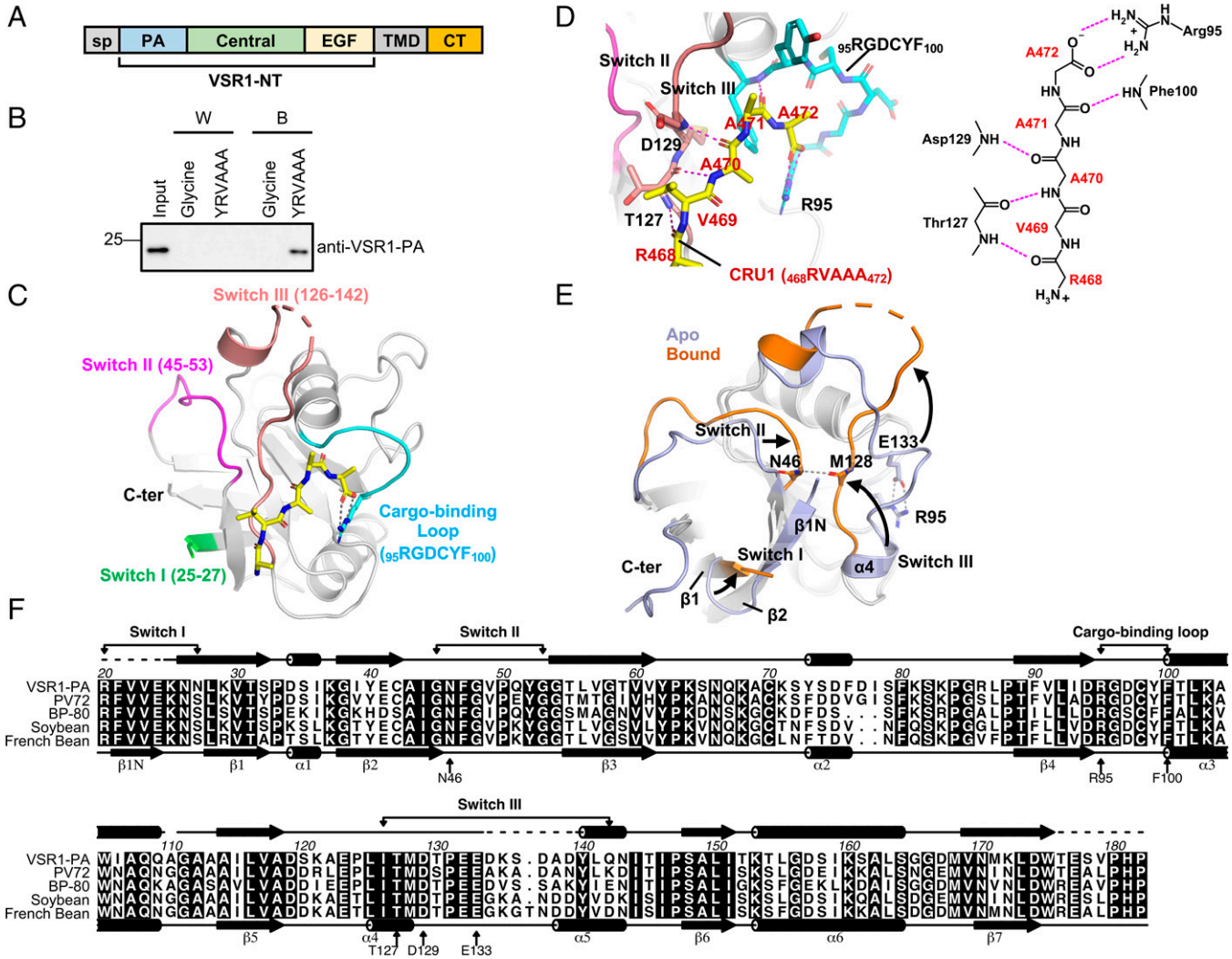

Fig. 1. Crystal structure of VSR1-PA in complex with the C-terminal pentapeptide ($_{468}$RVAAA$_{472}$) of CRU1. (*A*) Domain organization of VSRs. VSR1-NT consists of a protease-associated domain, a central domain, and three EGF repeats. sp, signal peptide. (*B*) Pull-down assay. *E. coli*–expressed VSR1-PA was incubated with NHS-resins coupled with the C-terminal peptide sequence of CRU1 (YRVAAA) or with glycine. A tyrosine residue was added to the N terminus of the peptide to facilitate the quantification of peptide concentration using A_{280}. After extensive washing to ensure VSR1-PA was not present in the last wash fractions (W), VSR1-PA bound (B) to the resins was analyzed by immunoblot with a VSR1-PA antibody. (*C*) Cartoon representation of the crystal structure of VSR1-PA in complex with the CRU1 C-terminal sequence, $_{468}$RVAAA$_{472}$ (yellow). Switch I, II, and III regions and the cargo-binding loop are color-coded green, magenta, salmon, and cyan, respectively. The complex structure determined at pH 6.5 is shown. (*D*) A close-up view of the detailed receptor–cargo interactions. $_{468}$RVA$_{470}$ of CRU1 forms a parallel β-sheet with $_{128}$TMD$_{129}$ of switch III. The last two residues of CRU1, $_{471}$AA$_{472}$, dock into a cradle formed by the conserved residues in the cargo-binding loop, $_{95}$RGDCYF$_{100}$. The backbone conformation of the bound cargo is maintained by a number of backbone–backbone hydrogen bonds (dotted lines). The C-terminal carboxyl group of CRU1 forms salt bridges with Arg95 in the cargo-binding loop. Intermolecular hydrogen bonds and salt bridges are summarized (*Right*). (*E*) VSR1-PA undergoes conformational changes upon binding of $_{468}$RVAAA$_{472}$. In the apo form of VSR1-PA (light blue), the cargo-binding site is occupied by switch III residues, where Glu133 forms salt bridges with Arg95. Cargo binding displaces the switch III residues away. N-terminal residues 20 to 24 that form strand β-1N in the apo form became disordered, making room for Asn46 in the switch II region to move toward and make a hydrogen bond with Met128 of switch III. Switch I residues (25 to 27) straighten up to form an antiparallel β-sheet with β-2. (*F*) Sequence of VSR1-PA, pumpkin PV72, pea BP-80, and soybean and French bean VSRs were aligned using the program MUSCLE (57). Secondary structure elements of the bound and apo forms of VSR1-PA are indicated above and below the alignment, respectively. Dotted lines indicate residues that are disordered in the crystal structures. Residues are numbered according to the VSR1-PA sequence.

Proline Substitutions in the C-Terminal Sequence of CRU1 Disrupt the Interaction with VSR1-PA. As structural analysis shows that the C-terminal pentapeptide sequence ($_{468}$RVAAA$_{472}$) of CRU1 interacts with VSR1-PA by forming a number of backbone–backbone hydrogen bonds and a short parallel β-sheet with the switch III residues of VSR1-PA (Fig. 1*D*), we hypothesized that proline substitutions in the C-terminal sequence of CRU1 will disrupt the backbone conformation and hydrogen bonds required for the interaction with VSR1-PA. To test this hypothesis, we introduced four proline substitutions into the C-terminal pentapeptide sequence of CRU1 and tested its interaction with VSR1-PA using an in vitro pull-down assay. Our results showed that VSR1-PA was pulled down by the C-terminal CRU1 peptide (YRVAAA) but not by the proline-substituted variant (YRPPPP) (Fig. 2*A*). Next, we performed a thermal shift assay by differential scanning fluorimetry (DSF) to test the interaction between VSR1-PA and the C-terminal sequence of CRU1 and its proline-substituted variant. The melting temperature (T_m) of VSR1-PA was significantly increased by 5.3 ± 0.7 °C when incubated with the YRVAAA peptide in the thermal shift assay (Fig. 2*C*), suggesting that VSR1-PA interacted with the C-terminal sequence of CRU1. In contrast, there was no significant change in the T_m of VSR1-PA

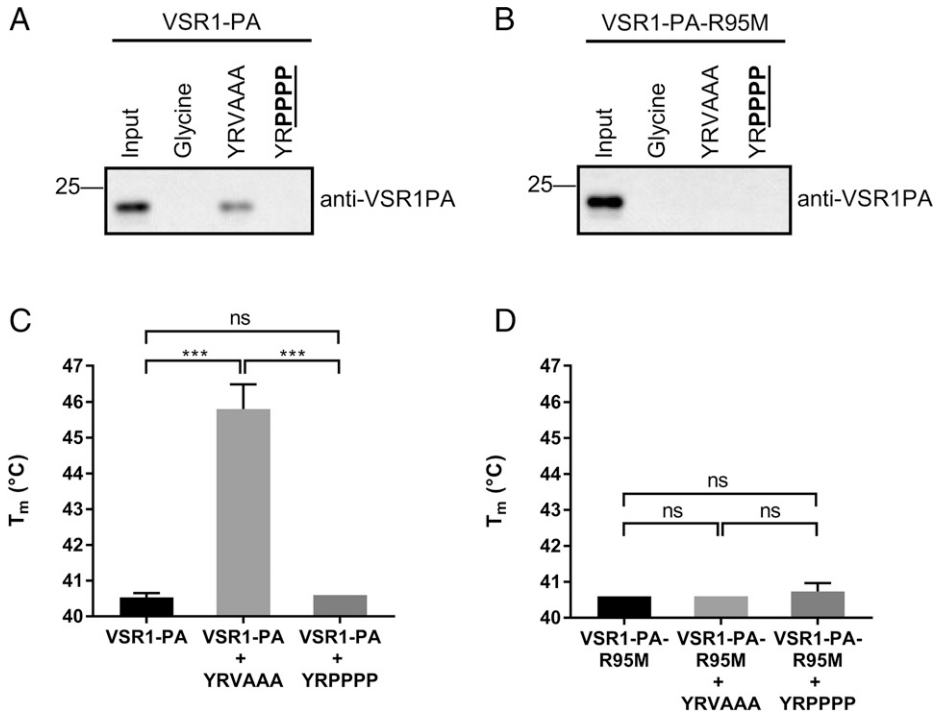

Fig. 2. Proline substitution of the C-terminal sequence of CRU1 and R95M of VSR1-PA disrupted receptor–cargo interaction. (*A* and *B*) Pull-down assay. *E. coli*–expressed VSR1-PA or VSR1-PA-R95M was incubated with NHS-resins coupled with the synthetic peptides from the C terminus of CRU1 (YRVAAA) and its proline substitution mutant (YRPPPP). After extensive washing, the presence of bound VSR1-PA was detected by immunoblot with a VSR1-PA antibody. (*C* and *D*) Thermal shift assay. Melting temperatures of VSR1-PA or VSR1-PA-R95M were measured with or without the addition of YRVAAA or YRPPPP by DSF. A significant increase of T_m was detected upon addition of the YRVAAA peptide but not with the addition of the YRPPPP peptide ($n = 3$; ***$P < 0.001$; not significant [ns], $P > 0.05$; error bars represent SD).

when it was incubated with the proline-substituted peptide (Fig. 2*C*). Taken together, our results suggest that the proline substitutions of the C-terminal sequence of CRU1 disturbed its interaction with VSR1-PA.

Arg95 on VSR1-PA Is Essential for Cargo Recognition. In the cargo-binding loop of VSR1-PA, the side chain of Arg95 forms salt bridges with the C-terminal carboxyl group of CRU1 (Fig. 1*D*). We argued that the salt bridges are crucial for VSR1-PA to recognize the C-terminal sequence of CRU1. To test this hypothesis, we substituted the positively charged Arg95 residue with methionine and tested if the R95M variant of VSR1-PA could interact with the C-terminal pentapeptide of CRU1. Our results showed that the R95M variant did not associate with NHS-resins coupled with the C-terminal peptide of CRU1 (Fig. 2*B*). Moreover, there was no significant change in the T_m of VSR1-PA-R95M when incubated with the C-terminal peptide of CRU1 in a thermal shift assay (Fig. 2*D*). These results suggest that breaking the salt bridges by R95M substitution disturbed receptor–cargo interactions between VSR1-PA and CRU1.

The C-Terminal Sequence of CRU1 Is a VSD. Next, we tested if the C-terminal sequence of CRU1 could serve as a ctVSD to promote sorting of cargo proteins to the vacuoles in plant cells. Secretory fluorescent protein fused with VSD sequences has been used as a reporter to test if VSD sequences can redirect the fluorescent protein to the vacuoles (19, 31, 39). We therefore transiently expressed secretory red fluorescent protein (spRFP) fused to the last 10 residues (SYGRPRVAAA) of the C-terminal sequence of CRU1 (spRFP-Cru1) in *Arabidopsis* protoplasts (Fig. 3*A*) and observed its subcellular localization with confocal microscopy (Fig. 3*B*). Compared with the confocal images of cells transformed with spRFP, the presence of the C-terminal sequence of CRU1 (spRFP-Cru1) redirected the spRFP's subcellular localization to the vacuoles in *Arabidopsis* protoplasts. In contrast, such a vacuolar sorting was disrupted when the last four residues of the CRU1 sequence were replaced by proline in the spRFP-Cru1PPPP (SYGRPRPPPP) variant (Fig. 3*B*). Results from the secretion assay further showed that spRFP-Cru1 was predominantly retained inside the protoplast, whereas the proline substitution increased the secretion of spRFP-Cru1PPPP into the culture medium (Fig. 3*C*). Our results suggest that the C-terminal sequence of CRU1 is sufficient to serve as a ctVSD that targets spRFP-Cru1 to the vacuoles.

ctVSD of CRU1 Is Recognized by VSR1-NT. Crystallographic analysis and mutagenesis experiments showed that the last four residues of CRU1 ctVSD are recognized by VSR1-PA (Figs. 1 and 2). We hypothesized that the ctVSD is also recognized by VSR1-NT. To address this hypothesis, we transiently expressed VSR1-NT in *Arabidopsis* protoplasts with a C-terminal T7 tag (21) and tested its interaction with the Cru1 peptide using a pull-down assay (Fig. 3*D*). Our results showed that VSR1-NT was pulled down by the wild-type C-terminal sequence of CRU1 but not its proline-substituted variant (Fig. 3*D*). We further showed that the interaction between VSR1-NT and the Cru1 peptide was disrupted by the R95M substitution in the cargo-binding loop (Fig. 3*D*). Our results on VSR1–ctVSD interaction using *Arabidopsis*-expressed VSR1-NT were consistent with those obtained using *E. coli*–expressed VSR1-PA, suggesting that the ctVSD of CRU1 is recognized by the PA domain of VSR1 in plant cells.

R95M Substitution of VSR1 Caused Missorting of ctVSD Cargos. As VSR1 can recognize the ctVSD of CRU1, we hypothesized that VSR1 is involved in the sorting of spRFP-Cru1 to the vacuoles. As previous reports suggested, overexpression of cargo binding–deficient mutants of VSR1 results in missorting and secretion of its cargo proteins (16, 21). To test the role of VSR1

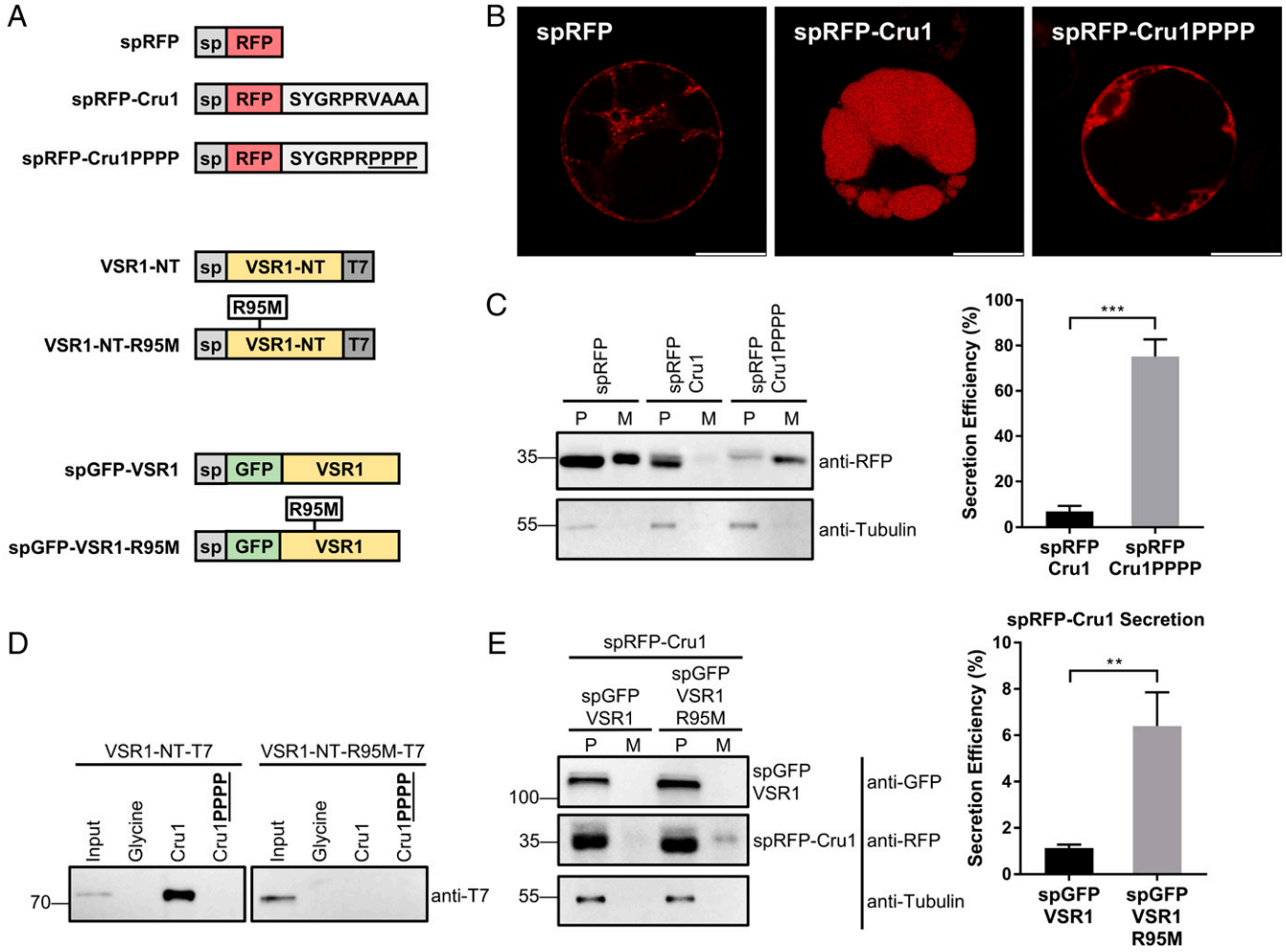

Fig. 3. C-terminal sequence of CRU1 functions as a ctVSD in *Arabidopsis* protoplasts. (*A*) Summary of the constructs used. Cru1: the C-terminal decapeptide sequence of CRU1. Cru1PPPP: the last four residues of Cru1 are replaced by proline. VSR1: full-length VSR1. (*B*) Confocal microscopy. spRFP, spRFP-Cru1, and spRFP-Cru1PPPP were transiently expressed in *Arabidopsis* protoplasts overnight before confocal imaging. (Scale bars, 20 μm.) Tagging of Cru1 at the C terminus induced vacuolar sorting of spRFP-Cru1, which was abolished by the proline substitutions in spRFP-Cru1PPPP. (*C*) Secretion assay. spRFP, spRFP-Cru1, and spRFP-Cru1PPPP were transiently expressed in *Arabidopsis* protoplasts. Protoplast fraction (P) and culture medium fraction (M) were collected and analyzed by SDS-PAGE and immunoblot with the respective antibodies. Cytosol leakage was monitored by the tubulin control. spRFP-Cru1 was predominantly retained in the protoplast, whereas spRFP-Cru1PPPP was largely secreted into the culture medium ($n = 3$; $^{***}P < 0.001$; error bars represent SD). (*D*) Pull-down assay. VSR1-NT and its variant VSR1-NT-R95M were transiently expressed in *Arabidopsis* protoplasts. After dialysis, the culture medium was mixed with NHS-resins coupled with Cru1 or Cru1PPPP peptides. Bound VSR1-NT and VSR1-NT-R95M were analyzed by SDS-PAGE and detected by immunoblot with a T7 antibody. R95M substitution on VSR1-NT or proline substitutions on the C-terminal Cru1 sequence abolished the receptor–cargo interactions. (*E*) R95M substitution of VSR1 increased the secretion of spRFP-Cru1. spRFP-Cru1 was coexpressed with GFP-tagged VSR1 or its R95M variant in *Arabidopsis* protoplasts. Protoplast fraction (P) and culture medium fraction (M) were separated after incubation and analyzed by SDS-PAGE and immunoblot with the respective antibodies. Cytosol leakage was monitored by the tubulin control. Coexpression of the R95M variant of VSR1 increased the secretion of spRFP-Cru1 ($n = 3$; $^{**}P < 0.01$; error bars represent SD).

in sorting the ctVSD of CRU1, we transiently coexpressed spRFP-Cru1 cargo protein with full-length wild-type or R95M-VSR1 tagged with GFP (spGFP-VSR1 and spGFP-VSR1-R95M) in *Arabidopsis* protoplasts (Fig. 3*E*). Our results showed that overexpression of the R95M mutant VSR1 increased the secretion of spRFP-Cru1 into the culture medium (Fig. 3*E*). Taken together, our results are consistent with the conclusion that the invariant Arg95 residue in the PA domain of VSR1 is involved in the specific interaction between VSR1 and ctVSD that sorts the cargo proteins to the vacuoles in *Arabidopsis* protoplasts.

Adding an Extra Proline Residue at the ctVSD Disrupted Receptor–Cargo Interaction and Caused Missorting of spRFP-Cru1. We tested the hypothesis that the C-terminal carboxyl group is essential for the receptor–cargo interaction by adding an extra proline residue to the C-terminal sequence of CRU1. In the crystal structure of the VSR1-PA–CRU1 complex, the C-terminal alanine residue adopts a backbone dihedral angle $\varphi = -152.7°$, which orientates its C-terminal carboxyl group to form salt bridges with Arg95 (Fig. 1*C*). We argued that addition of an extra proline residue, of which the ring structure restricts the φ dihedral angle to $\sim -60°$, would move the C-terminal carboxyl group away from the side chain of Arg95 and thus break the salt bridges. We showed that while addition of the YRVAAA peptide upshifted the T_m of VSR1-PA, there was no significant change in T_m upon addition of the YRVAAAP peptide (*SI Appendix*, Fig. S5*A*). This observation suggested that addition of the C-terminal proline disrupted the interaction between VSR1-PA and the ctVSD of CRU1 in vitro.

Next, we tested if *Arabidopsis*-expressed VSR1-NT can interact with the Cru1 peptide with an additional C-terminal proline residue (Cru1+P: SYGRPRVAAAP) with a pull-down assay. The addition of an extra C-terminal proline residue to Cru1

abolished its interaction with VSR1-NT (*SI Appendix*, Fig. S5*C*). We further added a C-terminal proline to spRFP-Cru1 to create the spRFP-Cru1+P variant (*SI Appendix*, Fig. S5*B*). When spRFP-Cru1+P was transiently expressed in *Arabidopsis* protoplasts, spRFP-Cru1+P could not reach the vacuoles (*SI Appendix*, Fig. S5*D*) and was largely secreted to the medium (*SI Appendix*, Fig. S5*E*). These results further supported that the C-terminal carboxyl group of CRU1 is essential for VSR–ctVSD interaction and for its vacuolar trafficking.

VSR1-NT Interacts with the C-Terminal Sequences of Cruciferin 4 and Vicilin-like Storage Protein 22. Cruciferin, like another well-studied seed storage protein, phaseolin, belongs to a vicilin/legumin protein family containing two cupin domains (40). We selected four isoforms of cruciferin (CRU1, CRU2, CRU3, CRU4) and three isoforms of vicilin-like storage proteins (VL21, VL22, VL43) in the *A. thaliana* genome for this study (Fig. 4*A*). To determine if these proteins carry ctVSDs that are recognized by VSR1, we coupled the C-terminal decapeptide sequences of these proteins to the NHS-resins and tested their interactions with *Arabidopsis*-expressed VSR1-NT in a pull-down assay (Fig. 4*B*). We also included the C-terminal sequence of phaseolin as a positive control. In addition to the previously identified ctVSD of CRU1, VSR1-NT interacted with the C-terminal sequences from CRU4 and VL22 (Fig. 4*B*).

We then further examined whether the two structural features mediating VSR1–ctVSD interaction, namely 1) Arg95 of VSR1 and 2) the backbone conformation of ctVSD, could be applied to the recognition of the C-terminal sequences of CRU4 and VL22. To this end, we introduced R95M substitutions to VSR1-NT and proline substitutions to the C-terminal sequences of CRU4 and VL22, and tested if these substitutions perturbed VSR1–ctVSD interactions by a pull-down assay (Fig. 4*C*). Consistent with the observations for VSR1–CRU1 interactions, both R95M substitution on VSR1-NT and proline substitutions on CRU4 and VL22 disrupted the interaction (Fig. 4*C*), suggesting the C-terminal sequences of CRU4 and VL22 are recognized by VSR1-NT using similar structural mechanisms.

ctVSD of VL22 Is Recognized by VSR1 for Vacuolar Sorting. To further test if the C-terminal sequences of CRU4 and VL22 can function as VSDs in plant cells, we fused the C-terminal decapeptide sequences of CRU4 and VL22 to the secretory RFP (spRFP-Cru4 and spRFP-VL22) and transiently expressed these constructs in *Arabidopsis* protoplasts. Our results showed that tagging the C-terminal sequence of CRU4 resulted in a partial redistribution of spRFP-Cru4 to the vacuoles, which was abolished by proline substitutions at the C terminus (spRFP-Cru4PPPP) (Fig. 4*D*). A large proportion of both spRFP-Cru4 and its proline-substituted variant, spRFP-Cru4PPPP, were secreted into the culture medium (Fig. 4*E*). Moreover, coexpression of spGFP-VSR1 or the spGFP-VSR1-R95M mutant did not affect the secretion of spRFP-Cru4 (Fig. 4*F*). Although the C-terminal sequence of CRU4 interacted with VSR1-NT in vitro, our results were not conclusive in supporting that CRU4 is a cargo protein recognized by VSR1.

On the other hand, tagging the C-terminal sequence of VL22, but not its proline-substituted variant (spRFP-VL22PPPP), redirected spRFP-VL22 to the vacuoles (Fig. 4*D*). The secretion assay showed that spRFP-VL22 was largely retained inside the protoplasts (Fig. 4*G*). In contrast, proline substitutions at the C terminus resulted in a significant increase in secretion of spRFP-VL22PPPP (Fig. 4*G*). We further showed that coexpression of spGFP-VSR1-R95M increased secretion of spRFP-VL22 into the culture medium (Fig. 4*H*). Taken together, our results suggest that the C-terminal sequence of VL22 served as a ctVSD that is recognized by VSR1 in *Arabidopsis* protoplasts.

P453A Substitution Converts Cru2 to a ctVSD that Is Recognized by VSR. In the crystal structure of VSR1-PA–CRU1, $_{468}$RVA$_{470}$ residues of CRU1 ctVSD form a parallel β-sheet with $_{127}$TMD$_{129}$ of the switch III region of VSR1-PA (Fig. 1*D*). In particular, the NH group of Ala470 at the antepenultimate position of CRU1 forms a hydrogen bond with the carbonyl group of Thr127 (Fig. 1*D*). It is noteworthy that the C-terminal sequence of Cru2, which could not interact with VSR1-NT, contains a proline residue (Pro453) at the antepenultimate position (Fig. 4 *A* and *B*). We hypothesized that Cru2 could not interact with VSR1-NT because Pro453 lacks the backbone NH group that could form a hydrogen bond with Thr127 of VSR1 and, therefore, the P453A variant of Cru2 should be recognized and sorted by VSR1.

We first tested if VSR1-NT could interact with the Cru2-P453A peptide by pull-down assay. We showed that VSR1-NT could be pulled down by Cru2-P453A but not by Cru2 (*SI Appendix*, Fig. S6*B*). Such interaction was disrupted by the R95M substitution in VSR1-NT. Next, we tested whether Cru2-P453A is a viable ctVSD sequence by fusing sequences of Cru2 or Cru2-P453A to the C terminus of secretory RFP (spRFP-Cru2 and spRFP-Cru2-P453A) (*SI Appendix*, Fig. S6*A*) and transiently expressing these constructs in *Arabidopsis* protoplasts to compare the subcellular localization and sorting in vivo. Consistent with the observation that Cru2 did not interact with VSR1-NT (Fig. 4*B*), spRFP-Cru2 did not reach the vacuoles (*SI Appendix*, Fig. S6*C*) and was largely secreted to the culture medium (*SI Appendix*, Fig. S6*D*). Coexpression with either VSR1 or its R95M variant did not impact the secretion level of spRFP-Cru2 (*SI Appendix*, Fig. S6*E*). On the contrary, spRFP-Cru2-P453A was sorted to the vacuoles and was largely retained in the protoplasts (*SI Appendix*, Fig. S6 *C* and *D*). The secretion level of spRFP-Cru2-P453A was also increased when it was coexpressed with the R95M variant of VSR1 (*SI Appendix*, Fig. S6*F*). Consistent with our structural insights on VSR1–ctVSD interaction, our observations support that the P453A substitution converts the C-terminal sequence of CRU2 to a ctVSD that is recognized by VSR1.

Acidic pH Weakens the Binding Affinity between VSR1-PA and ctVSD in vitro. Next, we tested if the binding affinity between VSR1-PA and ctVSD is pH-dependent. Thermal denaturation of VSR1-PA was followed by DSF in the presence of peptides (0 to 5 mM) derived from the ctVSD of CRU1 (YRVAAA) and VL22 (YSDRFV) at pH 7.1 and 6.0. DSF data were analyzed by isothermal analysis at 45 °C implemented in the web server FoldAffinity (41, 42) (https://spc.embl-hamburg.de/app/foldAffinity) to obtain the K_d values of VSR1-PA–ctVSD interactions (*SI Appendix*, Fig. S7 *A* and *B*). At pH 7.1, VSR1-PA bound YRVAAA (CRU1) and YSDRFV (VL22) peptides with K_d values of 2.6 ± 18 and 1.0 mM ± 20%, respectively. The K_d values for binding YRVAAA (CRU1) and YSDRFV (VL22) peptides increased to 9.6 ± 13 and 3.2 mM ± 9%, respectively, at pH 6.0 (*SI Appendix*, Fig. S7*B*), suggesting that acidic pH weakens the binding affinity between VSR1-PA and ctVSD in vitro.

Testing Whether ctVSD and ssVSD Could Compete for the Binding of VSR1-NT in vitro. We have previously shown that VSR1-PA binds to the residues (ADS) preceding the NPIR motif of aleurain-ssVSD (21). Here, we showed that VSR1-PA is also responsible for recognizing the ctVSD of Cru1. Taken together, both ctVSD and ssVSD could compete for the same binding site in the PA domain of VSR. To test this possibility, we immobilized VSR1-NT by loading it to NHS-resins coupled with aleurain-ssVSD (ADSNPIRPVT). Aleurain-ssVSD–bound VSR1-NT was eluted by the Cru1-ctVSD (SYGRPRVAAA) and aleurain-ssVSD peptides (*SI Appendix*, Fig. S8, *Upper*). In contrast, VSR1-NT was not eluted by the Cru1+P peptide (SYGRPRVAAAP), which could not interact with VSR1-NT

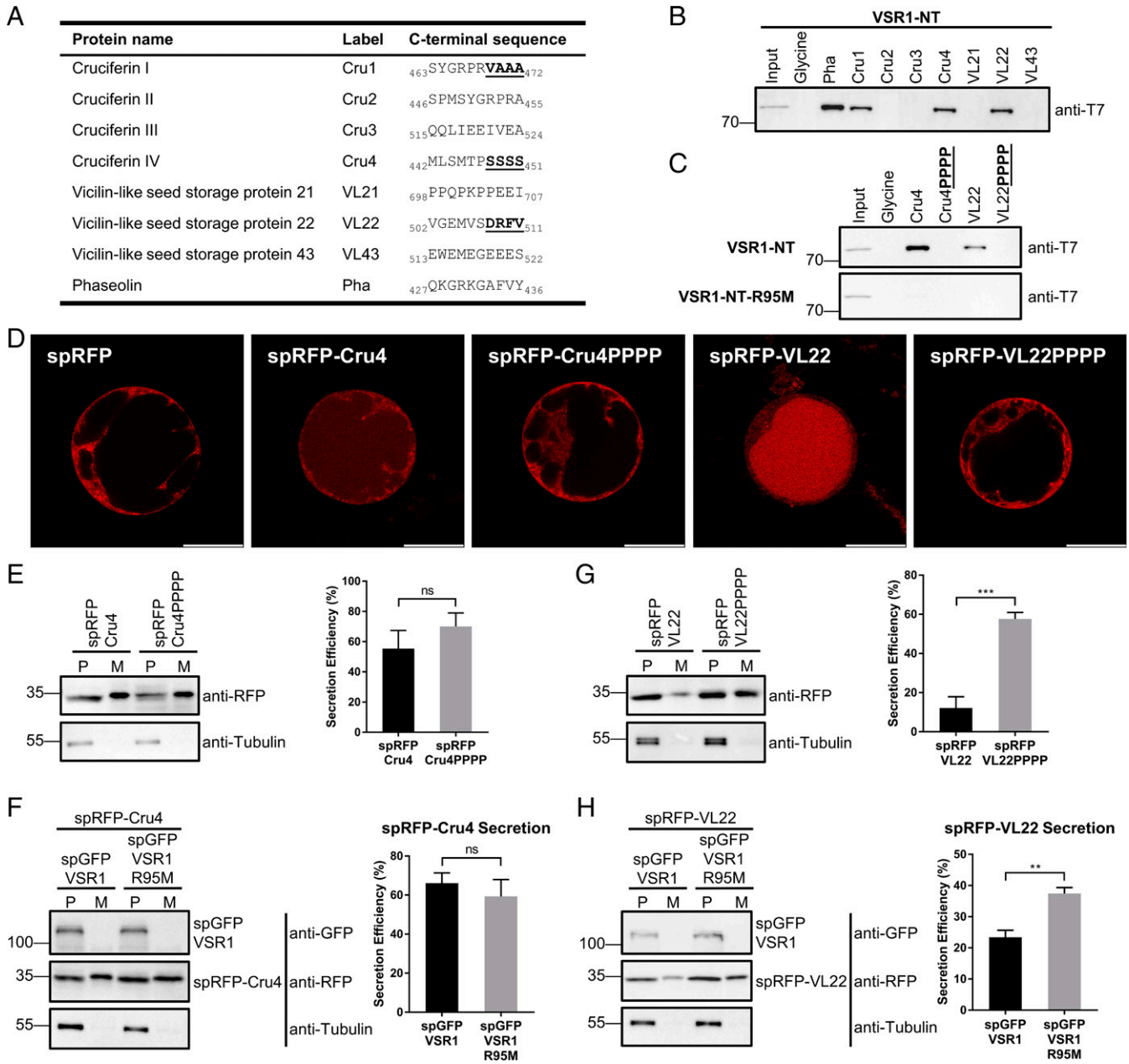

Fig. 4. Testing if VSR1 can recognize the C-terminal sequences of other vicilin/legumin-like proteins. (*A*) C-terminal sequences of other vicilin/legumin-like proteins in *Arabidopsis*. In this study, the last four residues (underlined) of Cru1, Cru4, and VL22 were replaced by proline to create the Cru1PPPP, Cru4PPPP, and VL22PPPP variants, respectively. (*B*) Pull-down assay. VSR1-NT was transiently expressed in *Arabidopsis* protoplasts. After dialysis, the culture medium was mixed with NHS-resins coupled with the C-terminal sequences of vicilin/legumin-like proteins listed in *A*. Bound AtVSR1-NT was analyzed by SDS-PAGE and detected by immunoblot using a T7 antibody. VSR1-NT was pulled down by C-terminal sequences from CRU4 and VL22. (*C*) VSR1-NT and VSR1-NT-R95M were transiently expressed in *Arabidopsis* protoplasts and tested for their interactions with the C-terminal sequences of CRU4 and VL22 and their proline-substituted variants by pull-down assay. The last four residues of the Cru4 and VL22 peptides were replaced by proline to create the Cru4PPPP and VL22PPPP variants. The pull-down assay was performed as described in *B*. (*D*) Confocal image analysis. The C-terminal decapeptides (or their proline-substituted variants) of CRU4 and VL22 were tagged with spRFP to create spRFP-Cru4, spRFP-Cru4PPPP, spRFP-VL22, and spRFP-VL22PPPP. These constructs were transiently expressed in *Arabidopsis* protoplasts overnight before confocal imaging. (Scale bars, 20 μm.) (*E*–*H*) Secretion assay. spRFP-Cru4, spRFP-VL22, and their proline-substituted variants were transiently expressed in *Arabidopsis* protoplasts (*E* and *G*) or coexpressed with spGFP-VSR1 or spGFP-VSR1-R95M (*F* and *H*). Protoplast fraction (P) and culture medium fraction (M) were collected and analyzed by SDS-PAGE and immunoblot with the respective antibodies. Cytosol leakage was monitored by the tubulin control. A large portion of spRFP-Cru4 and its proline-substituted variants was largely secreted into the medium (*E*). Moreover, coexpression with spGFP-VSR1 or spGFP-VSR1-R95M did not affect the secretion of spRFP-Cru4 (*F*). On the other hand, spRFP-VL22 was largely retained in the protoplasts, whereas spRFP-VL22PPPP showed a notable increase in secretion level (*G*). Coexpression of spGFP-VSR1-R95M and spRFP-VL22 notably increased the secretion of spRFP-VL22 (*H*) ($n = 3$; $^{**}P < 0.01$, $^{***}P < 0.001$; ns, $P > 0.05$; error bars represent SD).

(*SI Appendix*, Fig. S5*C*). These observations suggest that Cru1-ctVSD binds to the PA domain and displaces the aleurain-ssVSD from the binding site of VSR1-NT in vitro. On the other hand, Cru1-ctVSD–bound VSR1-NT was only eluted by ctVSD of CRU1 but not by aleurain-ssVSD and Cru1+P (*SI Appendix*, Fig. S8, *Lower*), suggesting that aleurain-ssVSD could not displace bound Cru1-ctVSD in vitro. To test if the vacuolar sorting of ssVSD and ctVSD cargos would interfere with each other,

we coexpressed secretory GFP tagged with N-terminal aleurain-ssVSD (sp-Aleu-GFP) (21) and spRFP-Cru1 in *Arabidopsis* protoplasts and observed their vacuolar sorting with confocal microscopy. We showed that coexpression of sp-Aleu-GFP and spRFP-Cru1 did not affect their trafficking to the vacuoles (*SI Appendix*, Fig. S9).

Discussion

Structural Insights into VSR–ctVSD Interactions. The main goal of this study was to understand the structural mechanism of how VSR recognizes the ctVSD often found in seed storage proteins. It has been long established that sorting determinants located at the C terminus of storage proteins are sufficient to target cargos to vacuoles. Genetic studies have showed that VSR1 is a sorting receptor for 12S globulins (15, 35). How the sequence information on ctVSD is "read" by VSR1 remains elusive. In this study, we showed that VSR1-PA is responsible for recognizing the ctVSD and determined the crystal structure of VSR1-PA in complex with the ctVSD ($_{468}$RVAAA$_{472}$) of CRU1, an isoform of 12S globulins. Structural insights were further supported by mutagenesis and functional studies, which suggest the following insights into VSR–ctVSD recognition:

1) The C-terminal carboxyl group of ctVSD is recognized by forming salt bridges with Arg95 of VSR1-PA. This structural insight is supported by the observations that R95M substitution of VSR1 (Fig. 3) and adding a proline residue to the ctVSD (*SI Appendix*, Fig. S5) abolished receptor–cargo interactions and vacuolar sorting of CRU1.
2) Pro453 of CRU2 prevents VSR–CRU2 interaction. The $_{468}$RVA$_{470}$ motif of CRU1 forms a parallel β-sheet with the $_{127}$TMD$_{129}$ motif of VSR1-PA. Since a proline residue lacks the backbone amide group and has a restricted backbone dihedral φ-angle (∼−60°) due to its ring structure, it is not favored at the C-terminal residues of the ctVSD. Supported by the observation that P453A substitution converts the C-terminal sequence of CRU2 to a ctVSD recognized by VSR1, we showed that CRU2 is not recognized and sorted by VSR1 because it contains a Pro453 at the antepenultimate position (*SI Appendix*, Fig. S6).
3) Hydrophobic residues at the last three residues of ctVSD form a hydrophobic cluster that can interact favorably with Tyr99. The complex crystal structure of VSR1-PA–ctVSD reveals that the last two residues of ctVSD could form direct hydrophobic interactions with Tyr99 of the cargo-binding loop of VSR1-PA (Fig. 5*A*), justifying the observation that ctVSD is often composed of mainly hydrophobic residues (15, 18, 24, 43–48). In this study, we additionally identified the C-terminal sequence of VL22 as a ctVSD that is recognized by VSR1 (Fig. 4). To understand how VSR recognizes other ctVSDs, we modeled the structures of VSR1-PA in complex with the C-terminal sequences of VL22, β-conglycinin, and phaseolin (Fig. 5 *B–D*). Our models show that hydrophobic residues of various sizes are well-accommodated at the cargo-binding site of VSR1-PA (Fig. 5 *B–D*). Tyr99 of the cargo-binding loop, being conserved among VSRs (Fig. 1*F*), makes hydrophobic interactions with the last two residues of ctVSD. As shown in the model of VSR1-PA–phaseolin, Phe434 at the antepenultimate position could additionally form hydrophobic interactions with Val435 and Tyr436 (Fig. 5*D*). This hydrophobic cluster can interact favorably with Tyr99. Our structural analyses suggest that hydrophobic residues are preferred at the last three positions of ctVSD.
4) Conserved Asp129 and Glu132 of the switch III region favor basic residues over acidic ones. Apart from hydrophobicity, we have observed that charge–charge interactions could also affect VSR–ctVSD interaction. In the cargo-binding site of VSR1-PA, there are two nearby negatively charged residues, Asp129 and Glu132 (Fig. 5). These two acidic residues are conserved among VSRs (Fig. 1*F*) and could form favorable charge–charge interactions with positively charged residues at the ctVSD. For example, our model suggests that Arg602 of β-conglycinin could form a salt bridge with Asp129 of VSR1-PA (Fig. 5*C*). That charge–charge interaction could contribute to VSR–ctVSD interaction is also supported by the observation that VSR1-NT failed to bind the C-terminal sequences of CRU3, VL21, and VL43 (Fig. 4*B*). As these sequences are rich in acidic residues (Fig. 4*A*), they could create charge–charge repulsion with Asp129 and Glu132 of VSR1-PA that disfavors negatively charged residues in the ctVSD sequences.
5) VSR binds CRU4 with only backbone interactions. We showed that VSR1-NT interacts with the C-terminal sequence of CRU4 in vitro. However, the CRU4 sequence is inefficient in vacuolar sorting when compared with the ctVSD of CRU1 and VL22 (Figs. 3 and 4). These observations highlight the importance of verifying in vitro physical interactions with an in vivo vacuolar sorting assay. Our model of the VSR–CRU4 complex suggests that the PA domain only makes interactions with the backbone atoms of CRU4 (*SI Appendix*, Fig. S10). At the last two residue positions, other ctVSD sequences (e.g., CRU1, VL22, β-conglycinin, phaseolin) all contain hydrophobic residues that can interact favorably with Tyr99 of VSR1-PA (Fig. 5). We speculate that CRU4 is an inefficient ctVSD because the C-terminal residues are replaced by polar serine residues, which form less favorable interactions with Tyr99 of VSR1-PA.
6) Acidic pH weakens VSR–ctVSD interactions. In the crystal structure of VSR1-PA, Arg95 forms salt bridges with Asp119 and Glu123 of VSR1-PA and with the carboxyl group of ctVSD (*SI Appendix*, Fig. S7*C*). It is conceivable that protonation of these carboxyl groups by acidic pH could weaken the VSR1-PA–ctVSD interactions. Supporting this insight, we showed that in vitro binding affinities of VSR1-PA to C-terminal peptides of CRU1 and VL22 decreased from pH 7.1 to 6.0 (*SI Appendix*, Fig. S7). That acidic pH weakens the receptor–cargo interaction could provide a plausible mechanism where VSR picks up their cargos as early as in the ER (with pH ∼7.1) and releases them in the TGN, where pH becomes slightly acidic (49–52).

Comparing the Recognition of ctVSD and ssVSD by VSR. We have previously determined the crystal structure of VSR1-PA in complex with ssVSD of barley aleurain and showed that the PA domain interacted with the residues (ADS) preceding the NPIR motif of the ssVSD (21). Supported by mutagenesis and modeling, we proposed that cargo binding induces a swivel motion of the C-terminal tail, which relocates the central domain to juxtapose with the PA domain to recognize the NPIR motif (Fig. 5*F*). The proline residue of the NPIR motif, being unfavorable at the cargo-binding site of the PA domain, also forces the motif to align with the central domain.

In this study, we showed that adding a proline residue to the C terminus of ctVSD disrupts the VSR–ctVSD interaction (*SI Appendix*, Fig. S5). We argue that the extra proline residue will break the salt bridges between the C-terminal carboxyl group and Arg95 and replace the interaction with a weaker hydrogen bond with the backbone amide. The weakened interaction is not compensated by the proline residue, which cannot form sequence-specific interactions with the central domain like the NPIR motif in ssVSD. Structure determination of VSR1-NT in complex with ssVSD may help to address how the PA and central domains cooperate to recognize ssVSDs that are located at the internal sequence positions. Nevertheless, as the PA domain recognizes the carboxyl group at the C terminus

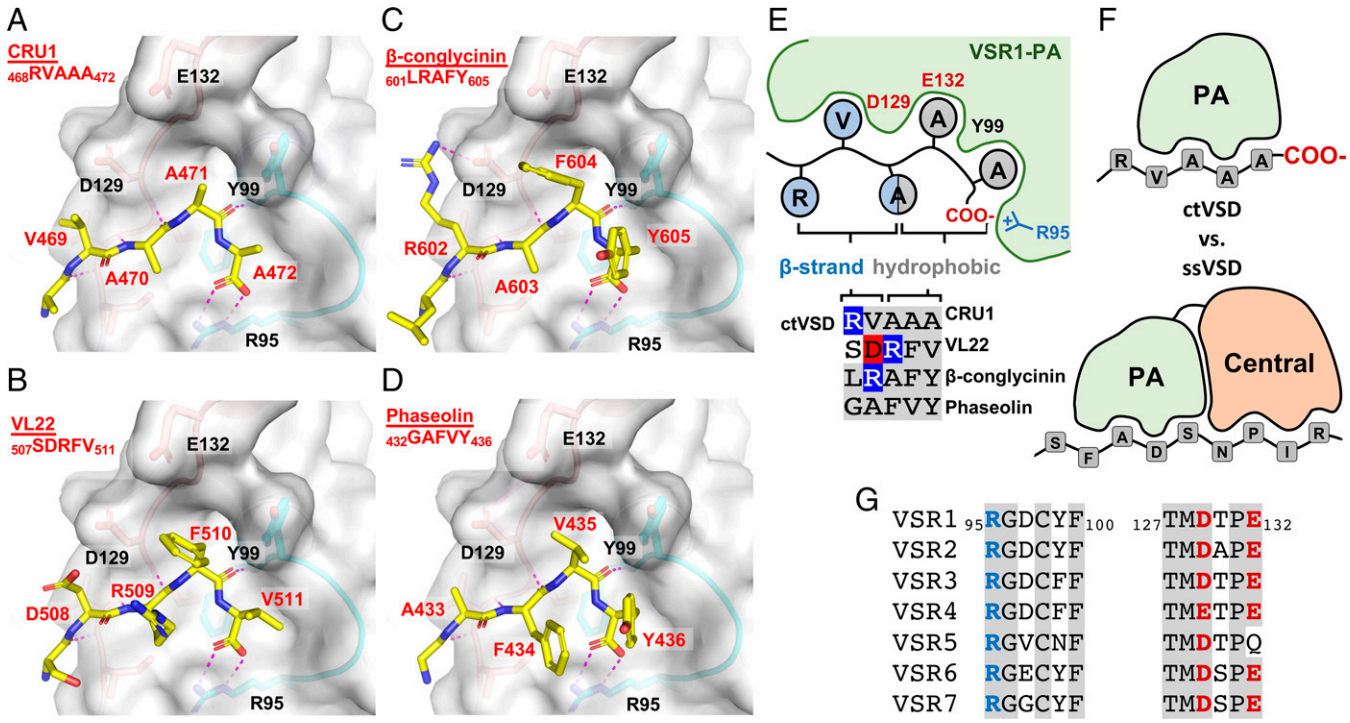

Fig. 5. Summary of sequence-specific recognition of ctVSD by VSRs. (*A–D*) Crystal structure of VSR1-PA in complex with ctVSD of CRU1 (*A*) and models of VSR-PA in complex with ctVSD of *Arabidopsis* VL22, soybean β-conglycinin (54), and phaseolin (31) (*B–D*). VSR1-PA is shown in surface representation. Residues in the cargo-binding loop, switch III region, and ctVSD are in cyan, salmon, and yellow, respectively. (*E*) Schematic diagram summarizing how VSR1-PA recognizes the sequence-specific information in ctVSD. Arg95 of VSR1-PA is responsible for recognizing the C-terminal carboxyl group of ctVSD. Tyr99 makes hydrophobic interactions with the last two residues of ctVSD. The $_{127}$TMD$_{129}$ motif of VSR1-PA forms a parallel β-sheet with $_{468}$RVA$_{470}$ of ctVSD. Conserved residues Asp129 and Glu132 located around the cargo-binding site should favor basic residues and disfavor acidic residues in the ctVSD. Hydrophobic, basic, and acidic residues are shaded in gray, blue, and red, respectively, in the alignment of ctVSD sequences. (*F*) Comparing the binding mode of ctVSD and ssVSD by VSR. (*G*) Residues involved in ctVSD recognition are highly conserved among *Arabidopsis* VSRs except VSR5, where Tyr99 and Glu132 are replaced by Asn and Gln, respectively.

of ctVSD, it is likely that the central domain is not directly involved in binding ctVSD (Fig. 5*F*).

Since both ctVSD and ssVSD bind to the PA domain of VSR, it raises the possibility that they may compete with each other for the binding of VSR. While we showed that ctVSD of CRU1 can displace the bound aleurain-ssVSD from VSR1-NT in vitro (*SI Appendix*, Fig. S8), vacuolar trafficking of ssVSD and ctVSD cargos are not affected by coexpression of sp-Aleu-GFP and spRFP-Cru1 (*SI Appendix*, Fig. S9), suggesting sorting of ssVSD and ctVSD cargos does not interfere with each other. It has been reported that some storage proteins could contain multiple VSDs (53–55). For example, β-conglycinin contains the C-terminal sequence of $_{612}$PLSSILRAFY$_{621}$ that is sufficient to target GFP to the PSV in soybean seeds (46). Further mutagenesis studies suggest that in addition to the AFY motif that served as a ctVSD, the SIL motif could also function as an ssVSD (54). Based on the structural insights into the VSR–VSD interactions reported, it is conceivable that both VSDs could be recognized by VSRs for vacuolar sorting.

In addition to VSRs, RMRs can also interact with ctVSD (26, 27). It is noteworthy that both VSRs and RMRs contain a PA domain in their N-terminal luminal region (36). Sequence alignment shows that the RGxCxF motif of the cargo-binding loop is conserved in both VSRs and RMRs (*SI Appendix*, Fig. S4). It is therefore tempting to speculate that the binding of RMRs and VSRs is similar to a way that the C-terminal carboxyl group of ctVSD is also recognized by the conserved Arg residue in the RGxCxF motif of RMRs. However, residue Tyr99 of VSR1 that forms hydrophobic interactions with the last two residues of ctVSD was replaced by either Ser or Thr in RMRs. Moreover, acidic residues Asp129 and Glu132 are conserved in only VSRs but not in RMRs. As the cargo-binding loop and these switch III residues form the cargo-binding pocket for ctVSD, it is likely the RMRs and VSRs have different sequence specificity toward ctVSD and they may bind to different cargo proteins. It is also possible that both VSRs and RMRs may sort cargo proteins via different sorting pathways. For example, it has been proposed that RMRs may be involved in the aggregation of storage proteins in the aggregation-sorting pathway, while VSRs are responsible for sorting soluble cargos via the receptor-mediated sorting pathway (28, 36).

Specific Receptor–Cargo Recognition Is Required for Vacuolar Sorting of ctVSD Cargos. We argued that the specific recognition between VSR1 and ctVSD shown in this study should also play an essential role in sorting storage proteins in *Arabidopsis* seeds. In a genetic study, Shimada and coworkers showed that in the *Arabidopsis vsr1* mutant, the storage proteins 12S globulin and 2S albumin were missorted and accumulated in the extracellular space in seeds (15). They also showed in a pull-down assay that VSR1 could bind to the C-terminal peptide of CRU1 and the binding was calcium-dependent. Zouhar and coworkers showed that the double mutants *vsr1vsr3* and *vsr1vsr4* further reduced the accumulation of the vacuolar form of 12S globulin in seeds (35), suggesting that there are functional redundancies among VSR1, VSR3, and VSR4 as sorting receptors for 12S globulin. Consistent with this view, residues in the cargo-binding loop and the switch III region that are involved in cargo recognition are highly conserved among VSR1, VSR3, and VSR4 (Fig. 5*G*). It is expected that these isoforms of VSRs should share similar sequence specificity toward cargo recognition in seeds.

Materials and Methods

Detailed materials and methods are reported in *SI Appendix, Materials and Methods*.

Structure Determination. Detail procedures for expression, purification, and structure determination of VSR1-PA are described in *SI Appendix, Materials and Methods*.

Differential Scanning Fluorimetry. Detailed procedures are described in *SI Appendix, Materials and Methods*. In brief, thermal shift assay and binding affinities were measured by DSF where 0 to 5 mM synthetic peptide was mixed with 50 μM protein samples in 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (pH 7.1) or 25 mM 2-(*N*-morpholino)ethanesulfonic acid (pH 6.0) buffer containing 2× Sypro Orange, 150 mM NaCl, 1 mM $CaCl_2$, and 1 mM Tris(2-carboxyethyl)phosphine. Dissociation constants for VSR1-PA–ctVSD interaction were determined by isothermal analysis (41) implemented in the FoldAffinity web server (42).

Transient Expression in *Arabidopsis* PSBD Protoplasts and Analyses. Transient expression of proteins with fluorescent tags and culture methods of *Arabidopsis thaliana* plant system biology dark-type culture (PSBD) protoplasts were conducted based on previous protocols (56). Plasmid construction, confocal imaging, secretion assay, and immunoblot analysis are described in *SI Appendix, Materials and Methods*.

Pull-Down Assay. Detailed procedures are described in *SI Appendix, Materials and Methods*. In brief, synthetic peptides derived from C-terminal sequences of seed storage proteins were coupled to NHS-activated Sepharose 4 Fast Flow resins (GE Healthcare) following the manufacturer's instructions. Purified protein samples of recombinant VSR1-PA or *Arabidopsis*-expressed VSR1-NT were loaded onto peptide-coupled NHS-resins. After washing, bound proteins were analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblot with VSR1-PA or T7 antibody (Abcam).

Data Availability. The crystal structures reported in this article of *A. thaliana* VSR1-PA in complex with the CRU1 C-terminal pentapeptide RVAAA at pH 6.5 and 9.0 have been deposited in the Protein Data Bank (ID codes 7F2I and 7F2D). *Arabidopsis* sequences used in this study can be accessed in The Arabidopsis Information Resource database (https://www.arabidopsis.org/) with the following accession numbers: VSR1, At3g52850; CRU1, At5g44120; CRU2, At1g03880; CRU3, At4g28520; CRU4, At1g03890; VL21, At2g18540; VL22, At2g28490; and VL43, At4g36700. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This study was supported by grants from the Research Grants Council of Hong Kong Special Administrative Region, China (CUHK14151416, C4041-18EF, AoE/M-05/12, AoE/M403/16, and C4033-19E) and The National Natural Science Foundation of China (91854201 and 31670179) and by direct grants from The Chinese University of Hong Kong.

1. Y. Wakasa, F. Takaiwa, "Seed storage proteins" in Brenner's Encyclopedia of Genetics, S. Maloy, K. Hughes, Eds. (Academic Press, 2013), pp. 346–348.
2. D. G. Robinson, J. M. Neuhaus, Receptor-mediated sorting of soluble vacuolar proteins: Myths, facts, and a new model. *J. Exp. Bot.* **67**, 4435–4449 (2016).
3. M. S. Otegui, R. Herder, J. Schulze, R. Jung, L. A. Staehelin, The proteolytic processing of seed storage proteins in *Arabidopsis* embryo cells starts in the multivesicular bodies. *Plant Cell* **18**, 2567–2581 (2006).
4. G. Hinz, S. Colanesi, S. Hillmer, J. C. Rogers, D. G. Robinson, Localization of vacuolar transport receptors and cargo proteins in the Golgi apparatus of developing *Arabidopsis* embryos. *Traffic* **8**, 1452–1464 (2007).
5. S. Castelli, A. Vitale, The phaseolin vacuolar sorting signal promotes transient, strong membrane association and aggregation of the bean storage protein in transgenic tobacco. *J. Exp. Bot.* **56**, 1379–1387 (2005).
6. T. Kirsch, N. Paris, J. M. Butler, L. Beevers, J. C. Rogers, Purification and initial characterization of a potential plant vacuolar targeting receptor. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3403–3407 (1994).
7. L. Jiang, T. E. Phillips, S. W. Rogers, J. C. Rogers, Biogenesis of the protein storage vacuole crystalloid. *J. Cell Biol.* **150**, 755–770 (2000).
8. L. Jiang, J. C. Rogers, Integral membrane protein sorting to vacuoles in plant cells: Evidence for two pathways. *J. Cell Biol.* **143**, 1183–1199 (1998).
9. L. J. Olson, J. Zhang, N. M. Dahms, J. J. Kim, Twists and turns of the cation-dependent mannose 6-phosphate receptor. Ligand-bound versus ligand-free receptor. *J. Biol. Chem.* **277**, 10156–10161 (2002).
10. J. L. Cereghino, E. G. Marcusson, S. D. Emr, The cytoplasmic tail domain of the vacuolar protein sorting receptor Vps10p and a subset of VPS gene products regulate receptor stability, function, and localization. *Mol. Biol. Cell* **6**, 1089–1102 (1995).
11. A. A. Cooper, T. H. Stevens, Vps10p cycles between the late-Golgi and prevacuolar compartments in its function as the sorting receptor for multiple yeast vacuolar hydrolases. *J. Cell Biol.* **133**, 529–541 (1996).
12. N. Paris *et al.*, Molecular cloning and further characterization of a probable plant vacuolar sorting receptor. *Plant Physiol.* **115**, 29–39 (1997).
13. S. U. Ahmed *et al.*, The plant vacuolar sorting receptor AtELP is involved in transport of NH(2)-terminal propeptide-containing vacuolar proteins in *Arabidopsis thaliana*. *J. Cell Biol.* **149**, 1335–1344 (2000).
14. D. Humair, D. Hernández Felipe, J. M. Neuhaus, N. Paris, Demonstration in yeast of the function of BP-80, a putative plant vacuolar sorting receptor. *Plant Cell* **13**, 781–792 (2001).
15. T. Shimada *et al.*, Vacuolar sorting receptor for seed storage proteins in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 16095–16100 (2003).
16. L. L. daSilva *et al.*, Receptor salvage from the prevacuolar compartment is essential for efficient vacuolar protein targeting. *Plant Cell* **17**, 132–148 (2005).
17. L. L. daSilva, O. Foresti, J. Denecke, Targeting of the plant vacuolar sorting receptor BP80 is dependent on multiple sorting signals in the cytosolic tail. *Plant Cell* **18**, 1477–1497 (2006).
18. P. K. Suen, J. B. Shen, S. S. M. Sun, L. W. Jiang, Expression and characterization of two functional vacuolar sorting receptor (VSR) proteins, BP-80 and AtVSR4 from culture media of transgenic tobacco BY-2 cells. *Plant Sci.* **179**, 68–76 (2010).
19. K. Fuji *et al.*, *Arabidopsis* vacuolar sorting mutants (green fluorescent seed) can be identified efficiently by secretion of vacuole-targeted green fluorescent protein in their seeds. *Plant Cell* **19**, 597–609 (2007).
20. L. Xiang, E. Etxeberria, W. Van den Ende, Vacuolar protein sorting mechanisms in plants. *FEBS J.* **280**, 979–993 (2013).
21. F. Luo *et al.*, How vacuolar sorting receptor proteins interact with their cargo proteins: Crystal structures of apo and cargo-bound forms of the protease-associated domain from an *Arabidopsis* vacuolar sorting receptor. *Plant Cell* **26**, 3693–3708 (2014).
22. S. Y. Bednarek, N. V. Raikhel, The barley lectin carboxyl-terminal propeptide is a vacuolar protein sorting determinant in plants. *Plant Cell* **3**, 1195–1206 (1991).
23. K. Matsuoka, K. Nakamura, Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 834–838 (1991).
24. J. M. Neuhaus, L. Sticher, F. Meins Jr., T. Boller, A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10362–10366 (1991).
25. T. Shimada, M. Kuroyanagi, M. Nishimura, I. Hara-Nishimura, A pumpkin 72-kDa membrane protein of precursor-accumulating vesicles has characteristics of a vacuolar sorting receptor. *Plant Cell Physiol.* **38**, 1414–1420 (1997).
26. M. Park, D. Lee, G. J. Lee, I. Hwang, AtRMR1 functions as a cargo receptor for protein trafficking to the protein storage vacuole. *J. Cell Biol.* **170**, 757–767 (2005).
27. Y. Shen *et al.*, The rice RMR1 associates with a distinct prevacuolar compartment for the protein storage vacuole pathway. *Mol. Plant* **4**, 854–868 (2011).
28. J. R. Ashnest, A. R. Gendall, Trafficking to the seed protein storage vacuole. *Funct. Plant Biol.* **45**, 895–910 (2018).
29. T. Kirsch, G. Saalbach, N. V. Raikhel, L. Beevers, Interaction of a potential vacuolar targeting receptor with amino- and carboxyl-terminal targeting determinants. *Plant Physiol.* **111**, 469–474 (1996).
30. K. Matsuoka, K. Nakamura, 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 (1999).
31. L. Frigerio, O. Foresti, D. H. Felipe, J. M. Neuhaus, A. Vitale, The C-terminal tetrapeptide of phaseolin is sufficient to target green fluorescent protein to the vacuole. *J. Plant Physiol.* **158**, 499–503 (2001).
32. N. Paris, J. M. Neuhaus, BP-80 as a vacuolar sorting receptor. *Plant Mol. Biol.* **50**, 903–914 (2002).
33. X. Cao, S. W. Rogers, J. Butler, L. Beevers, J. C. Rogers, Structural requirements for ligand binding by a probable plant vacuolar sorting receptor. *Plant Cell* **12**, 493–506 (2000).
34. E. Watanabe, T. Shimada, M. Kuroyanagi, M. Nishimura, I. Hara-Nishimura, Calcium-mediated association of a putative vacuolar sorting receptor PV72 with a propeptide of 2S albumin. *J. Biol. Chem.* **277**, 8708–8715 (2002).
35. J. Zouhar, A. Muñoz, E. Rojo, 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 (2010).
36. H. Wang, J. C. Rogers, L. Jiang, Plant RMR proteins: Unique vacuolar sorting receptors that couple ligand sorting with membrane internalization. *FEBS J.* **278**, 59–68 (2011).
37. J. H. Park, M. Oufattole, J. C. Rogers, Golgi-mediated vacuolar sorting in plant cells: RMR proteins are sorting receptors for the protein aggregation/membrane internalization pathway. *Plant Sci.* **172**, 728–745 (2007).
38. Y. Lee *et al.*, Functional identification of sorting receptors involved in trafficking of soluble lytic vacuolar proteins in vegetative cells of *Arabidopsis*. *Plant Physiol.* **161**, 121–133 (2013).
39. P. R. Hunter, C. P. Craddock, S. Di Benedetto, L. M. Roberts, L. Frigerio, Fluorescent reporter proteins for the tonoplast and the vacuolar lumen identify a single vacuolar compartment in *Arabidopsis* cells. *Plant Physiol.* **145**, 1371–1382 (2007).
40. M. C. Lawrence, T. Izard, M. Beuchat, R. J. Blagrove, P. M. Colman, Structure of phaseolin at 2.2 Å resolution. Implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins. *J. Mol. Biol.* **238**, 748–776 (1994).

41. N. Bai, H. Roder, A. Dickson, J. Karanicolas, Isothermal analysis of ThermoFluor data can readily provide quantitative binding affinities. *Sci. Rep.* **9**, 2650 (2019).
42. S. Niebling *et al.*, FoldAffinity: Binding affinities from nDSF experiments. *Sci. Rep.* **11**, 9572 (2021).
43. L. Frigerio, M. de Virgilio, A. Prada, F. Faoro, A. Vitale, Sorting of phaseolin to the vacuole is saturable and requires a short C-terminal peptide. *Plant Cell* **10**, 1031–1042 (1998).
44. G. Saalbach, M. Rosso, U. Schumann, The vacuolar targeting signal of the 2S albumin from Brazil nut resides at the C terminus and involves the C-terminal propeptide as an essential element. *Plant Physiol.* **112**, 975–985 (1996).
45. J. Shen *et al.*, An in vivo expression system for the identification of cargo proteins of vacuolar sorting receptors in *Arabidopsis* culture cells. *Plant J.* **75**, 1003–1017 (2013).
46. K. Nishizawa *et al.*, A C-terminal sequence of soybean beta-conglycinin alpha′ subunit acts as a vacuolar sorting determinant in seed cells. *Plant J.* **34**, 647–659 (2003).
47. A. Vitale, G. Hinz, Sorting of proteins to storage vacuoles: How many mechanisms? *Trends Plant Sci.* **10**, 316–323 (2005).
48. D. G. Robinson, P. Oliviusson, G. Hinz, Protein sorting to the storage vacuoles of plants: A critical appraisal. *Traffic* **6**, 615–625 (2005).
49. J. Shen *et al.*, Organelle pH in the *Arabidopsis* endomembrane system. *Mol. Plant* **6**, 1419–1437 (2013).
50. D. G. Robinson, P. Pimpl, Receptor-mediated transport of vacuolar proteins: A critical analysis and a new model. *Protoplasma* **251**, 247–264 (2014).
51. F. Künzl, S. Früholz, F. Fäßler, B. Li, P. Pimpl, Receptor-mediated sorting of soluble vacuolar proteins ends at the *trans*-Golgi network/early endosome. *Nat. Plants* **2**, 16017 (2016).
52. S. Früholz, F. Fäßler, Ü. Kolukisaoglu, P. Pimpl, Nanobody-triggered lockdown of VSRs reveals ligand reloading in the Golgi. *Nat. Commun.* **9**, 643 (2018).
53. N. Maruyama *et al.*, Multiple vacuolar sorting determinants exist in soybean 11S globulin. *Plant Cell* **18**, 1253–1273 (2006).
54. K. Nishizawa, N. Maruyama, S. Utsumi, The C-terminal region of alpha′ subunit of soybean beta-conglycinin contains two types of vacuolar sorting determinants. *Plant Mol. Biol.* **62**, 111–125 (2006).
55. D. D. Hegedus *et al.*, Multiple internal sorting determinants can contribute to the trafficking of cruciferin to protein storage vacuoles. *Plant Mol. Biol.* **88**, 3–20 (2015).
56. Y. Miao, L. Jiang, Transient expression of fluorescent fusion proteins in protoplasts of suspension cultured cells. *Nat. Protoc.* **2**, 2348–2353 (2007).
57. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).