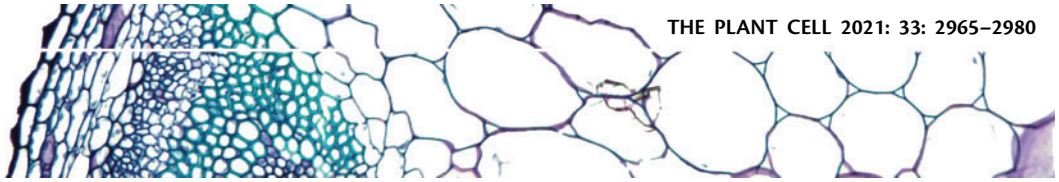
doi:10.1093/plcell/koab182



THE PLANT CELL 2021: 33: 2965–2980

Large-Scale Biology

A co-fractionation mass spectrometry-based prediction of protein complex assemblies in the developing rice aleurone-subaleurone

Youngwoo Lee [1], Thomas W. Okita [2] and Daniel B. Szymanski [1,3,*,†]

1 Department of Botany and Plant Pathology, Center for Plant Biology, Purdue University, West Lafayette, Indiana 47907, USA
2 Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164, USA
3 Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

*Author for correspondence: szymandb@purdue.edu
†Senior author

Y.L. and D.B.S. designed the project; Y.L. performed experiments; Y.L., T.W.O., and D.B.S. analyzed the data and wrote the article.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell) is: Daniel B. Szymanski (szymandb@purdue.edu).

Abstract

Multiprotein complexes execute and coordinate diverse cellular processes such as organelle biogenesis, vesicle trafficking, cell signaling, and metabolism. Knowledge about their composition and localization provides useful clues about the mechanisms of cellular homeostasis and system-level control. This is of great biological importance and practical significance in heterotrophic rice (*Oryza sativa*) endosperm and aleurone–subaleurone tissues, which are a primary source of seed vitamins and stored energy. Dozens of protein complexes have been implicated in the synthesis, transport, and storage of seed proteins, lipids, vitamins, and minerals. Mutations in protein complexes that control RNA transport result in aberrant endosperm with shrunken and floury phenotypes, significantly reducing seed yield and quality. The purpose of this study was to broadly predict protein complex composition in the aleurone–subaleurone layers of developing rice seeds using co-fractionation mass spectrometry. Following orthogonal chromatographic separations of biological replicates, thousands of protein elution profiles were subjected to distance-based clustering to enable large-scale multimerization state measurements and protein complex predictions. The predicted complexes had predicted functions across diverse functional categories, including novel heteromeric RNA binding protein complexes that may influence seed quality. This effective and open-ended proteomics pipeline provides useful clues about system-level posttranslational control during the early stages of rice seed development.

Introduction

Rice (*Oryza sativa*) is one of the three major food crops of the world (FAO, 2003), and its aleurone tissue and endosperm are an essential source of human nutrition. Endosperm development consists of coenocytic nuclear division, cytokinesis, and differentiation into the starchy endosperm and aleurone layer (Olsen, 2004; Wu et al., 2016). Like most cereal grains, rice has an aleurone composed of a single layer of cells that differentiates from the endosperm epidermis at 5 days after flowering (DAF) and is completely formed at 7 DAF (Krishnan and Dayanandan, 2003; Wu

Received March 23, 2021. Accepted July 01, 2021. Advance access publication July 16, 2021

Open Access

IN A NUTSHELL

Background: Many proteins act as components of multiprotein complexes that coordinate cellular processes such as vesicle trafficking, cell signaling, and metabolism. Moreover, the protein-protein interaction network functions as a dynamic system that can rearrange to enable productive responses to developmental or environmental changes. However, at present, knowledge about interaction networks is sparse.

Question: We set out to determine if co-fractionation mass spectrometry (CFMS) can be used to generate reliable predictions about protein complex compositions in developing aleurone-subaleurone layers of rice (*Oryza sativa*) seeds.

Findings: The CFMS pipeline utilizes biological replicates, reproducibility filters, and orthogonal separations of soluble extracts to reduce noise and the confounding effect of chance co-elution. Based on a distance-based clustering and a validated classification system, the compositions of 770 reliable rice protein complex predictions were generated. The predictions included both known complexes/subcomplexes with diverse functions and many new ones. This snapshot of protein multimerization provides a systems-level view into the development and physiology of tissues crucial to rice seed quality.

Next steps: This approach and the associated data have significant value for the research community to broadly test hypotheses about protein complex function across species and cell types. This open-ended proteomics pipeline is being further developed to analyze the compositions of membrane-associated complexes and those that change in response to mutation and environmental stress.

et al., 2016). During the grain filling stage, the aleurone and subaleurone tissues store proteins, lipids, vitamins, and minerals, whereas the endosperm mainly accumulates starch (Krishnan and Dayanandan, 2003; Becraft and Yi, 2010; Wu et al., 2016). The subaleurone tissue consists of four to six cell layers that are biochemically distinct from the rest of the starchy endosperm. These cells are much lower in starch content and accumulate the bulk of storage proteins, especially glutelins and prolamines. The aleurone also protects the endosperm during the grain filling stage, and its induced desiccation tolerance maintains stored starch in the endosperm and ensures seed survival (Fath et al., 2000; Young and Gallie, 2000; Bethke et al., 2001). Given the importance of the aleurone–subaleurone for human nutrition and seed development, we focused on this tissue in this high-throughput analysis of protein complex formation and composition using quantitative proteomics.

Data on protein multimerization and binding partner identity are some of the most valuable to analyze regulatory pathways and interactions among them. Genetic data indicate that subunits of a protein complex and interacting proteins share similar phenotypes and function as parts of signaling input–output modules (Yanagisawa et al., 2018). Protein multimerization is the cornerstone of cellular complexity, enabling mechanical tasks and the flow of genetic information that could never be achieved by individual proteins (Alberts, 1998; Marsh and Teichmann, 2015). For example, multiprotein complexes coordinate gene expression (Burd and Dreyfuss, 1994), organelle biogenesis (Li et al., 2019), vesicle trafficking (Kaksonen et al., 2005), metabolism (Weng et al., 2012), and signal transduction (Basu et al., 2008). Recent proteomic profiling studies indicated that more than one-third of all proteins exist as parts of stable protein complexes (Aryal et al., 2014, 2017; McBride et al., 2017, 2019; Lee and Szymanski, 2021). Due to the widespread occurrence and crucial roles of protein multimerization, numerous large-scale projects to characterize protein–protein interaction networks have been conducted by yeast two-hybrid analysis (Van Leene et al., 2007; Arabidopsis Interactome Mapping Consortium, 2011; Jones et al., 2014). However, comparisons of global protein–protein interaction studies have shown that combinations of approaches are needed. Overlap among interactome dataset types is low, and technical bias and the limitations of individual methods determine the extent to which they capture the full spectrum of physical interactions, which vary greatly in terms of affinity, cell type, or subcellular localization (Wodak et al., 2009; Rattray and Foster, 2019; Salas et al., 2020).

Protein correlation profiling, also known as co-fractionation–mass spectrometry (CF–MS), is gaining momentum as an effective approach to broadly analyze the multimerization behaviors of endogenous proteins (Kristensen et al., 2012; Aryal et al., 2014; McWhite et al., 2020; Salas et al., 2020). The major advantage of CF–MS is that it analyzes native protein complexes in an unbiased way, with no requirement for genetic transformation or gene cloning. It is a guilt-by-association protein chromatography method based on the expected indistinguishable elution profiles of subunits of stable protein complexes regardless of the separation method. This MS-based profiling of cell lysates separated by size exclusion chromatography (SEC) provides broad information on whether or not a protein is likely to multimerize (Aryal et al., 2014, 2017; Gilbert and Schulze, 2019), form distinct complexes at different subcellular locations (McBride et al., 2017), or evolve unique properties in diverse species (Lee and Szymanski, 2021).

The chance co-elution of noninteracting proteins is a confounding factor that increases the rate of false-positive predictions. One approach to improve accuracy is to carry out multiple fractionations and use presumed gold standards of evolutionarily conserved protein complexes in combination with machine learning-based predictions (Havugimana et al., 2012; Wan et al., 2015; McWhite et al., 2020). There is uncertainty about the extent to which gold-standard complexes exist as stable, fully assembled complexes (Aryal et al., 2014; McBride et al., 2017; Lee and Szymanski, 2021); however, the increasing number of validated known complexes will further empower these approaches. Our group predicted multimerization and complex composition based on experimental profile data alone (Aryal et al., 2014, 2017; McBride et al., 2017, 2019). In this workflow, biological replicates and automated peak detection algorithms (McBride et al., 2017, 2019) are used to remove unreliable profiles. The filtered data are used in distance-based clustering analyses to group proteins with the most similar elution profiles. This overall approach has been validated repeatedly using known complexes or subcomplexes, co-immunoprecipitation, and profiling a mutant in which the expression of a predicted novel subunit was knocked out (Aryal et al., 2014; McBride et al., 2017, 2019). This approach generates a valuable data resource for the community to develop and test hypotheses about protein interactions and system-level controls.

Here, we adopted this technology to predict protein complex assemblies from a dissected tissue with great developmental and agronomic importance: the aleurone–subaleurone cell layers of developing rice seeds. This pipeline begins with 1 mg of soluble protein extracts, whereupon 2,610 proteins were reproducibly profiled across two SEC and two ion-exchange chromatography (IEX) column separations. This combination of orthogonal separation strategies greatly reduces chance co-elution and can generate reliable protein complex composition predictions. The clustering analysis and systematic classification method predicted 771 protein complexes (proteins of interest can be searched in Supplemental Data Set S1), 170 of which correspond to self-interacting proteins. Numerous novel protein complexes involved in translation, cellular homeostasis, and tissue-specific physiology were predicted. These protein complexes could play essential roles in determining the fate of the aleurone, regulating endosperm development, and in the biosynthesis of seed reserves. These findings will facilitate a deeper understanding of seed development and quality.

Results and discussion

A high-quality CF–MS dataset

We used the CF–MS approach, coupled with biological replicates of SEC and IEX fractionations, to profile endogenous protein complexes in the aleurone–subaleurone layers of developing rice seeds (Figure 1). The expressed proteome in the developing seeds at 10 DAF was resolved under native conditions through SEC and orthogonal IEX separations to reduce chance co-elution and false-positive predictions

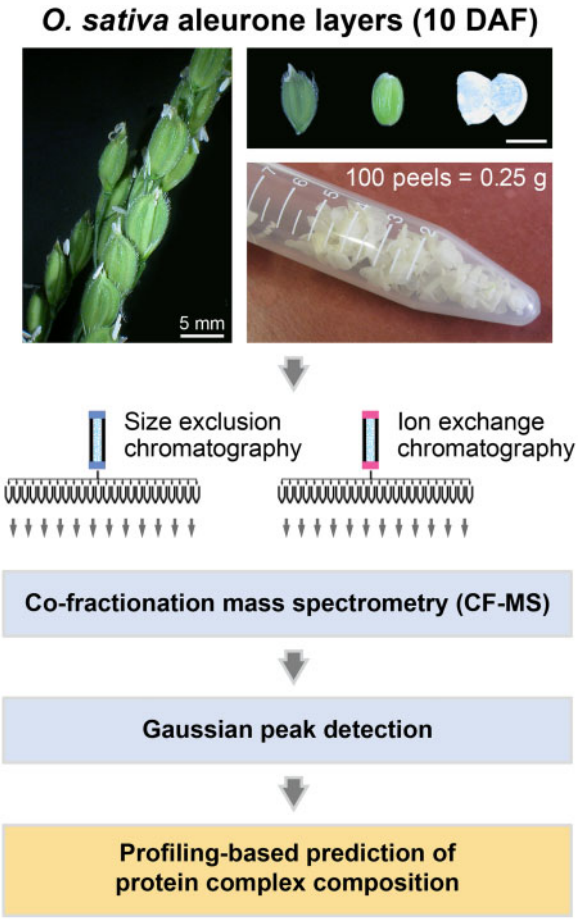


Figure 1 The CF–MS pipeline, composed of SEC and IEX separations, used to predict protein complex composition in the rice seed aleurone–subaleurone layers. The soluble cell fraction enriched from the isolated aleurone–subaleurone layers at 10 DAF is separated on a sizing column and a mixed-bed ion exchange column under nondenaturing conditions. Each column fraction is analyzed by LC–MS/MS for protein identification and quantification. Gaussian fitting is applied to choose reproducible peaks in the SEC and IEX datasets. M_{app} values are calculated for the reproducible SEC peaks. Profile-based clustering analysis is conducted to predict protein complex composition using the concatenated SEC and IEX datasets.

(McBride et al., 2019). We subjected biological duplicates of 24 SEC and 71 IEX fractions to label-free shotgun proteomics to obtain abundance profiles of thousands of resolved proteins. The abundance profiles of all peptides and proteins are provided in Supplemental Data Set S2. These elution profiles were subjected to Gaussian fitting to smooth the raw data and deconvolve multiple peaks, which likely reflect an individual protein being present in multiple protein complexes (McBride et al., 2017). With a 1% false discovery rate (FDR) for both the peptide and protein levels, 3,746 and 3,633 endogenous rice aleurone–subaleurone proteins were reproducibly identified from the SEC and IEX fractions, respectively (Figure 2, A and B). Overall, the SEC and IEX profile data were highly reproducible, as the Pearson correlation coefficients (PCCs) fell onto a diagonal across the SEC and IEX fractions (Figure 2, C and D). The elution peak locations

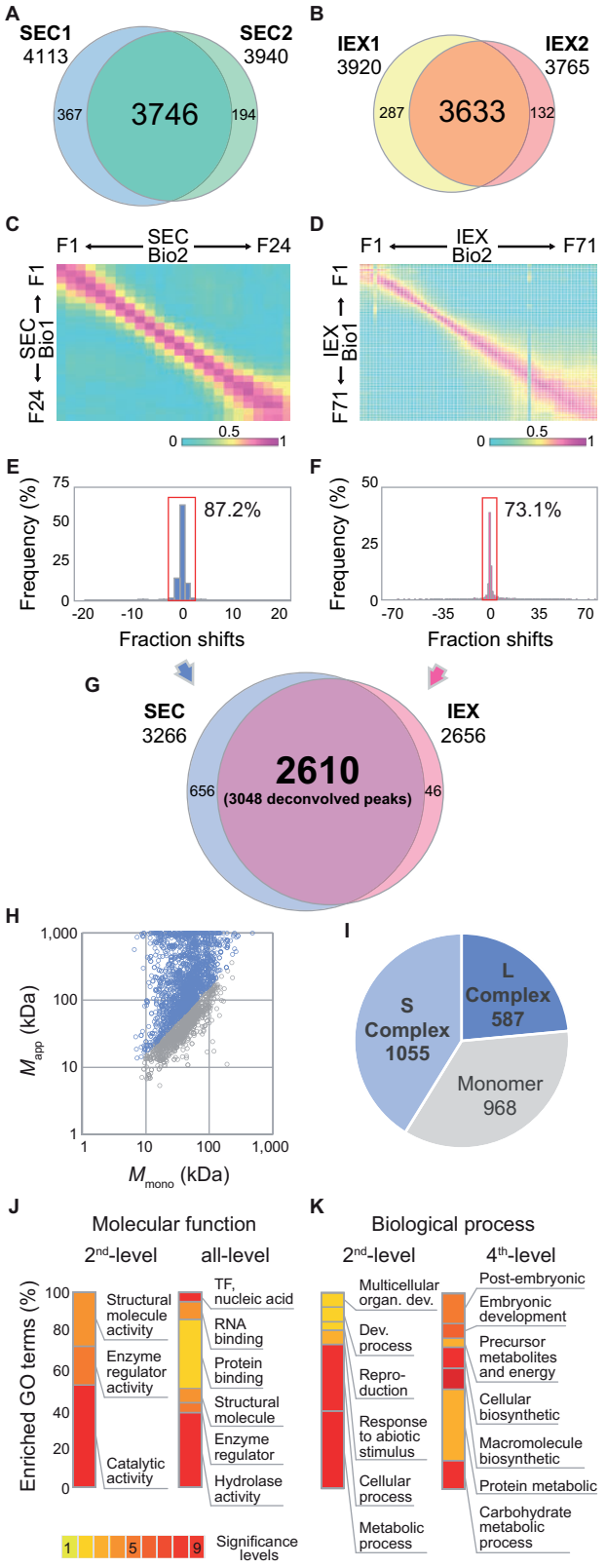


Figure 2 The CF–MS pipeline generated a highly reproducible dataset of protein complexes for the rice aleurone–subaleurone proteome. A–G, Reproducibility of the CF–MS datasets. Protein overlap between two biological replicates in SEC (A) and IEX (B). PCC of protein abundances across the SEC (C) and IEX (D) fractions. Column fraction shifts were measured based on the distances between the elution peaks of the replicates for all proteins in the SEC (E) and IEX (F) fractions. Reproducible proteins present within 2- or 4-fraction shifts are boxed with the ratio of reproducibility shown. G, Reproducible proteins that overlapped between the SEC and IEX datasets were selected for the prediction of protein complex composition. H, I, Protein complexes are common in the rice aleurone–subaleurone proteome. The quantification of protein multimerization was performed using M_{app} and R_{app}. H, A scatter plot of the M_{mono} and M_{app} of the reproducible proteins. Open circles in light blue are proteins with $R_{app} \geq 1.6$, while those in grey are proteins with $R_{app} < 1.6$. I, Distribution of protein multimeric states. Proteins are classified as M (monomer: $0.62 \leq R_{app} < 1.6$), S (small complex: $1.6 \leq R_{app} < 5$), or L (large complex: $R_{app} \geq 5$). J, K, The aleurone–subaleurone proteome shows functions in diverse processes. Overrepresented GO terms are highlighted according to the significance levels. Molecular function GO terms (J) and biological process GO terms (K) are visualized. Full GO analysis results are provided in Supplemental Figure S1.

were also reproducible between replicates (Figure 2, E and F). When we compared the peak locations for individual proteins, ~88% and 73% of the total protein peaks fell within two SEC fractions (representing ~40% size difference) and within four IEX fractions, respectively. The reproducible peak locations and their apparent mass (M_{app}) values are provided (Supplemental Data Set S3). To eliminate noisy profile data, the set of 2,610 proteins identified with reproducible peaks between both SEC and IEX replicates was chosen for protein complex prediction (Figure 2G).

Protein multimerization in the rice aleurone–subaleurone proteome

To estimate the proportion of proteins that eluted at larger than expected masses, we compared the distribution of expected monomeric masses (M_{mono}) with the distribution of M_{app} (Figure 2H). The scatter plot of M_{app} and M_{mono} was strongly skewed toward elevated M_{app} values, suggesting widespread multimerization. Using protein multimerization state, that is, R_{app} ($R_{app} = M_{app}/M_{mono}$) as a diagnostic for multimerization of individual proteins, more than half of the proteins fell into this class, with 37.1% detected as relatively small complexes and 22.5% as large complexes with R_{app} values of ≥ 5 (Figure 2I). Similar distributions of stable protein complexes were reported in different tissues and plant species (Aryal et al., 2014; McBride et al., 2017, 2019), although the types and sizes of complexes can vary within an orthologous group (Lee and Szymanski, 2021).

To gain insights into the types of proteins in the rice aleurone–subaleurone proteome, we performed gene ontology (GO) enrichment analysis, which revealed 70 significantly enriched terms at a 5% FDR (Figure 2, J and K; Supplemental Figure S1). In the molecular function GO category, enzyme activities including catalytic activity and enzyme regulator activity were overrepresented in the developing aleurone–subaleurone layer cells, reflecting many interesting enzymes, lipid binding, RNA binding, and signaling proteins present in the tissue (Figure 2J). In the biological process category, four GO terms were the most highly enriched: cellular biosynthetic

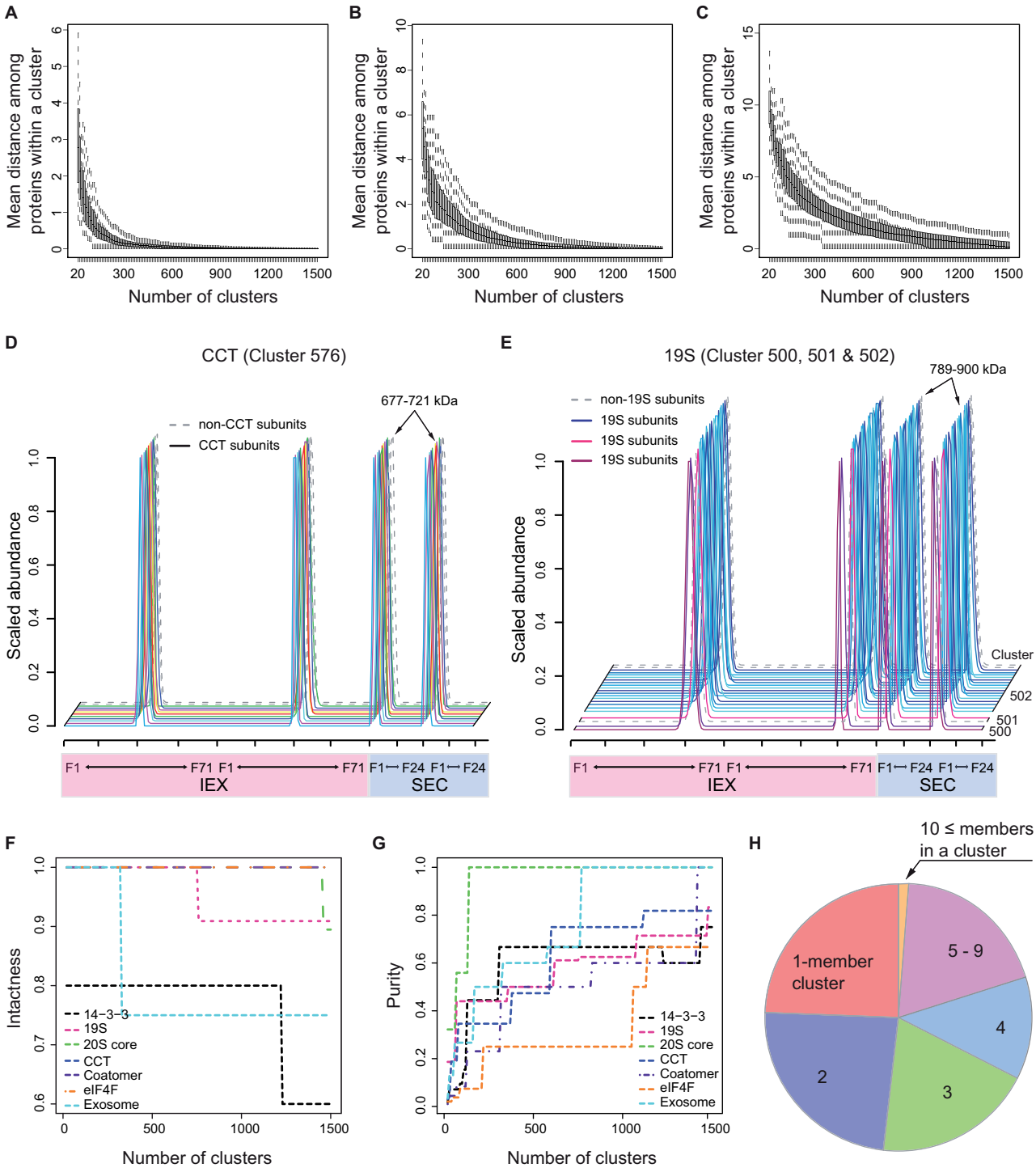


Figure 3 The strong resolving power of clustering analysis evaluated by the intrinsic and extrinsic tests. A–C, Intrinsic tests used to evaluate the resolving power for independent and concatenated protein profile datasets. Box-and-whisker plots visualize the distributions of average distances of elution profiles within the clusters in the SEC only (A), IEX only (B), and combined SEC and IEX (C) datasets as a function of ascending cluster number. D and E, The concatenated profiles of known protein complexes are visualized over the two IEX (pink) and two SEC (light blue) separations. Solid lines indicate elution profiles of CCT (D) and 19S proteasome complex (E) subunits. Dashed lines represent elution profiles of proteins that are not subunits of the known complexes. Numbers shown above the SEC peaks show M_{app} values of subunits in a given complex. Numbers to the right of the profiles are cluster numbers when multiple clusters are plotted. F and G, Extrinsic tests evaluating the resolving power for the independent and combined protein profile datasets as a function of ascending cluster number. F, The intactness of known complexes was measured. G, The purities of the known complexes were evaluated. Individual intactness and purity plots are provided in Supplemental Figure S2. H, The distribution of cluster size in the resulting clustering. The pie chart shows the distribution of predicted protein complexes based on the number of members in each cluster.

process, macromolecule biosynthetic process, carbohydrate metabolic process, and response to abiotic stimulus (Figure 2K). These overrepresenting terms reflect the notion that proteins in the aleurone–subaleurone layers promote the rapid biosynthesis of carbohydrates, proteins, and other seed storage substances. A similar pattern of GO term distribution was reported in a gene expression analysis of the developing aleurone in wheat (*Triticum aestivum*; Gillies et al., 2012). Moreover, two other highly enriched GO terms describe proteins that regulate embryonic development and postembryonic development (Supplemental Figure S1). During wheat seed development, these terms were highly enriched in aleurone layers compared to starchy endosperm cells (Gillies et al., 2012). These analyses indicate that a broad range of protein types with aleurone–subaleurone-enriched functions is captured in our dataset.

Profiling-based clustering and predicting protein complex compositions

To begin our distance-based clustering analysis, elution profiles with multiple peaks were deconvolved into separate profiles as previously described (McBride et al., 2019). Three hundred seventy-four proteins had 2 IEX peaks, and 32 had 3; these multipeak proteins were annotated with a peak number suffix in the locus identification (ID). Among these 406 multiple peak proteins, 57 had multiple SEC peaks. These proteins were assigned the same SEC peak with the largest M_{app}, enabling a protein to reside in multiple distinct protein complexes. In the end, the four SEC and IEX elution profile datasets were concatenated, and a total of 3,048 reproducible profile entries were used for further analysis.

As an intrinsic test of the enhanced resolution afforded by combining SEC and IEX, we analyzed the mean distances between clusters as a function of the numbers of clusters in the individual and combined datasets (Figure 3). When the elution profiles of the proteins in a cluster are similar to each other, the average distance among them is low. In the box plot of mean distance within a cluster for SEC or IEX alone, the third quartile approached zero at approximately 530 or 630 clusters, respectively (Figure 3, A and B). The combination of SEC and IEX fractionations increased the resolution so that the third quartile approached zero at 1,000 clusters (Figure 3C). These data demonstrate the utility of the combined SEC and IEX datasets and indicate that the dendrogram loses resolution beyond approximately 1,000 clusters. As a representative clustering result, we plotted dendrograms at a 1,000 cluster-cut (as shown in Supplemental File S1) and constructed a heatmap with the color code representing the relative protein abundance (Supplemental File S2).

Validation of protein complex prediction using known complexes

In the concatenated SEC and IEX separations, subunits of stable known complexes should co-elute and could serve as a partial validation of the prediction. Known chaperonin-containing TCP1 (CCT) folding complex subunits co-eluted and were assigned to cluster 576 with M_{app} of 677–721 kDa corresponding to the fully assembled complex (Figure 3D). Another validating known is 19S proteasome cap complex. Seventeen different 19S subunits with M_{app} of 789–900 kDa were also grouped into cluster 502 (14 of 19S subunits/16 of cluster members) and neighboring clusters 500 (2/2) and 501 (1/2) (Figure 3E). The 19S subunits in clusters 500 and 501 also had similar M_{app} to other subunits, but a slight shift in the IEX fractions assigned these three subunits into the adjacent clusters. Further discussion of the CCT and 19S complexes is provided in Supplemental File S3.

To further inform the decision of where to divide the dendrogram to generate a specific protein complex prediction using extrinsic data, we evaluated the intactness and purity of seven known complexes as a function of increasing cluster number (Figure 3, F and G; Supplemental Figure S2). The intactness of exosome and 19S proteasome subunits dropped in intactness at around 300 and 750 clusters but remained stable until 1,500 clusters. The 14-3-3 proteins had stable intactness until approximately 1,250 clusters. Coatomer, CCT, eIFs, and 20S proteasome maintained perfect intactness until around 1,450 clusters. Because purity can be utilized to detect false positives (McBride et al., 2019), the purity of known complexes was observed as cluster numbers increased to 1,400 clusters. The purity of most known complexes became stable after 1,100 clusters. The intactness and purity indices indicated that splitting the dendrogram into 1,000 clusters based on the intrinsic resolution test would be appropriate. Approximately 20% of clusters were predicted to be large complexes with 5–19 subunits, ~55% were 2- to 4-meric, and 224 proteins were assigned into single-entry clusters (Figure 3H). The reduced number of singletons compared to the predicted number of 400 (40% monomeric × 1000 clusters, estimation from Figure 2I) indicates the common occurrences of false positives. The baseline complex composition prediction from this study is provided in Supplemental Data Set S1A, with the caveat that true interactors may be located in nearby clusters, and clusters will also contain false positives due to chance co-elution. A classification system to guide the user is provided below, and we encourage readers to comment on the paper online as clusters are validated or refuted over time.

Protein complex heterogeneities and cross-validation using multiple peak entries

A small number of proteins displayed multiple peaks, and this could reflect the existence of heteromeric complexes with differing assembly states on the IEX column. It is also possible that functionally interchangeable orthologs/paralogs could assemble into homomers or ortholog-selective multimers with differing subunit stoichiometries and resolvable peaks on the IEX column. In the above scenarios, an accurate clustering result would place interacting proteins within

the same cluster in two separate instances. There were eight cases in which a pair of multiple peak proteins co-occurred in two distinct clusters (Supplemental Data Set S1B). In the first three cases listed, two proteins in a ribosomal protein S2 cluster, three in a DnaK family cluster, and two in another DnaK family cluster had two IEX peaks and two SEC peaks, neither of which corresponded to an expected monomer. This could be explained by protein assembly into two complexes with partial subunit overlap that were resolved on both columns, supporting true protein–protein interactions between these multiple peak proteins.

Other instances in which pairs of proteins had a single high mass SEC peak and multiple IEX peaks could reflect a complex that partially dissociated during high salt elution. Two proteins of the 19S proteasome cap complex (RPN12 [LOC_Os07g25420.1] and RPN9 [LOC_Os01g32800.2 and LOC_Os03g11570.1] in cluster 502) were also assigned into cluster 518. A solved structure (Lander et al., 2012), native MS analysis (Sharon et al., 2006), and a CF–MS prediction (Drew et al., 2017) did not show a direct interaction between RPN9 and RPN12. However, RPN9 and RPN12 possess significant sequence homologies with two interacting subunits of the COP9 signalosome, the CSN7 and CSN8 subunits, respectively (Kapelari et al., 2000; Fu et al., 2001), and the adjacency of RPN9 and RPN12 subunits in the lid structure (da Fonseca et al., 2012; Lander et al., 2012) could enable a stable physical interaction in some species. In another example, two eIFiso4F subunits showed similar behavior to the 19S pair. The large subunit eIFiso4G (LOC_Os04g42140.1) and the small cap-binding protein eIFiso4E (LOC_Os10g32970.1) co-occurred in two clusters with a single SEC peak. This result is consistent with their known direct physical interaction (Mayberry et al., 2011), further supporting the accuracy of the predictions in this study.

Two homologous pyruvate kinases (OsPKpα1 [LOC_Os07g08340.1] and OsPKpβ2 [LOC_Os10g42100.1]) had two IEX peaks, a single SEC peak, and co-occurred in two clusters. Mammalian pyruvate kinases are known homotetramers (Larsen et al., 1998; Christofk et al., 2008), while plant orthologs are obligate heterooligomers of ancient paralogs (Negm et al., 1995; Andre et al., 2007; Cai et al., 2018). The clustering pattern here could reflect partial disassembly of a multimeric form during high salt elution. *Arabidopsis thaliana* 14-3-3 paralogs showed IEX-resolved heteromerization patterns (McBride et al., 2019). Taken together, these co-clustering pairs provide clear cross-validations for our protein complex predictions.

Cross-species comparisons with published CF–MS datasets

We made cross-comparisons of our rice predictions with previously published CF–MS datasets from McWhite et al. (2020) and Arabidopsis leaf data from McBride et al. (2019). The clustering result presented here and that of McBride operated on protein groups defined by unique peptides. Both predictions were based on the profiled data from approximately 200 fractions generated from duplicates of SEC and IEX separations. The McWhite prediction was generated from approximately 2,000 fractions from 13 plant species and diverse tissues using four different separation methods. Orthologs and paralogs were merged into an averaged profile of a single ortholog group. Ortholog averaging obscures the commonly observed multimerization variability among paralogs and orthologs (Lee and Szymanski, 2021). The protein coverage of our rice dataset was three times greater than those of the previous studies due to the increasedliquid chromatography mass spectrometry (LC–MS) sensitivity and the use of protein groups. The key parameters for the three CF–MS studies are summarized in Supplemental Data Set S4A. The overlapping protein hits discussed here are summarized in Supplemental Data Set S4, B and C.

Differences in protein definitions and complexity make direct comparisons among studies difficult. In this comparison, we looked for two distinct protein/ortholog group members present within a single cluster in both the rice and the comparison studies. If the interaction was reproduced in both studies, they should fall within a single cluster in both instances. There were 31 rice pairs that were also present in the McBride Arabidopsis dataset, and 29% fell into a single cluster in both cases. There were 149 rice pairs that were in the McWhite Arabidopsis dataset, and 24% fell into a single cluster in both cases. Of the 144 rice pairs that were in the McWhite rice prediction, 25% co-occurred in the same cluster. Similar results are expected with the McWhite prediction regardless of species due to the ortholog merging. Pairs present across all three datasets were subunits in 20S, 19S, CCT, and coatomer complexes, showing highly conserved protein–protein interactions. The McBride prediction assigned subunits of exon junction complex, CCT, and HSP70 into a single cluster, while these complexes were resolved into three distinct clusters in our rice prediction and the McWhite datasets. Our clustering had more prediction resolution and protein coverage than both published CF–MS datasets, suggesting a better overall complex prediction.

Systematic classification of predicted rice protein complexes

We performed systematic cluster classification based on the sum of M_{mono} values of all proteins within one cluster (M_{calc}) and the measured M_{app} (Figure 4; Supplemental Data Set S1A). A reliable class contains proteins in the classes "homomer" and "possible homomer or heteromer/high subunit stoichiometry" categories. These proteins had elution profiles that placed them into small clusters of one to three proteins, yet on the SEC column, these proteins had very large M_{app} values (Figure 4A, area highlighted as light yellow). These data suggest a high subunit stoichiometry and/or the formation of homomers. Known homomers with M_{app} values corresponding to their expected stoichiometries and numerous novel homomers were identified from the single-entry clusters (Figure 4A–G). The indole acetic acid (IAA)-amino acid hydrolase ILR1-like 6 (ILL6) was assigned

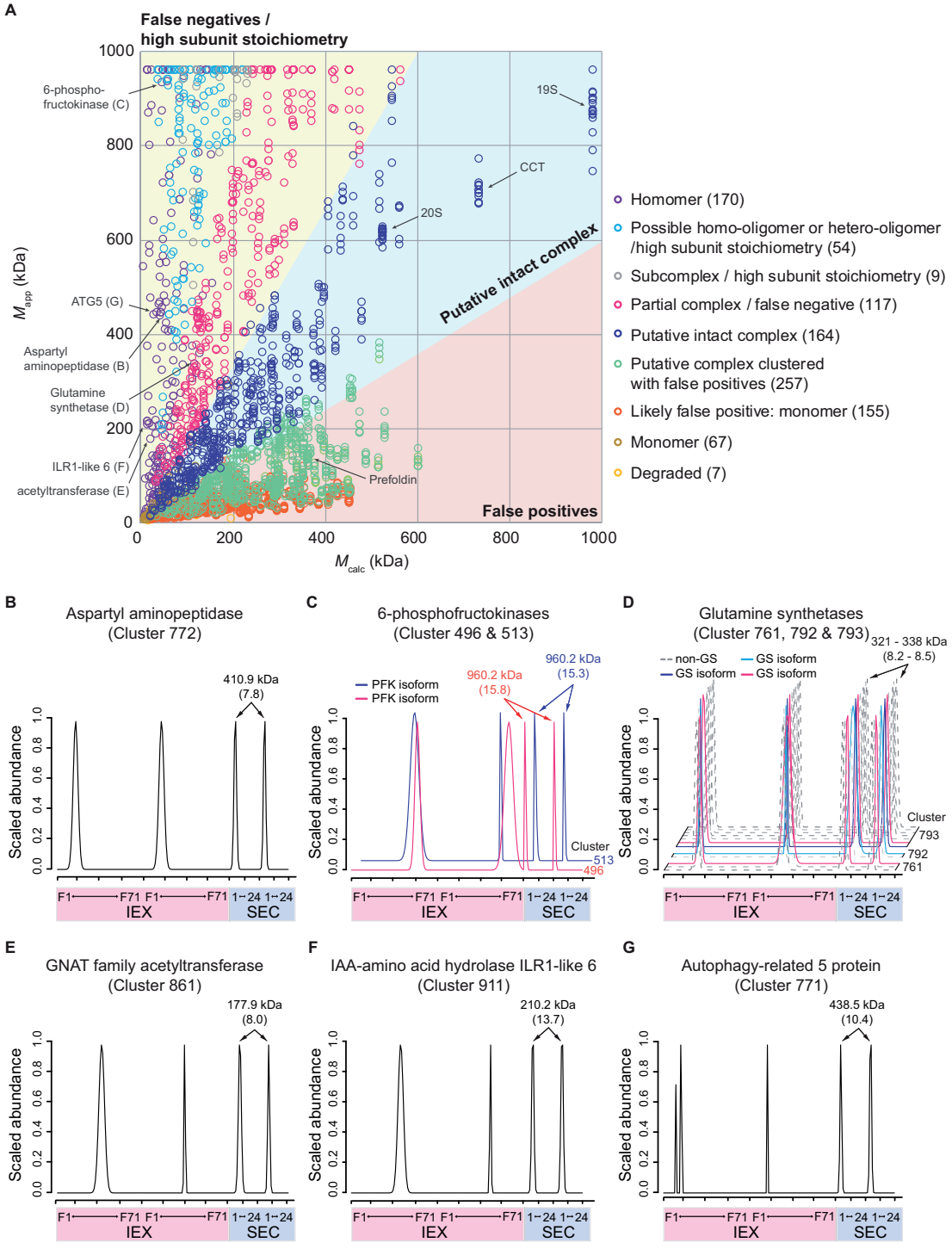


Figure 4 Systematic classification of the resulting clustering analysis to assess the stoichiometry of subunits in the predicted protein complexes. A, Clustering analysis was evaluated by comparing M_{calc} and M_{app}. The most reliable predictions are shown in the light blue area, where there are less than two-fold size differences between M_{calc} and M_{app}. Other reliable predictions are shown in the light yellow area, where M_{app} is two-fold greater than M_{calc}, including "false negatives" and "proteins with high subunit stoichiometry." Proteins present in the "putative complexes with false positives" category were saved from the false positives in the light pink region, where M_{calc} is two-fold greater than M_{app}. Number in parenthesis next to each category indicates the number of clusters that belong to the category. B–G, Many self-interacting homomers were assigned into single entry clusters and classified as homomers. B–D, The profiles of known homomers: Aspartyl aminopeptidase (B), 6-phosphofructokinase isoforms (C), and glutamine synthetase isoforms (D). Dashed lines represent elution profiles of proteins that are not subunits of the known complexes. E–G, The profiles of novel homomers: GNAT family acetyltransferase (E), IAA–amino acid hydrolase ILR1-like 6 (F), and autophagy-related 5 protein. (G), Profiles of homomers are visualized over the combined two IEX (pink) and two SEC (light blue) separations. Numbers shown above the SEC peaks show M_{app} values of homomers. Numbers in parenthesis indicate R_{app} values of homomers. Numbers to the right of the profiles are cluster numbers when multiple clusters are plotted.

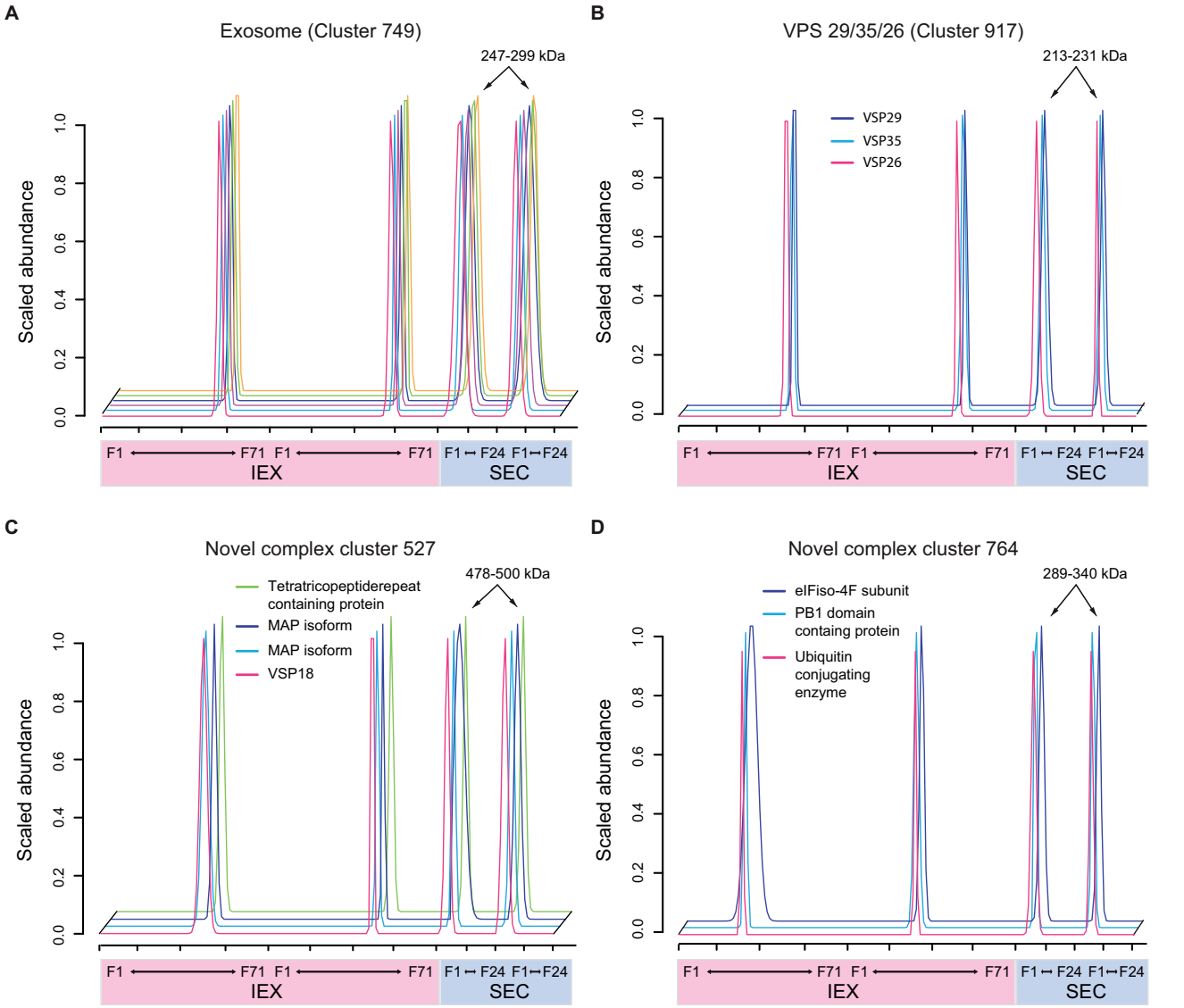


Figure 5 Elution profiling-based clustering placed known or novel protein complex subunits into the "putative intact complex" clusters. The profiles of known and novel protein complexes overlapped with each other across two IEX (pink) and two SEC (light blue) separations. The clusters were classified as "putative intact complexes." A and B, Known protein complexes. Elution profiles of five exosome subunits (A) and a heterotrimeric VPS complex (B) are visualized. C and D, Novel protein complexes. Elution profiles of a MAP containing heterotetrameric complex (C) and a heterotrimeric novel complex (D). Numbers shown above the SEC peaks show M_{app} values of subunits in a given complex.

into a single-member cluster with R_{app} of 13.7 (Figure 4F). In Arabidopsis, ILL6 is a member of an amidohydrolase family, which hydrolyzes not only IAA conjugates (LeClere et al., 2002; Zhang et al., 2016) but also jasmonoyl–isoleucine conjugates upon wounding (Widemann et al., 2013). ILL6 oligomerization might mediate crosstalk between the IAA and jasmonic acid signaling pathways (Zhang et al., 2016). A discussion of a subset of these interesting homomers, GCN5-related N-acetyltransferases and autophagy-related 5 protein, is provided in Supplemental File S3. This systematic, reliable classification method has the power to distinguish self-interacting proteins from single-entry clusters that resemble monomers in the hierarchical clustering.

Under the assumption of 1:1 subunit stoichiometries, another reliably predicted class is "putative intact complex" because the ratios of M_{calc} to M_{app} for these proteins are close to 1:1. One hundred sixty-four protein complexes fell into an interval along this diagonal (Figure 4A, graph sector highlighted in light blue; Supplemental Data Set S1A). As an example of the "putative intact complex" class, three subunits of the VPS29–VPS35–VPS26 trimeric cargo recognition core complex were assigned into cluster 917, with M_{app} of 213–231 kDa (Figure 5B). The M_{calc} value of the complex was within 40% of the average M_{app} of the VPS cluster. This subcomplex may serve as a regulated cytosolic pool of heterotrimers that assemble with a membrane-associated sorting nexin dimer

subcomplex into a fully functional retromer complex during retrograde transport from the endosome to the Golgi (Seaman et al., 1998). The ratio of M_{calc} to M_{app} and the known complex assembly support the reliability of this clustering analysis. In Supplemental File S3, the relevance of the other protein complex predictions in Figure 5 is discussed.

Approximately 22% of the rice aleurone–subaleurone protein clusters were classified into either the "homomer" or "possible homomer or heteromer/high subunit stoichiometry" category. Approximately 17% fell into the "putative intact complex" category, and 12.6% of the clusters were present in "partial complex/false negatives" and "subcomplex or high subunit stoichiometry." Collectively, ~51% of the clusters were annotated as reliable complex prediction classes (Figure 4A; Supplemental Data Set S1C). The least reliable clusters contained false positives due to chance co-elution (Figure 4A, light pink area). Approximately 26% of the clusters were in the "putative complex clustered with false positives" category, and ~23% of the proteins were filtered from clusters where they were categorized as "likely false positive: monomer," "monomeric," and "degraded" in this clustering analysis. A total of 657 out of 3,048 profiles (or 229 out of 1,000 clusters) were filtered out as likely monomeric proteins from this clustering, and the rest of them (70% of total profiles) were predicted in 771 different protein complexes with different reliabilities.

Predicted subunits of RBP-associated complexes

During the grain filling stage, putative *trans*-acting factors that recognize *cis*-acting elements in the mRNAs of storage proteins (two major rice storage proteins: glutelin and prolamine) maintain the restricted transport of messenger ribonucleoprotein (mRNP) complexes to specific subdomains of the endoplasmic reticulum (Okita et al., 1994; Choi et al., 2000; Crofts et al., 2004; Washida et al., 2012; Tian et al., 2020). In rice, 257 RNA binding proteins (RBPs) were experimentally identified as putative *trans*-acting factors expressed from at least 221 distinct genes (Hamada et al., 2003; Doroshenk et al., 2009; Morris et al., 2011). Even though their dynamic regulation of mRNP complex assembly and disassembly is critical for seed productivity, less is known about whether they are associated with multiprotein complexes or arranged in several smaller complexes. Our profiling identified 133 out of the 250 RBPs with 161 reproducible resolved peaks (27 RBPs exhibited multiple peaks) across the concatenated SEC and IEX separations (Supplemental Data Set S1A). The clustering analysis and classification strategy assigned 92 cytosolic RBPs (with 114 peaks) into 93 distinct protein complex clusters.

The scaffolding-nuclease protein Tudor-SN (LOC_Os02g32350.2), a central player in RNA storage and processing (Sami-Subbu et al., 2001; Gutierrez-Beltran et al., 2016), was clustered with chorismate synthase (CS; LOC_Os03g14990.1) and aspartyl/glutamyl-tRNA amidotransferase subunit B (LOC_Os11g34210.2; Figure 6C). In tobacco BY-2 cells, the same co-elution of Tudor-SN with CS was detected by a combination of ion exchange and gel filtration chromatography (Shan, 2018), and a two-hybrid interaction between rice Tudor-SN and isochorismate synthase (ICS) had been reported (Chou et al., 2017). Chorismate is a key metabolic precursor of salicylic acid (SA), phylloquinone (vitamin K_1), tetrahydrofolate (vitamin B_9), and aromatic amino acids (Tzin and Galili, 2010; Maeda and Dudareva, 2012). ICS converts chorismate to isochorismate en route to the synthesis of SA and vitamin K_1 (Tzin and Galili, 2010). Chorismate and SA metabolism are compartmentalized among the plastid and cytosol, with CS and ICS activities thought to reside solely in the plastid (Mousdale and Coggins, 1986; Strawn et al., 2007; Garcion et al., 2008). However, the functions of Tudor-SN are cytosolic. Perhaps, Tudor–SN complexes mediate feedback control of chorismate-dependent metabolites during mRNA processing steps (Lin et al., 2020). Tudor–SN complexes may also affect the subcellular localization and activity of CS and other enzymes that dictate the flux of chorismate. In addition to the Tudor–SN complex, a subset of novel RBP complexes shown in Figure 6, including RBP-Q-associated putative complex, RBP-T-associated putative EJC, and RBP-149 (eIF2α)-associated complex, is discussed in Supplemental File S3.

Multiprotein complexes coordinate metabolism and cell/tissue structure in the rice aleurone–subaleurone. Here we provided system-level protein complex prediction using a robust CF–MS approach that utilizes biological replicates, reproducibility filters, and orthogonal separations that increase reliability. Using a simple classification system, more than 700 novel protein complex predictions were made. Self-interaction was common, and this type of interaction can provide clues about allosteric control (Llorca et al., 2006) and paths to neofunctionalization over evolutionary time scales (Lee and Szymanski, 2021). Predicted novel heteromeric protein complexes are associated with protein translation, metabolism, signaling, and vesicle trafficking, all of which are crucial for seed development and quality. The data provided here can be broadly leveraged by the research community to generate testable hypotheses about the functional relevance of specific protein–protein interactions. This method can also be further developed to analyze how systems of protein complex assemblies change during development or in response to any desired experimental manipulation.

Materials and methods

Plant growth conditions and soluble protein extraction

The rice (*O. sativa* ssp. *japonica*) cultivar Kitaake was grown in a Conviron E15 growth chamber (Conviron) with a day/night setting of 26°C/22°C and 12/12 h at a light intensity of 300 μmol m^{-2} s^{-1} (fluorescent lamps: Philips F39T5/841/HO/ALTO; incandescent bulbs: GE 60 W light). Seed peels

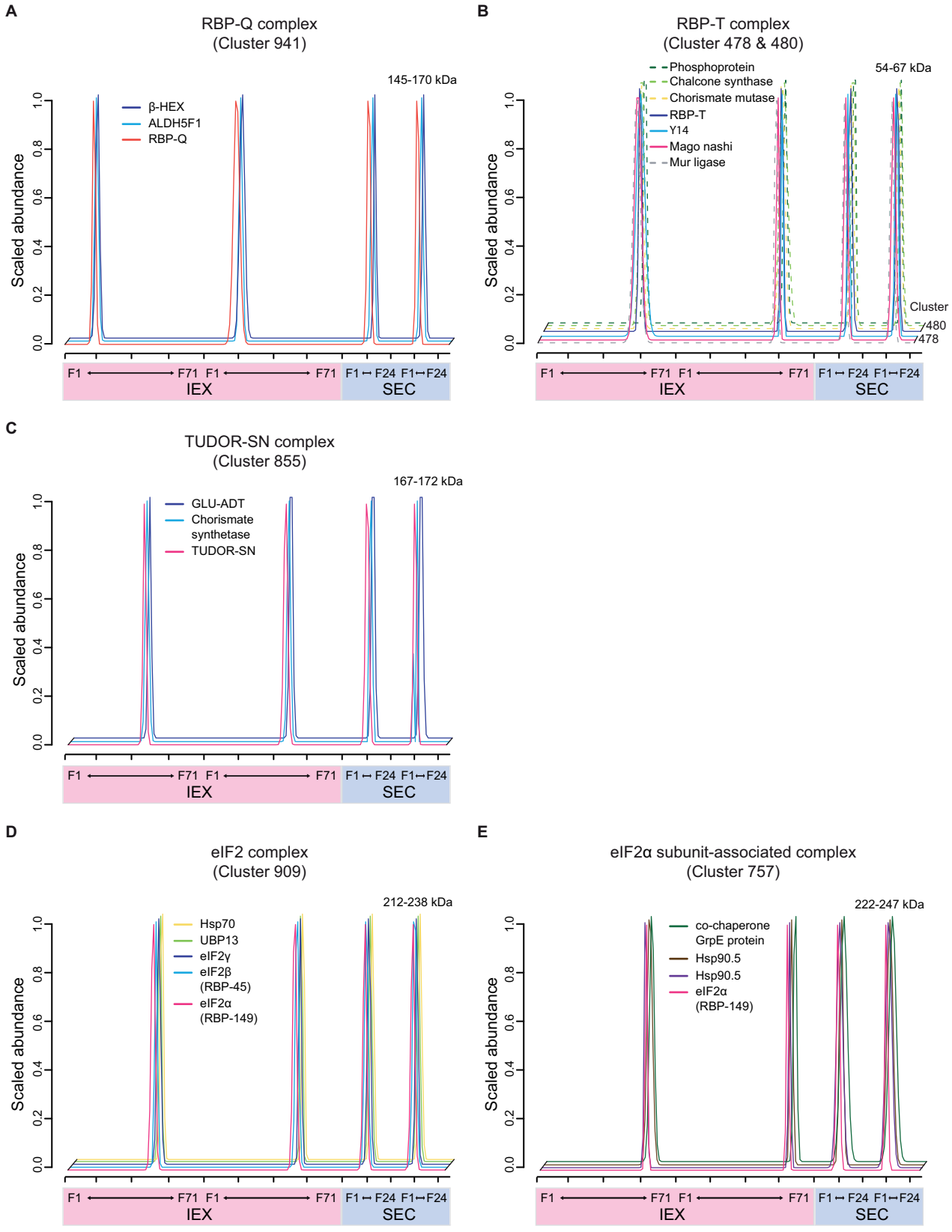


Figure 6 Association of RBPs into novel protein complexes. The elution profiles of subunits in RBP-associated novel protein complexes: RBP-Q associated putative complex (A), RBP-T associated putative EJC (C), Tudor-SN associated putative complex (C), eIF2 complex (D), and eIF2α associated complex (E). Dashed lines represent elution profiles of proteins that are not subunits of the known complexes. Numbers shown above the SEC peaks show M_{app} values of subunits in a given complex. Numbers to the right of the profiles are cluster numbers when multiple clusters are plotted.

containing purified aleurone–subaleurone cell layers were collected as described previously (Yang et al., 2018). Developing seeds at 10–12 DAF were pooled from panicles from 10 different plants. Seeds were dehulled, cut open, and the outer seed layers (pericarp and nucellus) were stripped away. The inner peels were washed in phosphate-buffered saline buffer to remove milky starchy endosperm and embryo. This left a semi-translucent tissue containing primarily aleurone and four to six layers of subaleurone cells. Approximately 250 mg of fresh subaleurone and aleurone tissue was prepared for cell fractionation in this project as described previously (Aryal et al., 2014, 2017; McBride et al., 2017, 2019). The collected tissue was disrupted by a Polytron homogenizer (Kinematica, New York, NY, USA) under 1 mL of ice-cold microsome isolation buffer solution (50 mM Hepes/KOH [pH 7.5], 250 mM sorbitol, 50 mM KOAc, 2 mM $Mg(OAc)_2$, 1 mM EDTA, 1 mM EGTA, 1 mM dithiothreitol [DTT], 2 mM PMSF and 1% [v/v] protein inhibitor cocktail [160 mg/mL benzamidine–HCl, 100 mg/mL leupeptin, 12 mg/mL phenanthroline, 0.1 mg/mL pepstatin A, and 0.1 mg/mL aprotinin]). To remove debris, the homogenate was centrifuged at 1,000 *g* using a Beckman Avanti 30 (Beckman, Palo Alto, CA, USA) for 10 min at 4°C. A soluble fraction was obtained by ultracentrifugation at 200,000 *g* for 20 min at 4°C on a Beckman Optima Ultracentrifuge (Beckman). Four sets of samples were prepared from independent pools of tissues and used as two biological replicates for the SEC and IEX separations.

SEC

From the soluble sample, endogenous proteins were fractionated by an ÄKTA FPLC system (Amersham Biosciences AB, Uppsala, Sweden) as described previously (Aryal et al., 2014, 2017; McBride et al., 2019). Approximately 1 mg of total cytosolic proteins was injected onto a Superdex® 200 10/300 GL column (GE Healthcare AB, Uppsala, Sweden). The SEC elution was performed with the mobile phase (50 mM Hepes/KOH [pH 7.5], 100 mM NaCl, 10 mM $MgCl_2$ 5% glycerol, and 1 mM DTT) at a flow rate of 0.65 mL/min. The elution chromatogram was monitored by measuring absorption at 280 nm. The column was calibrated using a Gel Filtration Markers Kit (MWGF1000; Sigma-Aldrich), determining a mass range from 669 to 29 kDa. SEC fractions of 500 µL were collected, and proteins were precipitated by cold acetone for LC–MS/MS analysis.

IEX

One milligram of the cytosolic proteins was fractionated on a PolyCATWAX 204CTWX0510 column (200 × 4.6 mm id, 5 µm, 1,000 A; PolyLC Inc., Columbia, MD, USA) using an UltiMate 3000 Standard HPLC System (Thermo Fisher Scientific Inc., Sunnyvale, CA, USA) as described previously (McBride et al., 2019). IEX separation was performed over a 100-min linear gradient elution program (from 0.0 to 1.5 M NaCl) at a flow rate of 1.0 mL/min. The absorbances of 214 and 280 nm were set to monitor protein elution. One hundred and seven sample fractions were collected every 22 s (~367 mL) between 3 min and 40 min. Each fraction was subjected to protein precipitation via a cold acetone method.

Sample preparation for LC–MS/MS analysis

Fractionated protein samples were digested for LC–MS/MS analysis using trypsin as described previously (McBride et al., 2017, 2019). Protein pellets were dissolved and denatured in 8 M urea for 1 h at room temperature, reduced in 10 mM DTT for 45 min at 60°C, and alkylated with 20 mM iodoacetamide for 45 min at room temperature in the dark. The urea concentration in the peptide solution was brought to 1.5 M for trypsin digestion by adding ammonium bicarbonate. The digested peptides were purified using Pierce® C18 Spin Columns (Thermo Fisher Scientific Inc., Rockford, IL, USA), and all samples were adjusted to an equal volume. Peptide concentrations were measured by bicinchoninic acid assay following the manufacturer's protocol (Thermo Fisher Scientific Inc.). The most concentrated sample had a peptide concentration of 0.2 µg/µL, and 5 µL of each sample was injected onto the LC–MS/MS system.

LC–MS/MS data acquisition

LC–MS/MS analysis was carried out as described previously (McBride et al., 2017). In brief, a Q-Exactive HF Hybrid Quadrupole-Orbitrap mass spectrometer in conjunction with reverse-phase HPLC–ESI–MS/MS using a Dionex UltiMate 3000 RSLC nano System (Thermo Fisher Scientific Inc.) was used. Peptides were resolved over a 125-min gradient at a flow rate of 300 nL/min. An MS survey scan was obtained from 350 to 1,600 mass/charge ratio range. MS/MS spectra were acquired by selecting the 20 most abundant precursor ions for sequencing with high-energy collisional dissociation normalized collision energy of 27%. A 15-s dynamic exclusion window was applied to reduce the number of times the same ion was sequenced.

Peptide identification and quantification

MaxQuant version 1.6.14.0 was used for relative protein abundance quantification and protein identification (Cox et al., 2014). The search was conducted as described (McBride et al., 2019). Raw files of total cytosolic, SEC, and IEX fractions were searched on MaxQuant together against the rice proteome Osativa_323_v7.0.protein.fa (Ouyang et al., 2007). The search parameters were as follows: cysteine carbamidomethylation was a fixed modification; oxidation on methionine and acetylation on protein N-terminus were variable modifications; up to two missed trypsin cleavages were accepted; 1% FDR at the protein and peptide level was chosen using a reverse decoy database; peptide abundance was calculated using the extracted ion current for both unique and razor peptides, and protein level signals were aggregated from peptide intensities using razor peptide signal allocation among protein groups; the match between runs function was set with a maximum matching time window of 0.7 min as default; all other parameters were set as default.

Reproducibility, Gaussian peak fitting, and R_{app} calculations

Reproducibility between two biological replicates was determined as described before (McBride et al., 2017, 2019). PCCs were estimated based on protein abundances in each fraction between duplicates and visualized by Data Analysis and Extension Tool. An optimized Gaussian fitting algorithm was applied to fit the chromatography resolution, and the Bayesian information criterion was utilized to prevent overfitting (McBride et al., 2017). The algorithm identified protein peaks when they had more than three nonzero fractions, with two being adjacent. Multiple reproducible peaks from a protein were split into multiple entries by labeling with a peak number on their locus IDs. The reproducible peaks were selected from the two replicates if they were present within two or four fraction shifts considering the increment rate between fractions in the SEC or IEX column, respectively. All nonreproducible peaks were eliminated from subsequent analyses. The fraction locations of the fitted peaks were used to determine the apparent mass (M_{app}) values of proteins using the SEC calibration curve obtained above. The protein multimerization state (R_{app}) was defined as the ratio of the M_{app} of a protein to the theoretical monomer mass (M_{mono}) of the protein. The R_{app} of $\geq$1.6 thresholds was applied to determine whether a protein was present in a complex.

Hierarchical clustering analysis

Hierarchical clustering was conducted as described previously (McBride et al., 2019). Briefly, a set of profiles that reflect the compositions of a protein complex was clustered based on protein elution similarity. This cluster analysis was carried out with IEX only, SEC only, and combined IEX and SEC datasets. The Euclidean distance was used as a metric for measuring similarity in profiles of a pair of proteins. A series of dendrograms over a wide range of cluster numbers was generated to determine an optimal cluster number value for the prediction.

Distance within clusters and purity and intactness of known protein complexes

To minimize false positives and negatives, the clustering results were evaluated based on distance within clusters and intactness and purity as described previously (McBride et al., 2019). The distance within a cluster indicates how much similar or dissimilar protein elutions are in the given cluster. A cluster center was first calculated as the average elution profiles of all proteins in a cluster. The mean distance of each protein in the cluster from the cluster center was then calculated.

Members of known protein complexes co-migrate. This characteristic of known complexes was applied to the determination of the final cluster number to predict protein complexes. Rice orthologs were searched against the CORUM database (Ruepp et al., 2009), which provides annotated protein complex information from mammalian organisms for the purity and intactness tests. Purity was calculated based on the ratio of the highest number of subunits for a known protein complex in a given group to the total number of proteins assigned to the group. Intactness was measured as the ratio of the number of identified subunits of a known complex assigned into one group to the total number of identified subunits in the known protein complex.

Validation and complex heterogeneity using external datasets and multiple peak proteins

To validate the clustering results, external datasets for Arabidopsis and rice CF-based protein complex predictions were downloaded from Supplemental Table S2 in McBride et al. (2019) and Supplemental Table S4 in McWhite et al. (2020), respectively. To facilitate comparisons among the studies, key parameters including the number of proteins, the number of clusters, and the definition of protein group and ortholog group used in the McBride et al. and McWhite et al. studies are summarized in Supplemental Data Set S4A. The Arabidopsis orthologs of rice proteins were searched using the Phytozome ortholog database (Goodstein et al., 2011) to map the McBride Arabidopsis complex prediction onto our rice dataset (column F in Supplemental Data Set S4B). For the McWhite prediction, their ortholog groups containing rice and Arabidopsis were similarly mapped onto our rice clustering dataset (Columns G–J in Supplemental Data Set S4C). For the rice data obtained in the current study, only protein groups in a cluster with two or more members but not classified as "false positive" or "degraded" were used for comparisons. When the same protein pair was present in the McBride and McWhite datasets, it was scored as "within a single cluster in both studies."

Another method to validate the clustering results using our rice data was to test for co-occurrence of multiple peak proteins in two distinct clusters (Supplemental Data Set S1B). Interpretations of the multiple IEX peaks were based on whether or not the proteins had either one or two distinct SEC peaks and the M_{app} values of the peaks.

Systematic classification of clustering results

Systematic classification was performed using a method similar to that described previously (McBride et al., 2019) based on M_{app}, the sum of M_{mono} of all proteins within one cluster (M_{calc}), the average of M_{app} within one cluster ($M_{app\text{-}avg}$), and R_{app} of all proteins in the given cluster. In the assumed scenario of 1:1 subunit stoichiometry, M_{calc} should be similar to the $M_{app\text{-}avg}$ of its members. These proteins were classified as "putative intact complex." In addition, clusters with likely high subunit stoichiometry contain reliable predictions. Single entry clusters were classified as "homomer" when the $R_{app} \geq 1.6$. Clusters with 2 or 3 protein members were defined as "possible homomer or heteromer/high subunit stoichiometry" if the protein had $M_{app} \geq (4 * M_{calc})$. Cluster members were classified as "subcomplex or high subunit stoichiometry" when $M_{app} \geq (4 * M_{calc})$ in the clusters with >3 protein members and as "partial complex/false negatives" when $M_{app} > 1.4 * M_{calc}$. When a protein within a cluster

had $M_{calc} > 1.4 * M_{app}$ and $R_{app} \geq 1.6$, the protein was defined as "putative complex clustered with false positives." If $R_{app} < 1.6$ and $M_{calc} > 1.4 * M_{app}$, the protein was flagged as "likely false positive: monomer." Single proteins were classified as "degraded" when the $R_{app} < 0.5$ and "monomeric" when $0.5 \leq$ the $R_{app} < 1.6$.

GO term analysis

The SEA tool in AgriGO version 2.0 was used for GO enrichment analysis (Tian et al., 2017). The enrichment was analyzed using Fisher's exact test at the 5% FDR level as Hochberg correction against all proteins in the MSU version 7.0 nonTE transcript ID (TIGR) background.

Data analysis

Gaussian fitting was applied using MATLAB_R2016a. Clustering analysis was performed using R version 3.5.1 (R Core Team, 2018) in RStudio version 1.1.463 (RStudio Team, 2018).

Accession numbers

The MS raw files have been deposited into the ProteomeXchange Consortium via PRIDE under accession code PXD022357. The mass spectra are available at the Protein Prospector with search key uij64faovq. The Gaussian fitting code and the clustering analysis code described in McBride et al. (2017) are available at (https://github.com/dlchenstat/Gaussian-fitting) and at (https://github.com/dlchenstat/ProteinComplexPredict), respectively.

Supplemental data

The following materials are available in the online version of this article.

Supplemental Figure S1. GO term enrichment analysis showing overrepresented terms in the sink-type aleurone–subaleurone layers.

Supplemental Figure S2. Extrinsic tests evaluating the resolving power for independent and combined protein profile datasets as a function of ascending cluster numbers.

Supplemental Data Set S1. Clustering results and protein complex prediction classifications.

Supplemental Data Set S2. Raw profiles of peptides and proteins.

Supplemental Data Set S3. Lists of reproducible protein peaks in SEC and IEX.

Supplemental Data Set S4. Comparisons of clustering results among plant CF–MS predictions.

Supplemental File S1. Elution profiles of clustered proteins and dendrograms of clusters.

Supplemental File S2. Clustering heatmap of the rice aleurone–subaleurone proteome.

Supplemental File S3. Supplemental text discussing the predicted novel and known protein complexes.

Funding

This work was supported by the National Science Foundation (NSF) Plant Genome Research Project 1444610 to T.W.O. and D.B.S. and 1951819 to D.B.S.. Y.L. was mainly supported by the above NSF funds and partially by a Bilsland Graduate Dissertation Fellowship from the Graduate School at Purdue University and a Center for Plant Biology (CPB) Graduate Research Award from the CPB at Purdue University.

Acknowledgments

We thank the Purdue Proteomics Facility and Dr. Uma Aryal for running the samples.

Conflict of interest statement. Authors declare no competing interests.

References

Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell **92**: 291–294

Andre C, Froehlich JE, Moll MR, Benning C (2007) A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in *Arabidopsis*. Plant Cell **19**: 2006

Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an *Arabidopsis* interactome map. Science **333**: 601

Aryal UK, McBride Z, Chen D, Xie J, Szymanski DB (2017) Analysis of protein complexes in *Arabidopsis* leaves using size exclusion chromatography and label-free protein correlation profiling. J Proteomics **166**: 8–18

Aryal UK, Xiong Y, McBride Z, Kihara D, Xie J, Hall MC, Szymanski DB (2014) A proteomic strategy for global analysis of plant protein complexes. Plant Cell **26**: 3867

Basu D, Le J, Zakharova T, Mallery EL, Szymanski DB (2008) A SPIKE1 signaling complex controls actin-dependent cell morphogenesis through the heteromeric WAVE and ARP2/3 complexes. Proc Natl Acad Sci **105**: 4044

Becraft PW, Yi G (2010) Regulation of aleurone development in cereal grains. J Exp Bot **62**: 1669–1675

Bethke PC, Fath A, Jones RL (2001) Regulation of viability and cell death by hormones in cereal aleurone. J Plant Physiol **158**: 429–438

Burd CG, Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. Science **265**: 615

Cai Y, Zhang W, Jin J, Yang X, You X, Yan H, Wang L, Chen J, Xu J, Chen W, et al. (2018) OsPKpα1 encodes a plastidic pyruvate kinase that affects starch biosynthesis in the rice endosperm. J Integr Plant Biol **60**: 1097–1118

Choi SB, Wang C, Muench DG, Ozawa K, Franceschi VR, Wu Y, Okita TW (2000) Messenger RNA targeting of rice seed storage proteins to specific ER subdomains. Nature **407**: 765–767

Chou HL, Tian L, Kumamaru T, Hamada S, Okita TW (2017) Multifunctional RNA binding protein OsTudor-SN in storage protein mRNA transport and localization. Plant Physiol **175**: 1608

Christofk HR, Vander Heiden MG, Wu N, Asara JM, Cantley LC (2008) Pyruvate kinase M2 is a phosphotyrosine-binding protein. Nature **452**: 181–186

Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics **13**: 2513–2526

Crofts AJ, Washida H, Okita TW, Ogawa M, Kumamaru T, Satoh H (2004) Targeting of proteins to endoplasmic reticulum-derived compartments in plants. The importance of RNA localization. Plant Physiol **136**: 3414

da Fonseca Paula CA, He J, Morris Edward P (2012) Molecular model of the human 26S proteasome. Mol Cell **46**: 54–66

Doroshenk KA, Crofts AJ, Morris RT, Wyrick JJ, Okita TW (2009) Proteomic analysis of cytoskeleton-associated RNA binding proteins in developing rice seed. J Proteome Res **8**: 4641–4653

Drew K, Müller CL, Bonneau R, Marcotte EM (2017) Identifying direct contacts between protein complex subunits from their conditional dependence in proteomics datasets. PLoS Comput Biol **13**: e1005625–e1005625

FAO (2003) Chapter 3. Prospects for aggregate agriculture and major commodity groups *In* Bruinsma J, ed, World Agriculture: Towards 2015/2030 – An FAO Perspective, Food and Agriculture Organization (FAO), London, pp 444

Fath A, Bethke P, Lonsdale J, Meza-Romero R, Jones R (2000) Programmed cell death in cereal aleurone. *In* E Lam, H Fukuda, J Greenberg, eds, Programmed Cell Death in Higher Plants, Springer Netherlands, Dordrecht, the Netherlands, pp 11–22

Fu H, Reis N, Lee Y, Glickman MH, Vierstra RD (2001) Subunit interaction maps for the regulatory particle of the 26S proteasome and the COP9 signalosome. EMBO J **20**: 7096–7107

Garcion C, Lohmann A, Lamodière E, Catinot J, Buchala A, Doermann P, Métraux JP (2008) Characterization and biological function of the ISOCHORISMATE SYNTHASE2 gene of Arabidopsis. Plant Physiol **147**: 1279–1287

Gilbert M, Schulze WX (2019) Global identification of protein complexes within the membrane proteome of Arabidopsis roots using a SEC-MS approach. J Proteome Res **18**: 107–119

Gillies SA, Futardo A, Henry RJ (2012) Gene expression in the developing aleurone and starchy endosperm of wheat. Plant Biotechnol J **10**: 668–679

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. (2011) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40**: D1178–D1186

Gutierrez-Beltran E, Denisenko TV, Zhivotovsky B, Bozhkov PV (2016) Tudor staphylococcal nuclease: biochemistry and functions. Cell Death Differ **23**: 1739–1748

Hamada S, Ishiyama K, Sakulsingharoj C, Choi SB, Wu Y, Wang C, Singh S, Kawai N, Messing J, Okita TW (2003) Dual regulated RNA transport pathways to the cortical region in developing rice endosperm. Plant Cell **15**: 2265

Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, et al. (2012) A census of human soluble protein complexes. Cell **150**: 1068–1081

Jones AM, Xuan Y, Xu M, Wang RS, Ho CH, Lalonde S, You CH, Sardi MI, Parsa SA, Smith-Valle E, et al. (2014) Border control—A membrane-linked interactome of *Arabidopsis*. Science **344**: 711–716

Kaksonen M, Toret CP, Drubin DG (2005) A modular design for the clathrin- and actin-mediated endocytosis machinery. Cell **123**: 305–320

Kapelari B, Bech-Otschir D, Hegerl R, Schade R, Dumdey R, Dubiel W (2000) Electron microscopy and subunit-subunit interaction studies reveal a first architecture of COP9 signalosome. J Mol Biol **300**: 1169–1178

Krishnan S, Dayanandan P (2003) Structural and histochemical studies on grain-filling in the caryopsis of rice (*Oryza sativa* L.). J Biosci **28**: 455–469

Kristensen AR, Gsponer J, Foster LJ (2012) A high-throughput approach for measuring temporal changes in the interactome. Nat Methods **9**: 907–909

Lander GC, Estrin E, Matyskiela ME, Bashore C, Nogales E, Martin A (2012) Complete subunit architecture of the proteasome regulatory particle. Nature **482**: 186–191

Larsen TM, Benning MM, Rayment I, Reed GH (1998) Structure of the Bis(Mg2+)−ATP−oxalate complex of the rabbit muscle pyruvate kinase at 2.1 Å resolution: ATP binding over a barrel. Biochemistry **37**: 6247–6255

LeClere S, Tellez R, Rampey RA, Matsuda SPT, Bartel B (2002) Characterization of a family of IAA-amino acid conjugate hydrolases from Arabidopsis. J Biol Chem **277**: 20446–20452

Lee Y, Szymanski DB (2021) Multimerization variants as potential drivers of neofunctionalization. Sci Adv **7**: eabf0984

Li L, Lavell A, Meng X, Berkowitz O, Selinski J, van de Meene A, Carrie C, Benning C, Whelan J, De Clercq I, et al. (2019) *Arabidopsis* DGD1 SUPPRESSOR1 is a subunit of the mitochondrial contact site and cristae organizing system and affects mitochondrial biogenesis. Plant Cell **31**: 1856–1878

Lin W, Zhang H, Huang D, Schenke D, Cai D, Wu B, Miao Y (2020) Dual-localized WHIRLY1 affects salicylic acid biosynthesis via coordination of ISOCHORISMATE SYNTHASE1, PHENYLALANINE AMMONIA LYASE1, and S-ADENOSYL-L-METHIONINE-DEPENDENT METHYLTRANSFERASE1 Plant Physiol **184**: 1884–1899

Llorca O, Betti M, González JM, Valencia A, Márquez AJ, Valpuesta JM (2006) The three-dimensional structure of an eukaryotic glutamine synthetase: functional implications of its oligomeric structure. J Struct Biol **156**: 469–479

Maeda H, Dudareva N (2012) The shikimate pathway and aromatic amino acid biosynthesis in plants. Ann Rev Plant Biol **63**: 73–105

Marsh JA, Teichmann SA (2015) Structure, dynamics, assembly, and evolution of protein complexes. Ann Rev Biochem **84**: 551–575

Mayberry LK, Allen ML, Nitka KR, Campbell L, Murphy PA, Browning KS (2011) Plant cap-binding complexes eukaryotic initiation factors eIF4F and eIFISO4F: MOLECULAR SPECIFICITY OF SUBUNIT BINDING. J Biol Chem **286**: 42566–42574

McBride Z, Chen D, Reick C, Xie J, Szymanski DB (2017) Global analysis of membrane-associated protein oligomerization using protein correlation profiling. Mol Cellular Proteomics **16**: 1972–1989

McBride Z, Chen D, Lee Y, Aryal UK, Xie J, Szymanski DB (2019) A label-free mass spectrometry method to predict endogenous protein complex composition. Mol Cell Proteomics **18**: 1588

McWhite CD, Papoulas O, Drew K, Cox RM, June V, Dong OX, Kwon T, Wan C, Salmi ML, Roux SJ, et al. (2020) A pan-plant protein complex map reveals deep conservation and novel assemblies. Cell **181**: 460–474.e414

Morris RT, Doroshenk KA, Crofts AJ, Lewis N, Okita TW, Wyrick JJ (2011) RiceRBP: A database of experimentally identified RNA-binding proteins in *Oryza sativa* L. Plant Sci **180**: 204–211

Mousdale DM, Coggins JR (1986) Detection and subcellular localization of a higher plant chorismate synthase. FEBS Lett **205**: 328–332

Negm FB, Cornel FA, Plaxton WC (1995) Suborganellar localization and molecular characterization of nonproteolytic degraded leukoplast pyruvate kinase from developing castor oil seeds. Plant Physiol **109**: 1461.

Okita TW, Li X, Roberts MW (1994) Targeting of mRNAs to domains of the endoplasmic reticulum. Trends Cell Biol **4**: 91–96

Olsen OA (2004) Nuclear endosperm development in cereals and *Arabidopsis thaliana*. Plant Cell **16**: S214

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L et al. (2007) The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res **35**: D883–D887

Core Team R (2018) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria

Rattray DG, Foster LJ (2019) Dynamics of protein complex components. Curr Opin Chem Biol **48**: 81–85

Team RStudio (2018) RStudio: Integrated Development Environment for R, RStudio, Inc., Boston, MA

Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW (2009) CORUM:

the comprehensive resource of mammalian protein complexes—2009. Nucleic Acids Res **38**: D497–D501

Salas D, Stacey GR, Akinlaja M, Foster LJ (2020) Next-generation Interactomics: considerations for the use of co-elution to measure protein interaction networks. Mol Cell Proteomics **19**: 1

Sami-Subbu R, Choi SB, Wu Y, Wang C, Okita TW (2001) Identification of a cytoskeleton-associated 120 kDa RNA-binding protein in developing rice seeds. Plant Mol Biol **46**: 79–88

Seaman MNJ, Michael McCaffery J, Emr SD (1998) A membrane coat complex essential for endosome-to-golgi retrograde transport in yeast. J Cell Biol **142**: 665–681

Shan H (2018) Characterization of P1 leader proteases of the Potyviridae family and identification of the host factors involved in their proteolytic activity during viral infection. Departamento de Biología Molecular, Universidad Autónoma de Madrid, Madrid, Spain

Sharon M, Taverner T, Ambroggio XI, Deshaies RJ, Robinson CV (2006) Structural organization of the 19S proteasome lid: insights from MS of intact complexes. PLoS Biol **4**: e267

Strawn MA, Marr SK, Inoue K, Inada N, Zubieta C, Wildermuth MC (2007) Arabidopsis isochorismate synthase functional in pathogen-induced salicylate biosynthesis exhibits properties consistent with a role in diverse stress responses. J Biol Chem **282**: 5919–5933

Tian L, Chou HL, Fukuda M, Kumamaru T, Okita TW (2020) mRNA localization in plant cells. Plant Physiol **182**: 97–109

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res **45**: W122–W129

Tzin V, Galili G (2010) The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. Arabidopsis Book **8**: e0132

Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, De Clercq A, Bonnet E, Laukens K, Remmerie N, et al. (2007) A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*. Mol Cell Proteomics **6**: 1226–1238

Wan C, Borgeson B, Phanse S, Tu F, Drew K, Clark G, Xiong X, Kagan O, Kwan J, Bezginov A, et al. (2015) Panorama of ancient metazoan macromolecular complexes. Nature **525**: 339–344

Washida H, Sugino A, Doroshenk KA, Satoh-Cruz M, Nagamine A, Katsube-Tanaka T, Ogawa M, Kumamaru T, Satoh H, Okita TW (2012) RNA targeting to a specific ER sub-domain is required for efficient transport and packaging of α-globulins to the protein storage vacuole in developing rice endosperm. Plant J **70**: 471–479

Weng JK, Li Y, Mo H, Chapple C (2012) Assembly of an evolutionarily new pathway for α-pyrone biosynthesis in *Arabidopsis*. Science **337**: 960–964

Widemann E, Miesch L, Lugan R, Holder E, Heinrich C, Aubert Y, Miesch M, Pinot F, Heitz T (2013) The amidohydrolases IAR3 and ILL6 contribute to jasmonoyl-isoleucine hormone turnover and generate 12-hydroxyjasmonic acid upon wounding in Arabidopsis leaves. J Biol Chem **288**: 31701–31714

Wodak SJ, Pu S, Vlasblom J, Seéraphin B (2009) Challenges and rewards of interaction proteomics. Mol Cell Proteomics **8**: 3–18

Wu X, Liu J, Li D, Liu CM (2016) Rice caryopsis development II: dynamic changes in the endosperm. J Integr Plant Biol **58**: 786–798

Yanagisawa M, Alonso JM, Szymanski DB (2018) Microtubule-dependent confinement of a cell signaling and actin polymerization control module regulates polarized cell growth. Curr Biol **28**: 2459–2466.e2454

Yang Y, Chou HL, Crofts AJ, Zhang L, Tian L, Washida H, Fukuda M, Kumamaru T, Oviedo OJ, Starkenburg SR, et al. (2018) Selective sets of mRNAs localize to extracellular paramural bodies in a rice glup6 mutant. J Exp Bot **69**: 5045–5058

Young TE, Gallie DR (2000) Programmed cell death during endosperm development. Plant Mol Biol **44**: 283–301

Zhang T, Poudel AN, Jewell JB, Kitaoka N, Staswick P, Matsuura H, Koo AJ (2016) Hormone crosstalk in wound stress response: wound-inducible amidohydrolases can simultaneously regulate jasmonate and auxin homeostasis in *Arabidopsis thaliana*. J Exp Bot **67**: 2107–2120