Computational and Structural Biotechnology Journal 19 (2021) 5117–5125

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

PROMISed: A novel web-based tool to facilitate analysis and visualization of the molecular interaction networks from co-fractionation mass spectrometry (CF-MS) experiments

Dennis Schlossarek [a], Marcin Luzarowski [a,*], Ewelina Sokołowska [a], Michał Górka [a], Lothar Willmitzer [a], Aleksandra Skirycz [a,b,*]

[a] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany
[b] Boyce Thompson Institute, 533 Tower Rd., Ithaca, NY 14853, United States

ARTICLE INFO

Article history:
Received 17 May 2021
Received in revised form 27 August 2021
Accepted 27 August 2021
Available online 01 September 2021

Keywords:
Co-fractionation mass spectrometry
CF-MS
Protein-protein complexes
Proteinmetabolite complexes
User-friendly app

ABSTRACT

Co-fractionation mass spectrometry (CF-MS)-based approaches enable cell-wide identification of protein-protein and protein-metabolite complexes present in the cellular lysate. CF-MS combines biochemical separation of molecular complexes with an untargeted mass-spectrometry-based proteomics and/or metabolomics analysis of the obtained fractions, and is used to delineate putative interactors. CF-MS data are a treasure trove for biological discovery. To facilitate analysis and visualization of original or publically available CF-MS datasets, we designed PROMISed, a user-friendly tool available online via https://myshiny.mpimp-golm.mpg.de/PDP1/ or as a repository via https://github.com/DennisSchlossarek/PROMISed. Specifically, starting with raw fractionation profiles, PROMISed (i) contains activities for data pre-processing and normalization, (ii) deconvolutes complex fractionation profiles into single, distinct peaks, (iii) identifies co-eluting protein–protein or protein–metabolite pairs using user-defined correlation methods, and (iv) performs co-fractionation network analysis. Given multiple CF-MS datasets, for instance representing different environmental condition, PROMISed allows to select for proteins and metabolites that differ in their elution profile, which may indicate change in the interaction status. But it also enables the identification of protein–protein and protein–metabolite pairs that co-elute together across multiple datasets. PROMISed enables users to (i) easily adjust parameters at each step of the analysis, (ii) download partial and final results, and (iii) select among different data-visualization options. PROMISed renders CF-MS data accessible to a broad scientific audience, allowing users with no computational or statistical background to look for novel protein–protein and protein–metabolite complexes for further experimental validation.

1. Background and summary

Comprehensive identification of protein–protein interactions (PPIs) is crucial for understanding the intricate mechanics behind all biological processes. In the past, a plethora of PPIs have been unraveled by large-scale studies using affinity purification coupled with mass spectrometry (AP-MS) and yeast-two-hybrid (Y2H) screens. Recent progress in mass-spectrometry proteomics has seen an advent of a complementary high-throughput method named co-fractionation mass spectrometry (CF-MS). In contrast to AP-MS and Y2H, CF-MS obviates the need for a protein bait and enables proteome-wide characterization of PPIs in a single experiment. CF-MS is based on separation of protein complexes, utilizing different biochemical techniques such as size-exclusion chromatography (SEC) [16,24], ion exchange (IEX) chromatography [13], blue native gels [11] or density-gradient centrifugation [7], followed by mass-spectrometry-based proteomics analysis of the obtained fractions. Similarity between separation profiles—referred to as co-fractionation—is used to delineate putative interactors. We have recently extended the CF-MS workflow to the

* Corresponding authors Boyce Thompson Institute, 533 Tower Rd., Ithaca, NY 14853, United States (A. Skirycz). Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany (M. Luzarowski).

E-mail addresses: schlossarek@mpimp-golm.mpg.de (D. Schlossarek), Luzarowski@mpimp-golm.mpg.de (M. Luzarowski), ESokolowska@mpimp-golm.mpg.de (E. Sokołowska), Willmitzer@mpimp-golm.mpg.de (L. Willmitzer), as4258@cornell.edu (A. Skirycz).

https://doi.org/10.1016/j.csbj.2021.08.042

analysis of protein–metabolite interactions (PMIs), dubbing our approach PROMIS (PROtein–Metabolite Interactions using Size separation). PROMIS combines SEC-based separation of protein–protein and protein–metabolite complexes followed by both proteomics and metabolomics analysis of the collected fractions. PROMIS builds upon the observation that metabolites remain in protein complexes during mild cell lysis and biochemical fractionation. Analogously to PPIs [13,24], PMIs are delineated by correlating the fractionation profiles of a protein–metabolite pair [17,21–23]. Table 1 list examples of CF-MS studies.

While a CF-MS protocol is not experimentally challenging, data analysis requires computational expertise that is not always present in an experimental group. Analysis of a CF-MS dataset entails normalization, the selection of single peaks from complex fractionation profiles, identification of co-fractionating molecule pairs, and finally charting of the interaction network. Here, we describe PROMISed (**PROMIS E**asy **D**ata analysis), a novel web tool designed to analyze, integrate, visualize and mine data obtained from CF-MS-based studies. Starting with datasets containing absolute or relative quantification of fractionated molecules (proteins, metabolites, nucleic acids), PROMISed allows users to i) perform preprocessing steps, including normalization, smoothing and replicate pooling, ii) split complex fractionation profiles into single peaks, iii) integrate data to identify co-fractionating molecules, e.g. protein–metabolite or protein–protein pairs, and iv) create and describe co-fractionation networks. Additionally, we implemented a statistical workflow which, given multiple datasets, looks for differential fractionation profiles indicative of a novel interaction, e.g. associated with a particular developmental stage or environmental condition. It is based on determining significant differences in the Manhattan distances calculated between fractionation profiles within the replicates and between experimental conditions. The main challenge with CF-MS is to differentiate true complexes from those merely having similar fractionation properties. For instance, in a single PROMIS experiment, every metabolite co-fractionates with several hundred proteins, of which possibly only one is a true binder. PROMISed enables users to identify pairs of molecules that co-migrate across multiple datasets; and in that way narrows down the list of putative interactors. In summary, PROMISed allows users with no computational experience to mine their own or available CF-MS datasets for novel protein–protein and protein–metabolite interactions.

2. Software description and methods

2.1. General design and implementation

PROMISed is accessible via a frontend web interface built using the shiny R package [5], as the backend of PROMISed is written in the R environment. The web tool aims to generate co-fractionation networks from raw fractionation profiles. The PROMISed user interface provides individual tabs to guide the user through the different steps of data analysis. Each step provides options to customize data analysis parameters and a plotting area displaying (intermediate) results, such as normalized profiles or protein–metabolite co-fractionation networks. For convenience, two example data files containing metabolite and protein fractionation profiles obtained upon cell lysate fractionation using SEC are available in order to demonstrate the functionality and requirements of PROMISed.

Table 1
Published datasets derived from co-fractionation-based methods. CN-PAGE: clear native PAGE, IEF: isoelectric focusing, IEX: ion-exchange chromatography, SEC: size-exclusion chromatography, SDG: sucrose density gradient. *1) the nine model species include: *Caenorhabditis elegans, Drosophila melanogaster, Mus musculus, Strongylocentrotus purpuratus, Homo sapiens, Xenopus laevis, Nematostella vectensis, Dictyostelium discoideum* and *Saccharomyces cerevisiae*; *2) plant species studied: *Arabidopsis thaliana, Brassica oleracea, Glycine* max*, Cannabis sativa, Solanum lycopersicum, Chenopodium quinoa, Zea mays, Oryza sativa* ssp. *japonica, Triticum aestivum, Cocos nucifera, Ceratopteris richardii, Selaginella moellendorf* and *Chlamydomonas reinhardtii.*

Method	Biological material	Identifications	Data available	Main focus	Reference
IEX, IEF, SDG	Human cell culture	3006 proteins, 622 putative complexes	Yes	A census of human soluble protein complexes	[13]
IEX	*E. coli*, *S. cerevisiae*		No	Target identification by chromatographic co-fractionation: monitoring of drug-protein interactions without immobilization or chemical derivatization	[4]
SEC	*A. thaliana*	713 cytosolic proteins	Yes	A proteomic strategy for global analysis of plant protein complexes	[2]
SEC	9 model species (*1)	13,386 protein orthologues	No	Panorama of ancient metazoan macromolecular complexes	[24]
SEC	*A. thaliana* cell suspension culture	5000 proteins, 140 metabolites	Yes	PROMIS, global analysis of protein–metabolite interactions using size separation in *Arabidopsis thaliana*	[23]
SEC	HEK293	2127 proteins, 462 complexes	Yes	Complex-centric proteome profiling by SEC-SWATH-MS	[14]
SEC	*A. thaliana*	3889 and 5563 proteins in two replicates, respectively	Yes	Global identification of protein complexes within the membrane proteome of Arabidopsis roots using a SEC-MS approach	[10]
SEC, IEX, IEF	13 plant species (*2)	141,910 unique proteins corresponding to 23,896 orthogroups	Yes	A pan-plant protein complex map reveals deep conservation and novel assemblies	[19]
SEC	Human cell culture, mouse embryonic stem cells	1012 (human) and 1165 (mESCs) RNA-associated proteins	No	Systematic discovery of endogenous human ribonucleoprotein complexes	[18]
CN-PAGE	*A. thaliana*	2338 (end of day) and 2469 (end of night) proteins	Yes	Protein complex identification and quantitative complexome by CN-PAGE	[11]
SDG	*A. thaliana*	216 ribosomal proteins	Yes	Separation and paired proteome profiling of plant chloroplast and cytoplasmic ribosomes	[7]
IEX, SEC, SDG	*Synechocystis* sp. PCC 6803	2062 proteins, 291 multiprotein complexes	Yes	Global landscape of native protein complexes in *Synechocystis sp.* Pcc 6803	[25]
SEC	*S. cerevisiae* (log-phase)	3982 proteins, 74 metabolites	Yes	Global mapping of protein–metabolite interactions in *Saccharomyces cerevisiae* reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity	[17]
SEC	*Chaetomium thermophilum*	3286 proteins, 257 metabolites	No	Coupling proteomics and metabolomics for the unsupervised identification of protein-metabolite interactions in *Chaetomium thermophilum*	[15]

As shown in Fig. 1, data analysis using PROMISed comprises i) a pre-processing step, reducing data noise and applying normalization and smoothing to ensure comparability between fractionation profiles obtained from independent CF-MS experiments, ii) statistical analysis based on the dis-elution score, identifying molecules whose fractionation profiles differ significantly between experimental conditions, iii) replicate pooling, creating a single fractionation profile for proteins or metabolites which replicates pass a chosen similarity threshold, iv) peak deconvolution, splitting complex fractionation profiles into single and distinct peaks for the downstream identification of co-fractionating molecules, v) data integration, creating a correlation matrix between the deconvoluted profiles which can be used to mine for co-fractionating partners or to vi) create co-fractionation networks.

PROMISed is available online via https://myshiny.mpimp-golm.mpg.de/PDP1/, located at the Max Planck Institute of Molecular Plant Physiology. Additionally, the source code as well as a docker image is available at https://github.com/DennisSchlossarek/PROMISed.

In the following we explain the main steps of the PROMISed workflow in detail, emphasizing the arguments adjustable by the user. A list of default settings is given in Table 2.

2.2. Input data requirements

The data files accepted by PROMISed are simple tables in a tab-delimited text format containing absolute or relative quantification of proteins and/or metabolites upon fractionation across collected fractions. Inexperienced users can easily generate the tables: the rows correspond to identified proteins or metabolites, each labeled with a unique name, and columns correspond to fractions obtained from different replicates and/or conditions. Column names should therefore include information about the name of the condition, the name and number of a replicate, and the number of a fraction, and should be constructed as:

Subsequent fractions of the same separation should be sorted in ascending order of fraction numbers. The tables may also contain columns with additional information, such as retention time and *m*/*z* for annotated metabolites or number of unique peptides used for protein identification. In such cases PROMISed asks the user to select a window of columns containing measured intensities in chromatographic fractions. Currently, additional information is ignored by PROMISed and will not be appended to the result tables. An example of CF-MS input data tables can be found in Supplementary Tables 1 and 2 or in the demo data available in PROMISed.

NameOfCondition_NameOfReplicate_#Replicate_ #Fraction

2.3. Pre-processing

Prior to statistical analysis and peak deconvolution, different data pre-processing steps can be applied, if necessary. These include data-noise reduction, normalization, and smoothing of fractionation profiles. In case of complex separation using size-exclusion chromatography, fractionation profiles of proteins and metabolites span several consecutive fractions, with the peak width depending on experimental setup such as SEC column pore size. Therefore, measured metabolite or protein abundance in one fraction, but not directly neighboring fractions, can be considered as data noise, which might interfere with downstream processes. *Remove Single Peaks* provides the option to remove this data noise by replacing single-fraction peaks with zeroes. An additional way of reducing data noise is *Profile Smoothing*, which is achieved here by applying a local 2nd-degree polynomial regression fitting controlled by the *Span Value* parameter. To reduce data noise even further, fractions with low relative intensity can be set to 0. This might be necessary if measured intensities are close to the detection limit of the instruments used, e.g. sensitivity of mass spectrometer.

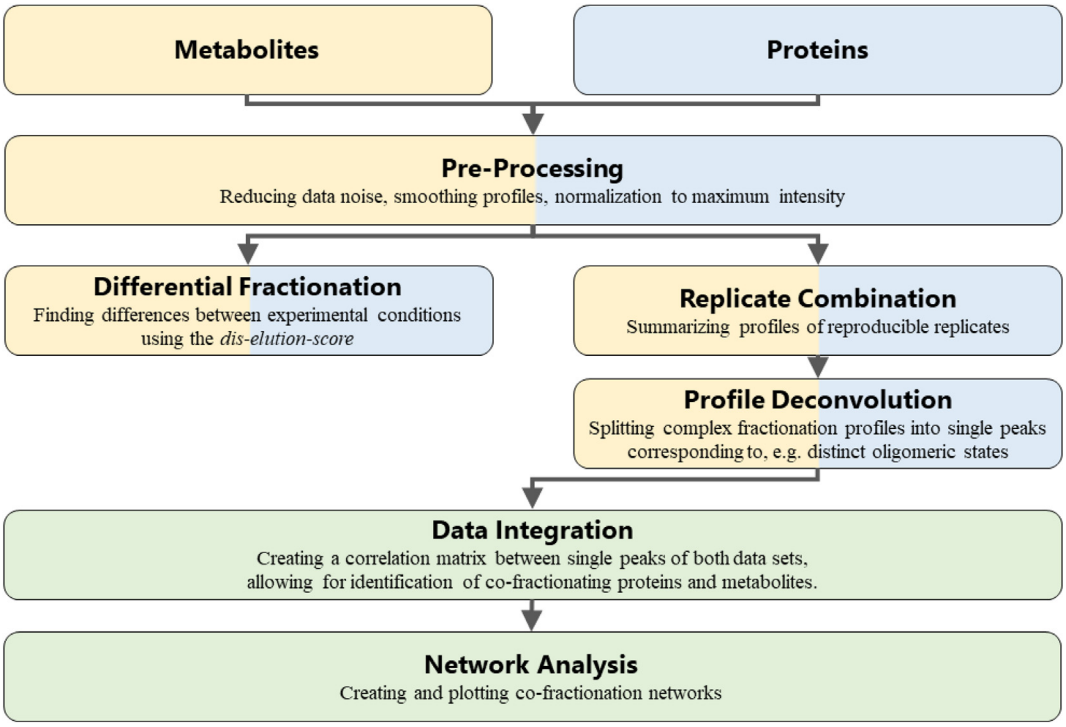


Fig. 1. Schematic workflow for the PROMISed web tool. PROMISed is designed to process and integrate two different CF-MS datasets, e.g. protein and metabolite profiles obtained by SEC. Separate metabolite and protein datasets are uploaded, then processed in parallel using the same methods and settings chosen by the user. After deconvolution, the datasets are integrated into a correlation matrix, which can be mined for co-fractionating pairs, and from which co-fractionation networks can be drawn.

Table 2
Default settings of the adjustable parameters in the PROMISed web tool.

Processing step	Parameter	Default
Pre-Processing	Remove Single Peaks	TRUE
	Normalize	TRUE
	Minimum Relative Intensity	0.10
	Profile Smoothing	TRUE
	Span Value	0.15
Replicate Combination	Correlation Method	Pearson
	Reproducibility Threshold	0.70
	Keep Single Replicates	FALSE
Deconvolution	Minimum Relative Intensity	0.20
	Minimum Intensity of Local Maxima	0.20
	Minimum Incline	0.80
Data Integration	Correlation Method	Pearson
	Correlation Threshold	0.70
Network Analysis	Filter Network	No Filter
	Node Colour	Cluster
	Layout	Force-directed
Differential Fractionation	p-Value Threshold	0.05

For data obtained from multiple, independently performed separations, normalization might be required to ensure comparability between fractionation profiles of different conditions or replicates. Normalization in PROMISed is performed separately for each fractionation profile, by normalizing every fraction to the fraction with maximum intensity within the fractionation profile. Normalization is required prior to statistical analysis of differentially eluting profiles, which uses Manhattan distances to compare peak intensities. All parameters and options included in the pre-processing step are listed below, and a schematic overview of their effects on fractionation profiles is shown in Fig. 2a.

Remove Single Peaks: replaces single values surrounded by zeroes with zero, eliminating data-noise "peaks" that only span one fraction.

Normalize: fractionation profiles are normalized to their maximum intensity.

Minimum Relative Intensity: sets a threshold for the minimum relative intensity (after normalization). Values below this threshold are replaced by 0.

Profile Smoothing: reduces data noise by using a local 2nd-degree polynomial regression fitting, controlled by *Span Value* as an additional parameter. Default *Span Value* is 0.15 for profiles of length 40. Sup. Fig. 1 gives an overview of minimal *Span Values* for profiles of different length.

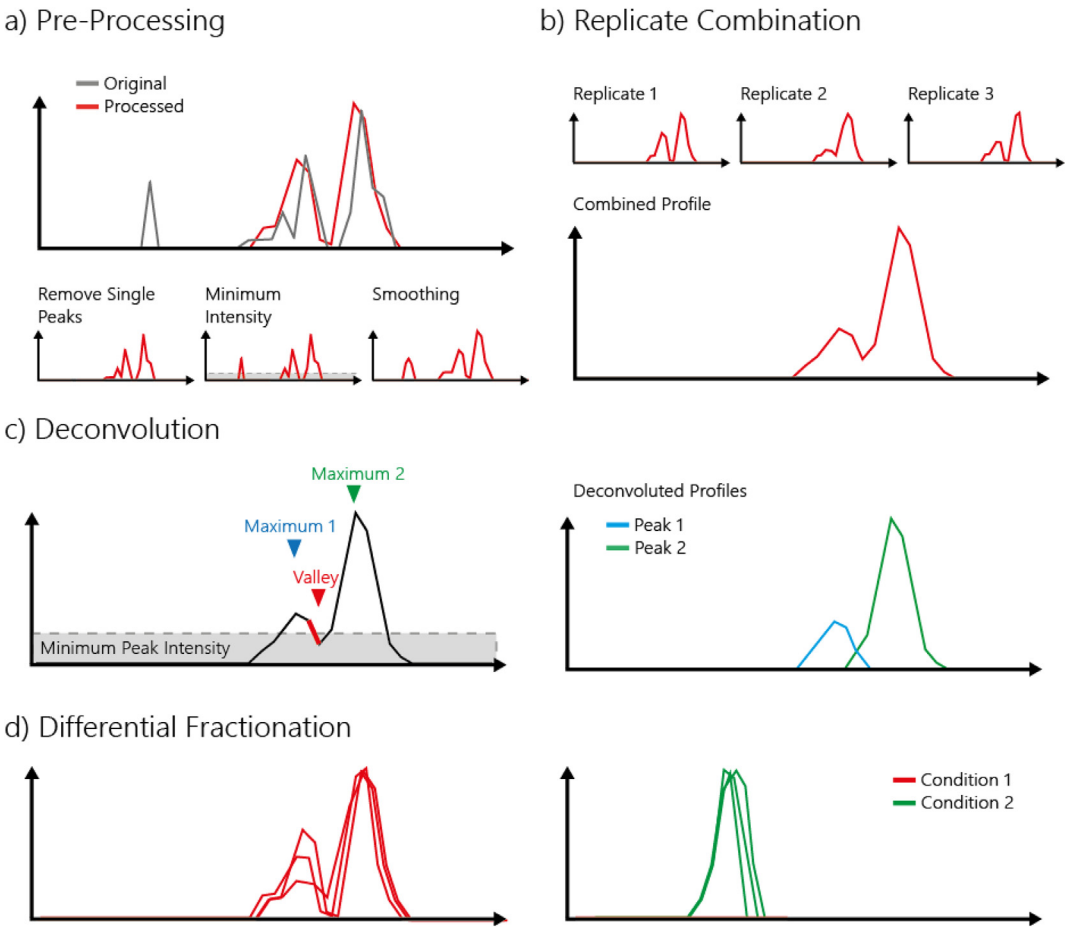


Fig. 2. Effect of processing options in PROMISed on fractionation profiles. a) Top: Resulting profile (red) after applying all pre-processing steps compared to the original profile (gray). Bottom: Effect of individual pre-processing options on the shape of the fractionation profile. b) Fractionation profiles of three replicates and the combined profile. c) Left: Identification of local maxima and valleys in the combined profile. Both local maxima are greater than the *Minimum Peak Intensity*. The valley is considered a cutting point, since the incline to the previous fraction (marked in red) exceeds the *Minimum Incline* threshold. The resulting peaks are shown on the right side. d) Example of a protein or metabolite profile fractionating differentially between two conditions (left and right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Span Value: controls the degree of smoothing by defining the number of neighboring fractions involved. Corresponds to proportion of total fractions and is given in a range of 0 to 1. Default is 0.15 for profiles spanning 40 fractions. In such examples, six neighboring fractions will be used for smoothing. Profiles differing in the number of fractions may require optimization of Span Values.

2.4. Replicate pooling

The pooling of replicates is achieved by summing up reproducible replicates of fractionation profiles (Fig. 2b). Reproducibility is calculated using either one of three correlation methods. An adjustable threshold gives the user control over the strictness of replicate combination. In addition, for datasets comprising two replicates, a single profile is treated as reproducible when the other replicate comprises an empty profile. Two of our previous studies exploiting PROMIS for system-wide detection of protein–metabolite complexes in *Arabidopsis thaliana* and *Saccharomyces cerevisiae* showed high reproducibility of fractionation (PCC > 0.9) between biological replicates [17,23].

Correlation Method: the user can choose between three methods of calculating correlation: *Pearson correlation*, *Kendall's tau* or *Spearman's rank correlation*.

Reproducibility Threshold: sets a threshold for the minimum correlation coefficient between replicates to sum up fractionation profiles.

Keep Single Replicates: allows to keep fractionation profiles of one replicate when the other replicates are uniformly 0. Used as default when uploaded data contains only one or two replicates.

2.5. Deconvolution

Since native proteins and metabolites can be involved in multiple protein–protein–metabolite complexes of different molecular weights, obtained CF-MS fractionation profiles are often complex and contain several local maxima. Profile deconvolution aims at splitting a complex fractionation profile into several profiles containing one peak each, corresponding to independent homomeric or heteromeric states of a protein or a protein partner of a metabolite. Profile deconvolution based on identifying local maxima was first used in studies aiming to identify protein–protein complexes using CN-PAGE [11] and later adapted for identification of protein–metabolite complexes using PROMIS [17]. In a nutshell, deconvolution is achieved by first identifying local maxima and subsequently finding the shape of the underlying fractionation peak (Fig. 2c).

Local maxima are found using the turnpoint function (pastecs R package [12] and filtered against the *Minimum Intensity of Local Maxima* parameter. Next, the shape of the fractionation peak is identified by "scanning" over the profile and making a decision at each fraction as follows: starting from a local maximum, each subsequent fraction is considered part of the underlying peak until its maximum normalized intensity is below *Minimum Relative Intensity* or until the fraction constitutes a major local valley, with an incline greater than *Minimum Incline* to one of the surrounding fractions. The fractions prior to the first local maximum, and fractions in-between the last fraction of an identified peak and the next maximum are assigned to the following true local maximum, as long as their intensity passes the *Minimum Relative Intensity* criteria. Profile deconvolution can be bypassed by checking the *No Deconvolution* box.

Minimum Relative Intensity: sets a threshold for the minimum relative. Values below this threshold are replaced by 0.

Minimum Intensity of Local Maxima: sets a threshold for the minimum relative intensity with which a local maximum is considered an independent peak.

Minimum Incline: minimum incline around a local minimum for cutting the peak into two.

2.6. Data integration

The integration of the two datasets is accomplished by calculating a correlation matrix between the deconvoluted profiles of both datasets. The method of correlation can be either *Pearson correlation*, *Kendall's tau* or *Spearman's rank correlation*. In our hands and for the purpose of screening for protein–metabolite complexes in *S. cerevisiae* using SEC, Pearson correlation and Kendall's tau served equally well for retrieving known protein–metabolite assemblies [17]. The resulting correlation matrix is filtered using the *Correlation Threshold* and can be downloaded as a text file. Proteins and metabolites passing the threshold are further considered as co-fractionating. Previous studies showed that a threshold of 0.7 (default) gives a good agreement between sensitivity and specificity for detection of known molecular complexes [17,23].

PROMISed offers multiple ways of mining the obtained results in a targeted and untargeted manner (see "Network analysis"). Focusing on a pre-selected molecule, in the "Data Integration" tab users have the possibility to identify co-fractionating partners of a chosen entry across all provided datasets ("Intersections of Conditions") or under a defined condition (e.g. for single species, developmental stage or growth conditions). Additionally, the deconvoluted fractionation profiles of a target from each dataset can be directly compared in one plot. Lastly, the protein or metabolite peaks with the highest correlation to the target can be plotted together with the target, and a subset of the correlation matrix is displayed in the user interface (Fig. 3).

Correlation Method: gives the user the option to choose between three methods of calculating correlation: *Pearson correlation*, *Kendall's tau* or *Spearman's rank correlation*.

Correlation Threshold: sets a threshold for the minimum correlation coefficient to consider metabolite and/or protein peaks as co-eluting.

2.7. Network analysis

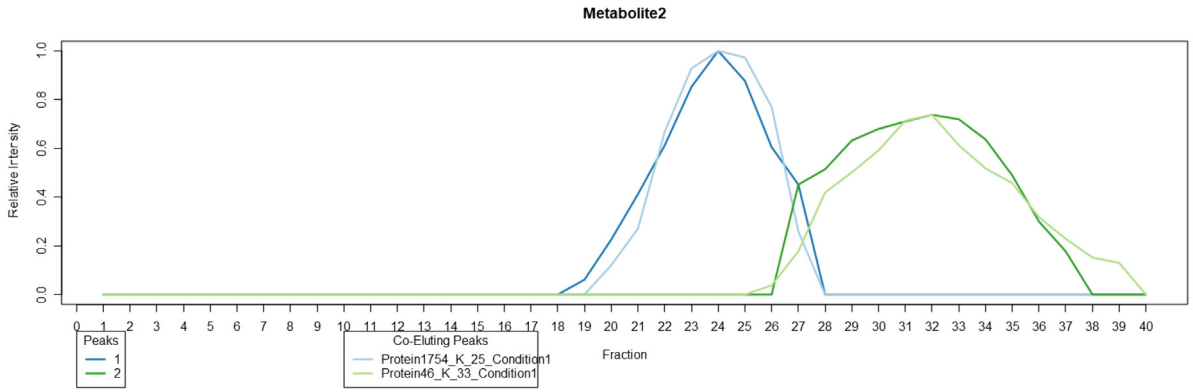
PROMISed constructs co-fractionation networks from the filtered correlation matrix using the igraph R package [6]. Networks are created as weighted correlation networks, depicting single protein or metabolite peaks as nodes, and the correlation coefficient between those peaks as edges. Networks can be restricted to selected entries instead of using the whole correlation matrix. To allow the user to recreate and analyze the obtained networks using different tools such as Cytoscape [20], the edgelists and nodelists can be readily downloaded.

The networks are displayed and made interactive using the visNetwork R package [1]. Nodes can be color-highlighted to depict either information about connectivity, calculated as k-coreness, or communities, identified using the Louvain method for community detection [3]. Additionally, the user can choose between three different options for the network's layout: network *components*, a *force-directed* layout, or *automatic* detection of the most appropriate layout. An example of network visualization is given in Fig. 4.

Filter Network: determines whether a network should be created using all available data (*No Filter*) or only around a selected protein or metabolite (*Selection 1* and *2*).

Node Color: changes node colors to highlight either communities (*Cluster*) (Fig. 4a) or node connectivity measured as k-Coreness (Fig. 4b). Default is a *Uniform* light blue.

Layout: controls the layout of the depicted network with two options: *Force-directed* uses the network layout algorithm by Fruchterman and Reingold [9], and *Circles* creates a circle of nodes for each network component.



Show 10 entries Search:

	Metabolite2_K_25_Condition1	Metabolite2_K_33_Condition1
Protein1754_K_25_Condition1	0.983	-0.209
Protein1238_K_24_Condition1	0.976	-0.211
Protein1825_K_26_Condition1	0.970	-0.218
Protein251_K_26_Condition1	0.968	-0.226
Protein694_K_25_Condition1	0.967	-0.224

Fig. 3. Top scored co-fractionating protein peaks of a selected metabolite. Upper panel: Deconvoluted fractionation profile of a selected metabolite (Metabolite2) in dark, and the deconvoluted protein peaks with the best scores (here: Pearson correlation) in light colors. The first Metabolite2 peak (blue) co-elutes with Protein1754, the second Metabolite2 peak (green) with Protein33. Lower panel: Table depicting correlation scores, here Pearson correlation, between the two peaks originating from the Metabolite2 fractionation profile and all protein peaks. The table is sorted to show the highest 5 scores for the first Metabolite2 peak, revealing Protein1754 as the best co-eluting protein peak with a PCC of 0.98. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

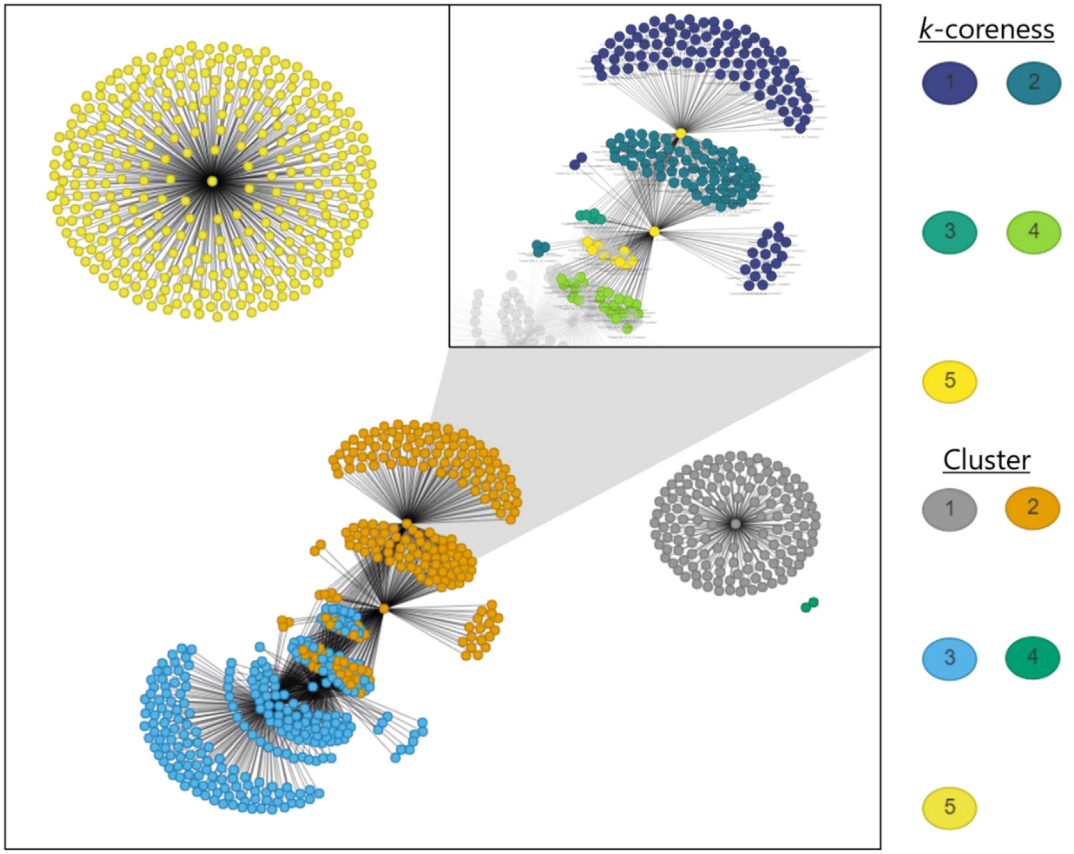


Fig. 4. Visualization of co-fractionation networks. The network is depicted using the Force-directed layout and Cluster color option. The inset shows Cluster 2 colored according to its k-coreness.

2.8. Differential fractionation

Describing differences in fractionation profiles (as shown in Fig. 2d) of the same protein or metabolite is a crucial, yet challenging step in identifying interactions responsive to experimental conditions. One attempt to meet this challenge was made by Mallam and colleagues, who developed a computational framework named DIF-FRAC score. The DIF-FRAC score compares a normalized Manhattan distance between fractionation profiles of treated and untreated proteins with those of known non-responsive proteins. This method proved to be powerful for the search of novel RNA-associated proteins in the presence and absence of RNA, where a true-negative list could be generated [18]. However, for most untargeted research questions this is not the case.

To circumvent this requirement, we developed the ***dis-elution score***, which relies on the statistical comparison of Manhattan distances between conditions with distances within conditions. In detail, we calculate the Manhattan distances of all combinations of replicates between conditions X and Y ($\|\vec{x} - \vec{y}\|_1$) and within the conditions ($\|\vec{x} - \vec{x}\|_1$ and$\|\vec{y} - \vec{y}\|_1$). The resulting vectors of Manhattan distances are then statistically compared using a one-way analysis of variance (ANOVA). A post-hoc Tukey test is then used to check whether $\|\vec{x} - \vec{y}\|_1$ is significantly larger than $\|\vec{x} - \vec{x}\|_1$ and $\|\vec{y} - \vec{y}\|_1$).

The "Differential Fractionation" tab allows the user to calculate the ***dis-elution score*** as a pairwise comparison between experimental conditions. The results can be downloaded in a table format, where FDR-corrected p-values of profiles passing the Tukey test are reported or viewed for each protein or metabolite in each pair of conditions depicted as a boxplot (see Metabolite 1–3 from example data). Please note that p-values shown on boxplots in the user interface are not FDR corrected and may differ from values in the results table.

p-Value Threshold: controls p-value threshold to filter dis-elution scores.

3. Promised captures predicted AHA2 interactors

To test the applicability of PROMISed to analyze previously published CF-MS experiments, we used the publicly available dataset from Gilbert&Schulze, 2018. The authors first investigated membrane protein–protein complexes isolated from the *Arabidopsis thaliana* roots and subsequently focused on one selected protein: H + -ATPase AHA2. Comparison of the 174 proteins co-eluting with AHA2 with the list of 32 proteins identified in the AHA2 pulldown experiment revealed an overlap of 13 proteins. We downloaded original protein elution profiles available as a supplement dataset to compare published analysis with the data analysis pipeline embedded in PROMIsed. The experiment comprises two biological replicas, which we analyzed separately, using the default settings, except for the Data Integration Correlation Threshold, which was set to 0.89 to replicate the author's high-confidence threshold. We then filtered the generated networks for AHA2 and obtained the network's edge-list. In other words, a list of co-fractionating proteins. In total, we identified 489 and 590 co-fractionating proteins in replicate one and replicate two, respectively, of which 161 were common. We then compared 161 identified interactors with the list of AHA2 protein partners reported in the original work (Fig. 5). The overlap contained 64 of the previously predicted AHA2 protein interaction partners, of which 19 were also retrieved in the pulldown experiment, which is six more than in the original analysis. The discrepancy originates from an additional AHA2 elution peak resulting from the deconvolution step.

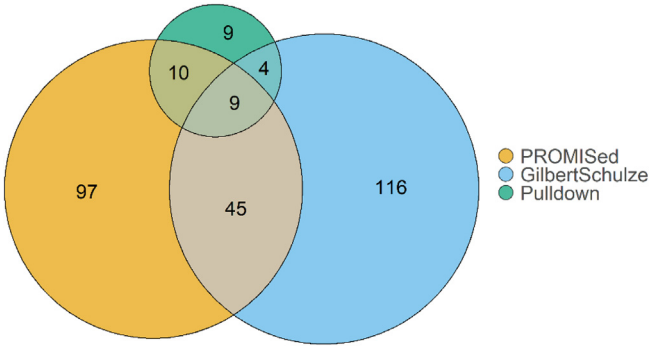


Fig. 5. PROMISed can capture previously predicted protein interactors of AHA2. Comparison of the number of proteins co-fractionating with AHA2 in the data from Gilbert&Schulze, 2018. The data was analysed using PROMISed default settings with the *Data Integration Correlation Threshold* of 0.89, to reproduce the original work. In total, PROMISed captured 161 interaction partners shared in both replicates, compared to the 174 in the original work (Gilbert&Schulze). Additionally, PROMISed captured 19 proteins that were also identified in a pull-down experiment performed in the original work.

4. Novelty and applications

Recent years have seen an advent of CF-MS-based approaches for the characterization of molecular complexes. Multiple datasets for both model and non-model organisms, spanning different developmental stages and environmental conditions have been published and are available to the scientific community. Moreover, interactive tools such as SECexplorer-cc (https://*sec*-explorer.shinyapps.io/hela_cellcycle/) [14], plant.Map (http://plants.proteincomplexes.org) [19] and YeastPMI (http://promis.mpimp-golm.mpg.de/yeastpmi) [17] provide an access to PPIs and PMIs for a handful of those published datasets. Here, we introduce a freely accessible and intuitive web tool designed for analysis, integration, visualization and mining of co-fractionation data, which does not require a computational or statistical background. PROMISed is not restricted by the identity of interactors (protein, metabolite etc.) or type of separation (SEC, IEX, native gel), and the only input it requires is raw fractionation profiles. PROMISed enables adjustment of multiple parameters at each analysis step, allowing tailoring of the process to a particular dataset or a biological question and making it more interactive and consequently more attractive to potential users. Moreover, at each data-processing step users can download and inspect the results file. When it comes to data mining, PROMISed can be used in several ways. In the simplest scenario, users can mine PROMISed-generated interaction networks (matrix) for the putative interactors of their protein or metabolite of interest. As already mentioned in the Introduction, a single CF-MS dataset may not be sufficient to differentiate true complexes from those merely having similar fractionation properties. A proven strategy to address this problem [19], implemented in PROMISed, is to look for repeating co-fractionation across multiple datasets, differing, for instance, in the separation method or source of the starting material. In addition to searching for repeating co-fractionation, another unique feature implemented in PROMISed allows users to look for proteins and metabolites that change their fractionation profile in multi-condition experiments. A difference in the fractionation profile can have different biological explanations, but it unequivocally attests to rewiring of the interactome. Overall, PROMISed is a unique tool designed for analysis, integration, visualization and mining of co-fractionation data, accessible to users without computational or statistical background.

5. Discussion and conclusion

The use of CF-MS-based approaches has led to numerous and significant biological insights. For example, by separating the soluble protein fraction of human cell culture lines using ion-exchange chromatography, authors found as many as 13,993 interactions, corresponding to 622 putative protein complexes; more than half of them (364) were previously unannotated [13]. Using a workflow dubbed DIF-FRAC, Mallam and colleagues identified 115 ribonucleoprotein complexes in human cell culture lines by comparing separation profiles of proteins obtained from lysates treated with RNAse to a non-treated control, inferring that 20% of human protein complexes contain an RNA component [18]. By expanding co-fractionation-based methods to protein–metabolite interactions, PROMIS has delineated putative protein interactors for over 140 metabolites in *A. thaliana* [23], and validated 225 previously predicted PMIs in yeast [17]. Following up on PROMIS results led to the identification of proteogenic dipeptides as novel regulators of metabolism [17,23].

While the number of methods using co-fractionation to study protein–protein and protein–metabolite interactions is growing, data-analysis pipelines and strategies greatly vary between research groups and are often highly specific for one experiment. For example, machine-learning approaches have been successfully used to create protein–protein interaction networks [13,19,24]. Additionally, a complex-centric experimental workflow using SEC-SWATH-MS has been developed [14], complemented by a machine-learning-based computational framework dubbed PCprophet [8]. However, machine-learning-based approaches require preselected data to train on, limiting transferability to other approaches and datasets.

In addition, neither approach has so far been applied to integrate the wide range of protein-binding metabolites. And finally, the complex-centric workflow, and especially any machine-learning approach, requires a certain level of bioinformatic skills, creating an entry barrier for many experimental researchers. We created a user-friendly application that does not require prior computational insight and that guides users through the processing steps, allows to adjust parameters of each activity to tailor various experimental setups, and helps to visualize co-fractionating pairs and interactome networks.

6. Code availability

PROMISed is available online via https://myshiny.mpimp-golm.mpg.de/PDP1/, located at the Max Planck Institute of Molecular Plant Physiology. Additionally, the source code as well as a docker image is available at https://github.com/DennisSchlossarek/PROMISed.

CRediT authorship contribution statement

Dennis Schlossarek: Conceptualization, Software, Methodology, Visualization, Writing – original draft, Formal analysis. **Marcin Luzarowski:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Ewelina Sokołowska:** Validation. **Michał Górka:** Methodology. **Lothar Willmitzer:** Resources, Project administration. **Aleksandra Skirycz:** Writing - review & editing, Project administration, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.08.042.

References

[1] Almende, B. V., Thieurmel, B., & Robert, T. (2019). visNetwork: Network Visualization using "vis.js" Library. R package version 2.0.9. Retrieved from: https://cran.r-project.org/package=visNetwork.

[2] U.K. Aryal Y.i. Xiong Z. McBride D. Kihara J. Xie M.C. Hall et al. A Proteomic strategy for global analysis of plant protein complexes 26 10 2014 2014 3867 3882 10.1105/tpc.114.127563.

[3] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008;2008 (10):1–12. https://doi.org/10.1088/1742-5468/2008/10/P10008.

[4] Chan, J. N. Y., Vuckovic, D., Sleno, L., Olsen, J. B., Pogoutse, O., Havugimana, P., … Emili, A. (2012). Target identification by chromatographic co-elution: Monitoring of drug-protein interactions without immobilization or chemical derivatization. Molecular and Cellular Proteomics, 11(7). https://doi.org/10.1074/mcp.M111.016642.

[5] Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2020). shiny: Web Application Framework for R. R package version 1.5.0. Retrieved from: https://cran.r-project.org/package=shiny.

[6] Csárdi G, Nepusz T. The igraph software package for complex network research. *InterJournal* 2006. https://doi.org/10.5281/zenodo.3630268.

[7] Firmino AAP, Gorka M, Graf A, Skirycz A, Martinez-Seidel F, Zander K, et al. Separation and paired proteome profiling of plant chloroplast and cytoplasmic ribosomes. Plants 2020;9(7):892. https://doi.org/10.3390/plants9070892.

[8] Fossati A, Li C, Uliana F, Wendt F, Frommelt F, Sykacek P, et al. PCprophet: a framework for protein complex prediction and differential analysis using proteomic data. Nat Methods 2021;18(5):520–7. https://doi.org/10.1038/s41592-021-01107-5.

[9] Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Software Pract Exp 1991;21(11):1129–64. https://doi.org/10.1002/spe.4380211102.

[10] Gilbert M, Schulze WX. Global identification of protein complexes within the membrane proteome of arabidopsis roots using a SEC-MS approach. J Proteome Res 2019;18(1):107–19. https://doi.org/10.1021/acs.jproteome.8b00382.

[11] Gorka M, Swart C, Siemiatkowska B, Martínez-Jaime S, Skirycz A, Streb S, et al. Protein complex identification and quantitative complexome by CN-PAGE. Sci Rep 2019;9(1):1–14. https://doi.org/10.1038/s41598-019-47829-7.

[12] Grosjean, P., & Ibanez, F. (2018). pastecs: Package for Analysis of Space-Time Ecological Series. R package version 1.3.21. Retrieved from: https://cran.r-project.org/package=pastecs.

[13] Havugimana P, Hart GT, Nepusz T, Yang H, Turinsky A, Li Z, et al. A census of human soluble protein complexes. Cell 2012;150(5):1068–81. https://doi.org/10.1016/j.cell.2012.08.011.

[14] Heusel M, Bludau I, Rosenberger G, Hafen R, Frank M, Banaei-Esfahani A, et al. Complex-centric proteome profiling by SEC - SWATH - MS. Mol Syst Biol 2019;15(1):1–22. , https://doi.org/10.15252/msb.20188438.

[15] Li Y, Kuhn M, Zukowska-Kasprzyk J, Hennrich ML, Kastritis PL, O'Reilly FJ, et al. Coupling proteomics and metabolomics for the unsupervised identification of protein – metabolite interactions in Chaetomium thermophilum. PLoS ONE 2021;16(7):e0254429. https://doi.org/10.1371/journal.pone.0254429.

[16] Liu X, Yang W-C, Gao Q, Regnier F. Toward chromatographic analysis of interacting protein networks. J Chromatogr A 2008;1178(1-2):24–32. https://doi.org/10.1016/j.chroma.2007.10.067.

[17] Luzarowski M, Vicente R, Kiselev A, Wagner M, Schlossarek D, Erban A, et al. Global mapping of protein–metabolite interactions in Saccharomyces cerevisiae reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity. Commun Biol 2021;4(1). https://doi.org/10.1038/s42003-021-01684-3.

[18] Mallam AL, Sae-Lee W, Schaub JM, Tu F, Battenhouse A, Jang YJ, et al. Systematic discovery of endogenous human ribonucleoprotein complexes. Cell Rep 2019;29(5):1351–1368.e5. https://doi.org/10.1016/j.celrep.2019.09.060.

[19] McWhite CD, Papoulas O, Drew K, Cox RM, June V, Dong OX, et al. A pan-plant protein complex map reveals deep conservation and novel assemblies. Cell 2020;181(2):460–474.e14.

[20] Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003. https://doi.org/10.1017/CBO9781107415324.004.

[21] Sokolowska EM, Schlossarek D, Luzarowski M, Skirycz A. PROMIS: global analysis of PROtein-metabolite interactions. Curr Protoc Plant Biol 2019;4(4):. https://doi.org/10.1002/cppb.20101e20101.

[22] Veyel, D., Kierszniowska, S., Kosmacz, M., Sokolowska, E. M., Michaelis, A., Luzarowski, M., … Skirycz, A. (2017). System-wide detection of protein-small molecule complexes suggests extensive metabolite regulation in plants. Scientific Reports, 7(July 2016), 1–8. https://doi.org/10.1038/srep42387.

[23] Veyel D, Sokolowska EM, Moreno JC, Kierszniowska S, Cichon J, Wojciechowska I, et al. PROMIS, global analysis of PROtein-metabolite

interactions using size separation in Arabidopsis thaliana. J Biol Chem 2018;293(32):12440–53. https://doi.org/10.1074/jbc.RA118.003351.
[24] Wan C, Borgeson B, Phanse S, Tu F, Drew K, Clark G, et al. Panorama of ancient metazoan macromolecular complexes. Nature 2015;525(7569):339–44. https://doi.org/10.1038/nature14877.
[25] Xu C, Wang B, Yang L, Zhongming Hu L, Yi L, Wang Y, et al. Global landscape of native protein complexes in Synechocystis sp. PCC 6803. Genom Proteom Bioinform 2021. https://doi.org/10.1016/j.gpb.2020.06.020.