



OPEN

Functional implication of the homotrimeric multidomain vacuolar sorting receptor 1 (VSR1) from *Arabidopsis thaliana*

HaJeung Park¹, BuHyun Youn², Daniel J. Park³, Sathyanarayanan V. Puthanveetil⁴ & ChulHee Kang⁵✉

The vacuolar sorting receptors (VSRs) are specific to plants and are responsible for sorting and transporting particular proteins from the *trans*-Golgi network to the vacuole. This process is critically important for various cellular functions, including storing nutrients during seed development. Despite many years of intense studies on VSRs, a complete relation between function and structure has not yet been revealed. Here, we present the crystal structure of the entire luminal region of glycosylated VSR1 from *Arabidopsis thaliana* (AtVSR1) for the first time. The structure provides insights into the tertiary and quaternary structures of VSR1, which are composed of an N-terminal protease-associated (PA) domain, a unique central region, and one epidermal growth factor (EGF)-like domain followed by two disordered EGF-like domains. The structure of VSR1 exhibits unique characteristics, the significance of which is yet to be fully understood.

The proper functioning of eukaryotic cells requires the sorting and targeting of proteins synthesized in the endoplasmic reticulum (ER). From the ER, correctly folded/assembled proteins are transferred to the Golgi, which initiates the first step in protein sorting and trafficking. In plant cells, vacuolar sorting receptors (VSRs) are responsible for the sorting of proteins from the *trans*-Golgi network (TGN) to prevacuolar compartments (PVCs) and finally to their respective vacuoles¹. VSRs are type-I transmembrane proteins with ~ 600 amino acids (80 kDa without the N-terminal signaling peptide)². VSRs recognize specific signal sequences of the cargo proteins called Vacuolar Sorting Determinants (VSDs) through its N-terminal protease-associated (PA) domain^{3–5}. This recognition by VSRs is known to be pH-dependent; the cargo protein binds to the VSR via the VSD in the TGN in a neutral pH, and the cargo protein is released upon exposure to an acidic pH⁶. An alternative model presented by Kunzl et al. suggests that the binding and release of cargo occur at the ER/Golgi and TGN/early endosome, respectively, through the influence of pH and Ca²⁺ changes⁷. In addition, PV72, a homolog of AtVSR1 in potatoes has a calcium-dependent cargo binding within the pH range of 5.5–7⁸.

The VSDs of various vacuolar proteins are mainly classified into two groups, sequence-specific vacuolar sorting determinants (ssVSDs)^{9,10} and C-terminal vacuolar sorting determinants (ctVSDs)¹¹. The proaleurain peptide (SSSFADSNPIRPVTDRAASTYC) present in a plant cysteine protease is the prime example of a ssVSD with the presence of a central 'NPIR' motif^{9,10}. Any alteration from this specific sequence motif can negatively affect AtVSR1-binding. For instance, a mutated peptide with glycine in place of isoleucine in the NPIR motif renders the peptide unable to compete with the proaleurain peptide for binding to VSR1^{9,10}. On the other hand, the characteristics of ctVSDs appear to be contextual without any distinctive known sequence features, though it has been proposed that the VSR1-binding motif is approximately five residues long, with the last three being hydrophobic, and it is located at the C-terminus of a cargo protein^{5,12}. A recent study by Park et al. proposed that soluble proteins carrying a ctVSD are transported by RMRs, not by AtVSR1¹³.

The structural details of PA domain of AtVSR1 in complexes with both ssVSD and ctVSD were reported^{4,5,12}. The structures show that the peptides interact with the PA domain of AtVSR1 in sequence non-specific manners, as the peptides bind the PA domain through their backbone only without any side chain involvement.

¹X-Ray Core, UF Scripps Biomedical Research, University of Florida, Jupiter, FL 33458, USA. ²Department of Biological Sciences, Pusan National University, Busan 46241, Republic of Korea. ³Burnett School of Biomedical Sciences, University of Central Florida, Orlando, FL 32827, USA. ⁴Department of Neuroscience, UF Scripps Biomedical Research, University of Florida, Jupiter, FL 33458, USA. ⁵Department of Chemistry, Washington State University, Pullman, WA 99164, USA. ✉email: chkang@wsu.edu

Furthermore, the PA-ssVSD complex showed that the PA interacted with the preceding residues of the NPIR motif, while the motif itself was invisible, probably because of structural disorder⁴. Despite these structural studies, the involvement of the PA domain in recognizing VSD motifs remains unclear.

Herein, we report a crystal structure of full-length VSR1 luminal domain from *Arabidopsis thaliana* (AtVSR1). The structure displays a distinct organization of constituent domains with unique interactions among them. Although the structure does not contain a bound peptide, it still provides valuable insights into the ligand-binding mechanism of VSRS. Additionally, the crystal contacts and domain structures offer further information about the other characteristics and potential functions of this unique class of proteins.

Results

The overall structure

The luminal part of AtVSR1 from residue 1 to 560 was expressed from *Drosophila* S2 insect cells. The full sequence is disclosed in the Supplemental Information. The purified protein was crystallized in a space group P2₁3, which diffracted up to 2.6 Å¹⁴. The crystal lattice packing indicated that one molecule in the asymmetric unit was closely associated with neighboring molecules through crystallographic three-fold symmetry. The resulting crystal structure was composed of an N-terminal protease-associated (PA) domain, a unique central region, and three C-terminal epidermal growth factor (EGF)-like repeats³ (Fig. 1A). The resolved structure had disordered regions at the C-terminus, where density could not be uniquely traced for proper model building after the first EGF-like domain. At the beginning of the refinement, a clear electron density attached to the sidechain of Asn289 in the central domain was observed and assigned as two N-acetylglucosamine (GlcNAc).

The structural model starts at Phe21 and ends at Ala465 and the overall structure is separated into three distinct domains: the PA domain (Phe21-Trp175), the thioredoxin (TRX) domain (Val186-Phe396), and the EGF-like domain (Glu409-Ala462) (Fig. 1B). The central domain is named as TRX domain, as a DALI search picked γ -interferon-inducible lysosomal thiol reductase (PDB ID 6NWX) as the closest 3D structure with a high Z-score of 14.6, followed by bacterial disulfide isomerase A (PDB ID 3BCI) with a Z-score of 14.0¹⁵. The rest of the DALI results are all TRX-fold proteins, most of which are involved in protein disulfide bond reduction (Table 1).

The PA and the TRX domains are connected by a 7-residue linker with no significant interaction between them. In contrast, the TRX and the EGF-like domains are connected through a 12-residue linker and display tight interdomain interactions between them through an extensive hydrogen bond (H-bond) network (Suppl. Table S1). At the core of the interdomain interface, the side chain of Asp303 in TRX forms H-bonds with the side chains of Arg440 and Tyr457 in EGF-like domains (Fig. 2A,B). At the periphery, an additional salt bridge between Lys 221 and Glu414 is observed (Fig. 2C), and the main chain carbonyl of Arg440 also forms an H-bond with the amine side chain of Lys213 (Fig. 2B). Notably, no significant hydrophobic interaction was found in this interdomain interaction. The Cys405 in the linker establishes a disulfide bond with Cys393 of the TRX domain potentially stabilizing the linker.

The PA domain is composed of approximately 120 residues in length and organized as a central β -barrel with nine strands in two β -sheets, and two α -helices that are located at the periphery (Suppl. Fig. S1). Similarly shaped PA domains are known to exist in several protease classes such as subtilases, aminopeptidases, bacterial

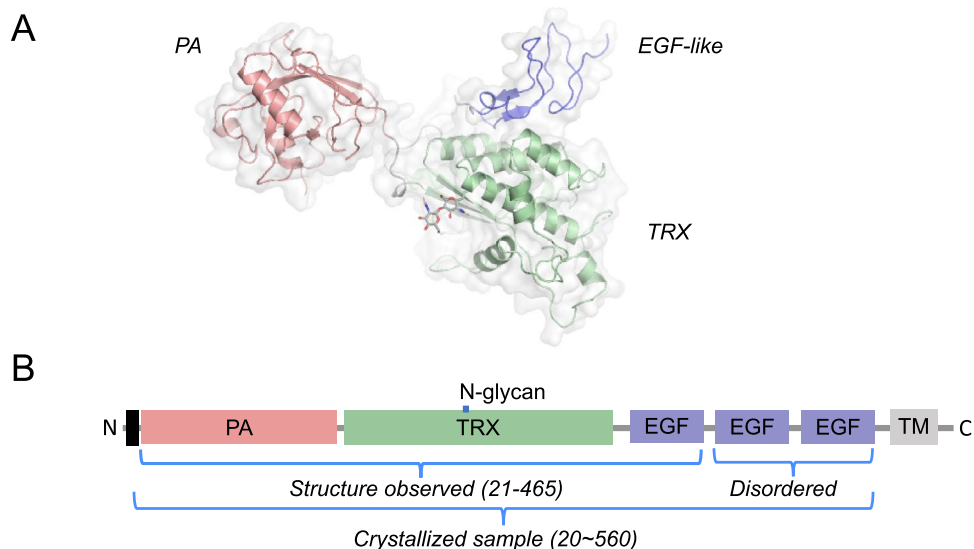


Figure 1. The overall structure and the schematic of AtVSR1. **(A)** The overall structure of AtVSR1 is shown as a cartoon model. The color of each domain matches that of the schematic. Each domain is also labeled. **(B)** The schematic of AtVSR1 shows the domain organizations, the information of the crystallization construct, and the structure resolved portion. N-glycan indicates the location where glycosylation is observed. The black rectangle represents the signaling peptide.

No.	PDB ID-Chain	Z	RMSD (Å)	Alignment length	%ID	Description	Organism
1	6nwx-A	14.6	3.0	162	10	γ -interferon-inducible lysosomal thiol reductase	<i>Mus musculus</i>
2	3bci-A	14.0	2.7	153	14	DsbA	<i>Staphylococcus aureus</i>
3	3bck-A	13.9	2.7	152	14	DsbA T153V	<i>Staphylococcus aureus</i>
4	3bd2-A	13.8	2.7	152	14	DsbA E96Q	<i>Staphylococcus aureus</i>
5	2in3-A	12.6	2.9	155	10	PDI	<i>Nitrosomonas europaea</i>
6	6ghb-D	12.3	2.9	159	9	YjbH	<i>Geobacillus kaustophilus</i>
7	3eu3-A	12.2	3.3	156	14	BdbD	<i>Bacillus subtilis</i>
8	3gha-A	12.1	3.2	156	14	BdbD	<i>Bacillus subtilis</i>
9	3eu4-A	12.1	3.3	156	14	BdbD	<i>Bacillus subtilis</i>
10	3gh9-A	12.1	3.3	156	14	BdbD	<i>Bacillus subtilis</i>
11	5vyo-B	11.6	3.5	153	18	DsbA	<i>Burkholderia pseudomallei</i>
12	4jr4-A	11.6	3.0	157	13	DsbA	<i>Mycobacterium tuberculosis</i>
13	4jr6-A	11.6	3.0	157	12	DsbA	<i>Mycobacterium tuberculosis</i>
14	3kzq-B	11.6	3.5	159	11	Putative protein	<i>Vibrio parahaemolyticus</i>
15	5hfi-A	11.5	2.9	154	9	Disulfide reductase DsbM	<i>Pseudomonas aeruginosa</i>
16	5vyo-C	11.5	3.5	153	18	DsbA	<i>Burkholderia pseudomallei</i>
17	4k6x-B	11.5	3.0	156	13	DsbA-like oxidases	<i>Mycobacterium tuberculosis</i>

Table 1. Structural similarity search by the Dali server.

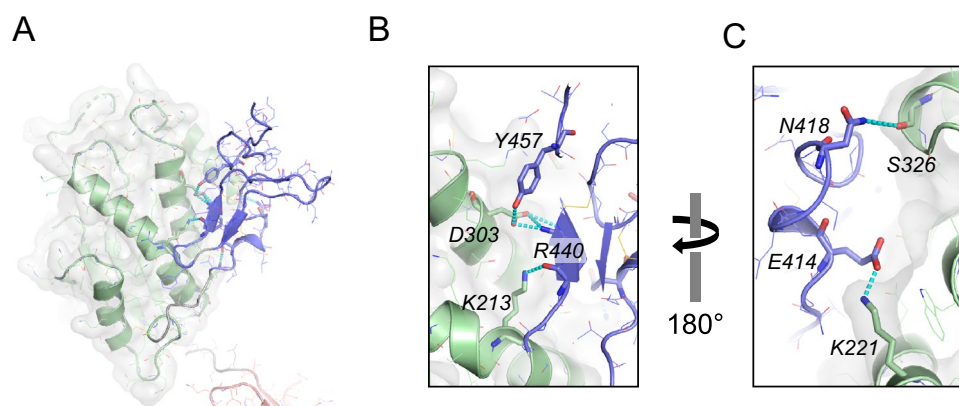


Figure 2. Domain interaction between the TRX and the EGF-like domain. (A) The overall structure showing the interaction between the two domains. (B,C) Detailed hydrogen (H)-bond interactions are shown. Panel (A) and (B) are presented in the same view. Panel (C) provides a back side view of panel (A) and (B). The TRX domain and EGF-like domains are colored green and blue, respectively, and the hydrogen bonds are colored cyan. The key residues are depicted as sticks.

endopeptidases, as well as two families of sorting receptors, RMRs and VSRs^{16,17}. However, the exact function of the PA domain has been examined in only a handful of proteins and speculated that the PA domain may serve as a protease-interacting or -regulating domain based on multiple studies¹⁷. Previous structural studies revealed that the PA domain of AtVSR1 harbors the binding sites for ctVSD and a portion of ssVSD^{4,5,12}. The shallow pocket is established between β_4 - α_3 loop and α_4 - α_5 loop of PA domain where α_4 unfurls and its residues are pushed away by the bound peptide^{4,5,12}. The residues between β_5 and α_5 (Ser120-Asp137) appear highly flexible as they are disordered in our crystal structure (Suppl Figs. S1 and S2). A portion of the corresponding region is also reported to be disordered in the peptide-bound states⁴; however, the same portion is fully structured in the apo-form (PDB ID 4TJV) due to stabilization through interaction with the C-terminus and crystallographic contacts. Apart from this C-terminal linker, the PA domain of our AtVSR1 structure overlaps well with the previous apo-form PA domain structure (PDB ID 4TJV), with a root-mean-square-deviation (RMSD) of 1.22 Å over 148 Ca positions of the residues from 26 to 173.

The TRX domain of our AtVSR1 structure reveals four internal disulfide bonds plus an above-mentioned additional disulfide bond between Cys393 and Cys405, which connects it to the EGF-like domain. Although there is no direct interaction between the intramolecular PA and TRX domains, the two domains display extensive interactions with those domains of two neighboring molecules related by a crystallographic threefold symmetry, forming a cyclical domain-swapping (Fig. 3A). This homotrimer occurs without any conformational change in

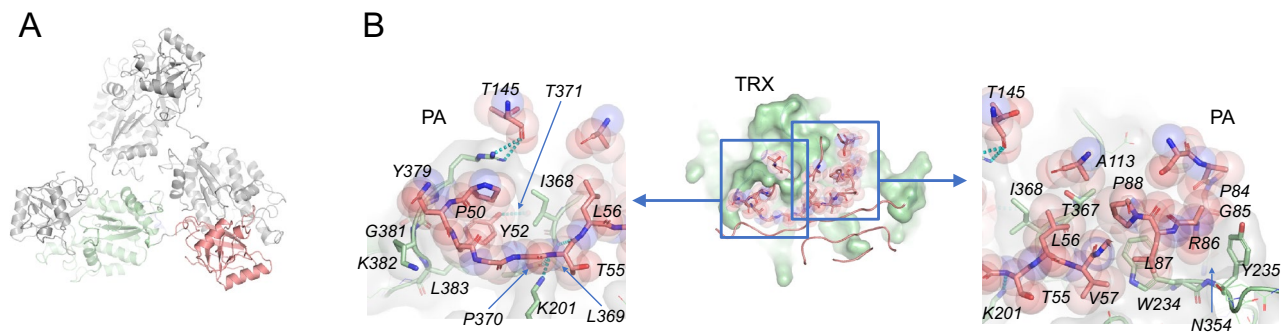


Figure 3. The crystallographic structure exhibits threefold symmetry through the swapping of the PA and TRX domains. **(A)** The asymmetric unit molecule is colored salmon and green, while the other two symmetry-related molecules are colored grey. **(B)** The detailed interaction between the two domains involves extensive hydrophobic interactions and tight van-der-Waals complementarity. The overall interface measures approximately $1,040 \text{ \AA}^2$. The PA and TRX domain residues responsible for the interactions are labeled and presented as sticks with transparent spheres (PA) and surface (TRX) (left and right panels). The PA and TRX domains are colored salmon and green, respectively.

the PA domain, evidenced by the average RMSD of 0.64 \AA between our structure and the structures of the PA domain alone (PDB ID 4TJV, 4TJX, and 8HYG).

AlphaFold2 model

The full length model of AtVSR1, generated by AlphaFold2 (AF2), has been deposited in the AlphaFold protein structure database with an accession number AF-93026-F1¹⁸. Most of the regions in the model have per-residue-confidence-scores (predicted Local Distance Difference Test, pLDDT) of 0.9 or higher, indicating that the model is highly reliable. Each domain of the model superimposes well with that of the crystal structure (Suppl Fig. S3A). However, the interdomain interactions of the AF2 model are completely different from those observed in our crystal structure (Suppl Fig. S3B,C). The inaccuracy of quaternary structure prediction is a well-known limitation of AF2^{19,20}. Despite the caveat, the AF2 model structure still offers valuable insight into how the regions not observed in the crystal structure may look like.

The oligomerization of AtVSR1

PDBEPIA analysis of the crystallographic *intermolecular* interaction between the PA and the TRX domains reveals that the interface score is 1.0, indicating that AtVSR1 is highly likely to form a trimer in solution²¹. The total buried surface area between the PA and the TRX domains is approximately $1,040 \text{ \AA}^2$, and the interface is formed by 24.6% and 14% of the residues from the PA domain and the TRX domain, respectively. In comparison, the intramolecular interaction between the TRX and EGF-like domains has a PISA interface score of 0, indicating the observed interaction is not significant. The intermolecular interaction between the PA and the TRX domains displays notable hydrophobic interactions mediated by both ends of the loops connected to $\beta 3$ of the PA domain (Fig. 3B). At one end, the residue Tyr52 of PA domain is nestled in a hydrophobic pocket in TRX domain established by Ile368, Pro370, Thr371, Tyr379, Gly381, Lys382 and Leu383. In addition, the side chain of Arg380 is packed against the nearby hydrophobic residue, Pro50, while also forming an H-bond interaction with the side chain of Thr145. The main chain carbonyl and amine of Thr55 exchanges H-bonds with the main chain of Leu369. The hydroxyl group of Tyr52 also forms a H-bond with the main chain carbonyl oxygen of Thr371 (Fig. 3B, left). On the other end, a loop formed by Pro84, Gly85, and Arg86 is nestled within a valley fabricated by Trp234, Tyr235, and Gln354. Aggregation of hydrophobic residues, Leu56, Val57, Leu87, Pro88, and Ala113 connects the loops (Fig. 3B, right). Considering the resolved crystal structures of the TRX domain, the previously hypothesized *intramolecular* PA and TRX domain interaction through the same interface appears unlikely. The linker between the PA and the TRX domains is composed of 8 residues with approximately 26 \AA in length. Even when we expanded the linker boundary in our modeling attempt, the proposed intramolecular interaction between the PA and the TRX domains was not possible without disrupting the adjacent residues and dihedral angle violations in linker residues (Suppl Fig. S4). This further supports the possibility of an intermolecular interaction that results in a domain-swapped trimer *in vivo*.

The EGF-like domain

It is noticeable that there is no significant intermolecular interaction by the EGF-like domains in the crystal lattice (Suppl Fig. S5). The two additional EGF-like domains that are tandemly located at the C-terminus are not visible in the crystal structure, indicating their high flexibility. It is likely that the string of EGF-like domains is oriented towards the ER membrane and provides flexibility for the PA and TRX domains, enabling them to promote intermolecular interactions (Suppl. Fig. S3A). The interaction between the TRX and EGF-like domains is held by multiple H-bond interactions. There are 4 residues each from the TRX domains and the EGF-like domains involved in this interaction. Together, they form 7 H-bonds, out of which 4 are salt bridges (Fig. 2, Suppl. Table S1).

Discussion

Trimeric nature of AtVSR1

Cao et al. conducted elaborate experiments digesting the luminal domain of AtVSR1 to define three protease-resistant domains and their involvement in ssVSD binding. Although the proposed multi-domain nature of AtVSR1 was accurate, interpreting the results based on the estimated molecular weights by SDS-PAGE has some drawbacks³. In the report, they suggested the luminal domain of AtVSR1 could be monomeric in solution based on its elution profile being similar to bovine serum albumin (BSA) on a gel filtration chromatography³. To have similar hydrodynamics to BSA, AtVSR1 must be tightly packed, suggesting that it could exist as a monomer with an intramolecular interaction between the PA and the TRX domains. In addition, all three EGF-like domains should be tightly packed with the other domains in the same molecule. However, our crystal structure shows that the AtVSR1 has an extended conformation with significant dynamic flexibility of the EGF-like domains. Significantly, the same chromatographic profile of Cao et al. also displayed a shoulder peak with a high molecular weight aggregate, which could be attributed to AtVSR1 multimers.

Evidence of AtVSR1 forming a trimer was reported by Kim et al. where hemagglutinin-tagged AtVSR1 (AtVSR1:HA) is expressed in transgenic plants, followed by fractionation of the protein extract using Superdex 200 HR 10/30, and identified the existence of 240 KDa and ~80–100 KDa species that responded to anti-HA antibody (Ab)²². Subsequent experiments confirmed the 240 KDa species to be homopolymer of AtVSR1. They further dissected the regions responsible for homotrimer formation and concluded that both the transmembrane and C-terminal cytosolic domains were necessary, while the luminal domain was not. This conclusion was based on a series of co-immunoprecipitation assays using deletion and substitution mutants of AtVSR1. It is possible that the detergents contained in the assay buffer were not suitable for the oligomerization of the luminal domain. Nevertheless, AtVSR1 with the C-terminal cytosolic domain mutants that cannot form oligomers suffer in vacuolar trafficking efficiency due to localization to the Golgi apparatus instead of the prevacuolar compartment.

Considering the structural evidence of homotrimer formation through the luminal PA and TRX domains present in full-length luminal domain of AtVSR1, the formation of the oligomer through the transmembrane and C-terminal cytosolic domains might be enhanced by the homotrimer of the luminal domain.

Domain-swapping among homopolymers as observed in our AtVSR1 is not uncommon, in which proteins establish a dimer, a cyclic multimer, or an open-ended aggregate^{23,24}. The swapping can be performed via exchanging a secondary structural element or an intact domain of the participating monomeric subunit. Diphtheria toxin is a classic example that forms a dimer through the latter method²⁵, and the conformation in our AtVSR1 structure also swaps domains similarly. While diphtheria toxin can form a monomer through *intramolecular* domain interaction of the receptor domain, the only viable option for AtVSR1 is to form a homotrimer (or possibly a homomultimer) through *intermolecular* interactions, as opposed to the previously proposed monomer through intramolecular domain interaction. The presence of restrictive linker residues prevents the existence of a functional monomer with the compacted PA-TRX domains. Additionally, due to its cyclic domain swapping, AtVSR1 may also form additional oligomers.

Further analysis of VSR1 in solution using size exclusion chromatography-multi-angle light scattering (SEC-MALS) or sedimentation equilibrium (SE) analysis is necessary to clarify the oligomeric state in solution and the functional implications of oligomerization in the luminal domain.

VSD-recognition mechanism of AtVSR1

The previous structural study by Luo et al. showed that a part of the loop between $\beta 5$ and $\beta 6$ in the PA domain is displaced by an ssVSD peptide containing an NPIR motif⁴. Unexpectedly, this complex structure showed that the interaction between PA domain and a ssVSD peptide is not mediated by the NPIR motif, but rather by three preceding residues of the NPIR motif. Although sensitivity to mutation of the penultimate residue before the NPIR motif was shown by pull-down and other biological assays, a detailed structural analysis showed a sign of promiscuity in the ssVSD peptide binding, as little side chain involvement was observed.

The promiscuity of interaction is evidenced by another structural study by Tsao et al., where the authors found that a ctVSD peptide from CRU1 binds to the AtVSR1 PA domain in the same way as the ssVSD NPIR-peptide barley aleurain. ctVSD-peptides are composed of four C-terminal residues without a known consensus^{5,12}. The structural study revealed a favored tendency; namely, a basic residue is preferred in position 1, and hydrophobic residues (except for proline) are favored in the three remaining positions. Intriguingly, the C-terminal carboxyl group of ctVSD is recognized by Arg95 of PA domain, which is conserved throughout the VSR isoforms. Therefore, the binding site of VSR appears to be attuned to a C-terminal region rather than an internal sequence motif.

Based on the crystal structure of PA domain, Luo et al. hypothesized that the *internal* NPIR peptide recognition by AtVSR1 is established by two separate domains: the three residues preceding the NPIR-binding motif in the PA domain and the NPIR-binding motif in the TRX domain⁴. As a mechanism to bring two motifs together of two separated domains from a same AtVSR1 molecule, the authors pointed out the 180° swing motion triggered by the three-residue peptide binding to the PA domain⁴.

Being estimated from our crystal structure of the luminal domain, the shortest distance between the ctVSD-binding site of the PA domain and the either intra- or inter-molecular TRX domain is approximately 30 Å, which is too far for an NPIR peptide to bind (Suppl. Fig. S6). If the domain-swapped trimer presented here exists in vivo and the NPIR-binding site is located in the TRX domain, it is still unlikely that the ctVSD-binding site at the PA domain is involved in ssVSD recognition.

The plausible function of the TRX domain

The TRX domain observed in our AtVSR1 structure shows a typical fold observed among protein disulfide isomerase, PDI-fold²⁶, featuring a four-helix bundle attached between the classic TRX fold of $\beta_1\alpha_1\beta_2$ and $\beta_3\beta_4\alpha_2$

topologies (Suppl. Figs. S1 and S7). As noticed in many PDIs, β_4 -strand in the TRX domain of AtVSR1 is not prominent. The helical bundle observed in the TRX domain of AtVSR1 has two disulfide bonds; one between helices α_1 and α_4 , and another between helices α_2 and α_3 . There is a 40-residue-long insertion unique to the TRX domain between β_1, α_2 and the helical bundle, which features an α -helix and a 3_{10} helix with two disulfide bonds (Suppl. Fig. S7). The archetypal TRX-fold containing PDI enzymes have a dithiol CxxC active site at the beginning of α_1 to catalyze the reduction of a disulfide bond. Superposition of the TRX domain with the PDBs retrieved by the DALI search shows that the dithiol CxxC motif is structurally conserved among those PDIs, even though AtVSR1 has an additional residue between the two conserved cysteines as $^{198}\text{Cxxx}^{\text{C}202}$ (Suppl. Fig. S7).

Some of the proteins that are sorted by VSR1 contain disulfide bonds. For example, cruciferin 1 and 3, which are 12S globulins in Arabidopsis, have two and one disulfide bond, respectively. Additionally, aleurain also has two disulfide bonds. Thus it is tempting to speculate that VSR1 engages in the quality control of the cargo proteins with disulfide before or during the sorting process. Furthermore, the formation of vacuoles in seed storage is a rapid process, which is typically completed within 24–48 h²⁷. Rapid synthesis of the storage proteins may compromise protein folding, and the TRX domain of VSR1 may function as a secondary mechanism to ensure folding of proteins with proper disulfide bond configurations.

An intriguing aspect of TRX domains in general is their ability to recognize protein substrates. Whether the TRX domain of VSR1 is an active enzyme or nature is simply re-purposing the fold²⁸ for target (ssVSD) binding needs further investigation.

Ca²⁺ and pH dependence of the EGF-like domain

Ca²⁺-coordination property of the EGF-like domains is achieved by 5 residues, 3 of which through conserved side-chain carboxyl/carboxamide and 2 through the main-chain carbonyls^{29–31}. Comparison of the EGF-like domain of AtVSR1 to known Ca²⁺-binding EGF-like domains show those residue positions are conserved, suggesting the EGF-like domain of VSR1 is likely to bind a Ca²⁺ (Suppl. Fig. S8). Deleting the EGF-like domain in PV72, which is a homolog of AtVSR1 from pumpkin seed, displayed a tenfold decreased affinity for the NPIR peptide compared to the full length. Furthermore, the binding of the NPIR peptide of PV72 is dependent on Ca²⁺, supporting the significance of Ca²⁺-binding motif³².

Multiple studies have consistently shown pH-dependent cargo binding and unloading of VSRs. In a recent study, it was demonstrated that the NHX5 and NHX6 antiporters play a critical role in maintaining proper pH homeostasis in vacuoles, and VSR-cargo binding and trafficking rely on pH homeostasis maintained by the two vesicular antiporters⁶. The findings further highlight the importance of pH-dependent receptor-cargo interactions in protein trafficking and provide insights into the role of NHX antiporters in regulating VSR function.

Noticeably, the interaction between the TRX and EGF-like domains is held together by 6 H-bonds, 3 of which are salt bridges (Suppl. Table S1). The pK_a values of Glu and Asp are 4.4 and 4.0, respectively. Therefore, at low pH, the strength of the salt bridges is likely to be reduced significantly. Likewise, binding of Ca²⁺ to the EGF-like domain is expected to cause a conformational change in the domain, which may influence the interaction with the TRX domain. Further study is needed to fully understand the involvement of the TRX and EGF-like domains, as well as the role of the salt bridges involved, in NPIR peptide binding.

Conclusion

Decades ago, when we embarked on determining the structure of AtVSR1 generated from the insect cell expression system, we anticipated that it would provide answers to questions regarding the protein's function. The structure of the full-length luminal domain of AtVSR1 with glycosylation indeed provides fascinating insights into its function. It also raises more intriguing questions that require further investigation. Based on our current structural study, we argue whether the proposed model by Luo et al. accurately represents the physiological condition⁴, that is whether the ctVSD and ssVSD sites are in the PA and TRX domain in tandem at AtVSR1. We instead speculate that the ssVSD binding might be present near the TRX and EGF-like domain interface. This is due to the fact that ssVSD binding is affected by pH and Ca²⁺ ion, and the interaction between the TRX and EGF-like domain is dependent on multiple H-bonds. The Ca²⁺ binding site of the TRX domain may serve as a pH/Ca²⁺ sensor. Further research is needed to elucidate the specific region responsible for ssVSD binding in AtVSR1. Furthermore, whether the TRX domain of AtVSR1 functions as a disulfide bond isomerase needs to be investigated. The PFAM search did not find any other examples of a receptor protein having a chaperone domain. The VSRs may be a truly unique receptor system if the TRX domain functions as disulfide bond isomerase. Given the rapid nature of vacuolar sorting in cells, it is possible that quality control is not given high priority. In this regard, VSRs may serve a dual role of binding cargo and also playing a catch-up function of cargo refolding during the sorting process. Future research should investigate this unique plant system to enhance our understanding of the perspective it offers to all life forms.

Methods

Protein expression and purification

Information regarding cloning and protein production can be found in the literature^{3,3,14}. In brief, the plasmid containing the luminal domain of AtVSR1 with a C-terminal 6XHis-tag was transfected into Drosophila S2 cells with the Drosophila Expression System kit. Stably transformed cell lines were selected with hygromycin and transferred to serum-free medium over a month period. Expression of the recombinant proteins was induced with 0.5 mM copper sulfate for 72 h. AtVSR1 was purified from S2 cell medium using His affinity column followed by a proaleurain peptide-affinity column.

Crystallography

The details of the crystallization method and data collection information were disclosed by Rogers et al.¹⁴. However, the 2.6 Å dataset featured in the paper was lost, and a dataset from a crystal diffracted to 3.5 Å Bragg spacings was used in this study and was successfully phased by molecular replacement using the PA domain of AtVSR1 (PDB ID 4TJV) as a search model³³. The crystal contained one molecule in the asymmetric unit. The top solution of the MR search had a translation function z-score (TFZ) of 32.6, and the initial phase map showed boundaries for two additional domains as having positive density, all leading to the conclusion that the MR solution was correct. The additional model building was done with Coot, and the refinements were done in Phenix-refine^{34,35}. Even at 3.5 Å resolution, most side chain electron densities were legible, and therefore the residues were placed in position. The final structure had a total of 427 residues with N-glycosylation at Asn289. R_{work} and R_{free} were 18.03%, and 21.60%, respectively. Regularization of the coordinate for structural analysis was done in the Maestro protein preparation workflow (Schrödinger, LLC). Structural figures were created with PyMol (Schrödinger, LLC). The data processing and refinement statistics are presented in Suppl. Table S2. The coordinate has been deposited (PDB ID 9B1R).

Data availability

Data is provided within the manuscript or supplementary information files.

Received: 13 January 2024; Accepted: 23 March 2024

Published online: 26 April 2024

References

- Kang, H. & Hwang, I. Vacuolar sorting receptor-mediated trafficking of soluble vacuolar proteins in plant cells. *Plants* **3**, 392–408 (2014).
- Paris, N. *et al.* Molecular cloning and further characterization of a probable plant vacuolar sorting receptor. *Plant Physiol.* **115**, 29–39 (1997).
- Cao, X., Rogers, S. W., Butler, J., Beevers, L. & Rogers, J. C. Structural requirements for ligand binding by a probable plant vacuolar sorting receptor. *Plant Cell* **12**, 493 (2000).
- Luo, F. *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).
- Tsao, H.-E. *et al.* Structural insights into how vacuolar sorting receptors recognize the sorting determinants of seed storage proteins. *Proc. Natl. Acad. Sci.* **119**, e2111281119 (2022).
- Reguera, M. *et al.* pH regulation by NHX-type antiporters is required for receptor-mediated protein trafficking to the vacuole in arabidopsis. *Plant Cell* **27**, 1200–1217 (2015).
- Künzl, F., Frühholz, S., Fäßler, F., Li, B. & Pimpl, P. Receptor-mediated sorting of soluble vacuolar proteins ends at the trans-Golgi network/early endosome. *Nat. Plants* **2**, 16017 (2016).
- Watanabe, E., Shimada, T., Kuroyanagi, M., Nishimura, M. & Hara-Nishimura, I. Calcium-mediated association of a putative vacuolar sorting receptor PV72 with a propeptide of 2S Albumin*. *J. Biol. Chem.* **277**, 8708–8715 (2002).
- Kirsch, T., Saalbach, G., Raikhel, N. V. & Beevers, L. Interaction of a potential vacuolar targeting receptor with amino- and carboxyl-terminal targeting determinants. *Plant. Physiol.* **111**, 469–474 (1996).
- Kirsch, T., Paris, N., Butler, J. M., Beevers, L. & Rogers, J. C. Purification and initial characterization of a potential plant vacuolar targeting receptor. *Proc. Natl. Acad. Sci.* **91**, 3403–3407 (1994).
- Shimada, T. *et al.* Vacuolar sorting receptor for seed storage proteins in Arabidopsis thaliana. *Proc. Natl. Acad. Sci.* **100**, 16095–16100 (2003).
- Shimada, T., Kuroyanagi, M., Nishimura, M. & Hara-Nishimura, I. A pumpkin 72-kDa membrane protein of precursor-accumulating vesicles has characteristics of a vacuolar sorting receptor. *Plant Cell Physiol.* **38**, 1414–1420 (1997).
- Park, J. H., Oufattole, M. & Rogers, J. C. 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).
- Rogers, S. W., Youn, B., Rogers, J. C. & Kang, C. Purification, crystallization and preliminary crystallographic studies of the ligand-binding domain of a plant vacuolar sorting receptor. *Acta Crystallogr. Sect. D* **60**, 2028–2030 (2004).
- Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **32**, e4519 (2023).
- Mahon, P. & Bateman, A. The PA domain: A protease-associated domain. *Protein Sci.* **9**, 1930–1934 (2000).
- Luo, X. & Hofmann, K. The protease-associated domain: A homology domain associated with multiple classes of proteases. *Trends Biochem. Sci.* **26**, 147–148 (2001).
- Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- McCafferty, C. L., Pennington, E. L., Papoulas, O., Taylor, D. W. & Marcotte, E. M. Does AlphaFold2 model proteins' intracellular conformations? An experimental test using cross-linking mass spectrometry of endogenous ciliary proteins. *Commun. Biol.* **6**, 421 (2023).
- Gao, M., An, D. N., Parks, J. M. & Skolnick, J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat. Commun.* **13**, 1744 (2022).
- Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
- Kim, H. *et al.* Homomeric interaction of AtVSR1 is essential for its function as a vacuolar sorting receptor. *Plant Physiol.* **154**, 134–148 (2010).
- Kundu, S. & Jernigan, R. L. Molecular mechanism of domain swapping in proteins: An analysis of slower motions. *Biophys. J.* **86**, 3846–3854 (2004).
- Bennett, M. J., Choe, S. & Eisenberg, D. Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci.* **91**, 3127–3131 (1994).
- Bennett, M. J., Choe, S. & Eisenberg, D. Refined structure of dimeric diphtheria toxin at 2.0 Å resolution. *Protein Sci. Publ. Protein Soc.* **3**, 1444–63 (1994).
- Martin, J. L. Thioredoxin—a fold for all reasons. *Structure* **3**, 245–250 (1995).
- Cao, H., Duncan, O. & Millar, A. H. Protein turnover in the developing Triticum aestivum grain. *N. Phytol.* **233**, 1188–1201 (2022).
- Chothia, C. One thousand families for the molecular biologist. *Nature* **357**, 543–544 (1992).
- Handford, P. A. *et al.* Key residues involved in calcium-binding motifs in EGF-like domains. *Nature* **351**, 164–167 (1991).
- Knott, V., Downing, K. A., Cardy, C. M. & Handford, P. Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1. *J. Mol. Biol.* **255**, 22–27 (1996).

31. Wouters, M. A. *et al.* Evolution of distinct EGF domains with specific functions. *Protein Sci.* **14**, 1091–1103 (2005).
32. Watanabe, E. *et al.* An ER-localized form of PV72, a seed-specific vacuolar sorting receptor, interferes the transport of an NPIR-containing proteinase in Arabidopsis leaves. *Plant Cell Physiol.* **45**, 9–17 (2004).
33. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
34. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D* **75**, 861–877 (2019).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D* **66**, 486–501 (2010).

Acknowledgements

We thank the staff at the SSRL beamline for synchrotron data collection. We would like to express sincere gratitude to Dr. John C. Rogers and Sally W. Rogers for their initial work for the study and providing purified AtVSR1, and Dr. Marisa Otegui (University of Wisconsin-Madison) for her insightful discussion of the manuscript. Original research was supported by NSF (MCB-2043248), USDA-NIFA (2023-67013-39629, 2003-35318-13672).

Author contributions

HP collected and analyzed data, and wrote manuscript. BY prepared crystals and analyzed data. DJP analyzed data and wrote manuscript. SVP analyzed data. CK initiated the project, secured funding, analyzed data, and wrote manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-57975-2>.

Correspondence and requests for materials should be addressed to C.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024