### Title

I’m Walking into Spiderwebs: Making Sense of Protein—Protein Interaction Data




### Journal

Journal of Proteome Research, 23(8)

### ISSN

1535-3893



### Authors

Skawinski, Chase LS

Shah, Priya S



### Publication Date

2024-08-02

### DOI

10.1021/acs.jproteome.3c00892

Peer reviewed

Journal of proteome research



Review

# I'm Walking into Spiderwebs: Making Sense of Protein−Protein Interaction Data

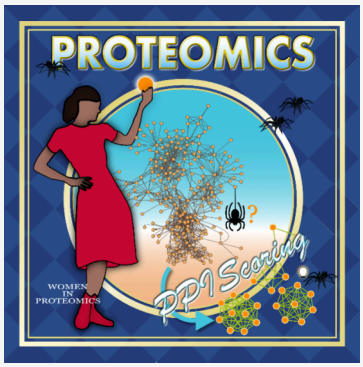
Chase L. S. Skawinski and Priya S. Shah*

Cite This: J. Proteome Res. 2024, 23, 2723−2732

Read Online

ACCESS | Metrics & More | Article Recommendations

**ABSTRACT:** Protein−protein interactions (PPIs) are at the heart of the molecular landscape permeating life. Proteomics studies can explore this protein interaction landscape using mass spectrometry (MS). Thanks to their high sensitivity, mass spectrometers can easily identify thousands of proteins within a single sample, but that same sensitivity generates tangled spiderwebs of data that hide biologically relevant findings. So, what does a researcher do when she finds herself walking into spiderwebs? In a field focused on discovery, MS data require rigor in their analysis, experimental validation, or a combination of both. In this Review, we provide a brief primer on MS-based experimental methods to identify PPIs. We discuss approaches to analyze the resulting data and remove the proteomic background. We consider the advantages between comprehensive and targeted studies. We also discuss how scoring might be improved through AI-based protein structure information. Women have been essential to the development of proteomics, so we will specifically highlight work by women that has made this field thrive in recent years.




**KEYWORDS:** *protein−protein interactions, mass spectrometry, proteomic scoring*

## INTRODUCTION

Coordinated interactions between proteins mediate a myriad of cellular functions throughout life. These protein−protein interactions (PPIs) can occur directly between two proteins, but many comprise multiple proteins that work together as a complex to execute specific functions. Protein complexes form as quaternary structures from the non-covalent interactions of multiple proteins via hydrogen bonds and van der Waals forces. Studying PPIs can reveal mechanisms of cellular homeostasis,[1,2] dynamic cellular processes (e.g., cell signaling, division, or differentiation),[3] and disease.[4,5] Here, we review how PPIs can be identified using mass spectrometry (MS)-based approaches, the caveats associated with analyzing such data rigorously, and the opportunities for future innovation in the field. Instead of providing a comprehensive review of such a broad field for this special issue focused on women in proteomics, we highlight work by women when possible (see notes).

## MS-BASED APPROACHES IN PPI IDENTIFICATION

MS identifies proteins using unique mass spectra, or mass-to-charge ratio fingerprints. This can be done using a "bottom-up" or "top-down" method. Bottom-up proteomics requires enzymatic digestion of a protein sample into peptide fragments, while top-down proteomics will use whole proteins or large fragments. In both cases, peptides (for bottom-up) or intact proteins (for top-down) are separated by liquid chromatography (LC) prior to ionization and fragmentation for MS analysis. For top-down proteomics, intact proteins can undergo size separation prior to LC-MS, which helps the mass spectrometer to identify proteins or proteoforms in each sample. Many groups have recently published reviews on these topics, including a bottom-up review by Plubell et al.,[6] a top-down review by Po and Eyers,[7] and a general mass spectrometry review by Shuken.[8] Data-dependent acquisition (DDA) is often the focus of reviews because of its simplicity and wide use as an MS data acquisition technique. However, data-independent acquisition (DIA) has gained popularity due to its increased sensitivity in detecting lower-abundance peptides. This is possible because of computational advancements that identify peptides from complex mixtures containing multiple precursor ions. DIA MS was recently reviewed by Ludwig et al.[9] and Zhang et al.[10]

Regardless of the method selected, once mass spectra are acquired, they are mapped back to proteins based on a peptide search against the organism of study. To perform the mapping process, computational analysis software, such as DIA-NN,[11] MSFragger,[12] Skyline,[13] MaxQuant,[14] Protein Prospector,[15] Spectronaut,[16] and many more, will map the spectra to

**Special Issue:** Women in Proteomics and Metabolomics

**Published:** April 1, 2024





https://doi.org/10.1021/acs.jproteome.3c00892
J. Proteome Res. 2024, 23, 2723−2732

Table 1. Summary of Mass Spectrometry Approaches With Associated Benefits and Limitations of Each Approach

| approach | benefits | limitations |
|---|---|---|
| AP-MS | modular and scalable for use with multiple baits, can be used with proximity labeling for spatially and temporally resolved PPIs | protein interactions that are not biologically relevant may occur after cell lysis during standard AP-MS, unstable interactions may not be detected, direct vs indirect interactions and specific protein complexes cannot be distinguished |
| CF-MS | can identify proteins participating in multiple complexes, can be combined with AP approaches to increase resolution | protein interactions that are not biologically relevant may occur after cell lysis, protein complexes can be obscured by the co-fractionation of different complexes at the same time |
| TPCA | directly identifies protein complex members, compatible with multiplexed labeling to increase throughput | experimental conditions may alter melt temperatures, extensive pre-work may be required to generate melt curves |
| XL-MS | proteins and interactions are fixed *in situ* prior to lysis, improved protein complex identification over traditional AP-MS | crosslinked peptides require specialized analysis and FDR management |

peptides that theoretically match based on known amino acid sequences. The software then maps the peptides to proteins. Since spectra mapping to peptides can give false positives, decoy peptides, created through a variety of computational techniques, can be used to estimate the false discovery rate (FDR) and ensure it is acceptably low.

Variations on LC-MS provide varying degrees of resolution in PPI identification, depending on the study's goals. Methods available include classic affinity purification−mass spectrometry (AP-MS), cofractionation mass spectrometry (CF-MS), thermal proximity coaggregation (TPCA), and cross-linking mass spectrometry (XL-MS). Here, we provide brief explanations of these four approaches while discussing the advantages and caveats of each (Table 1).

As the name implies, AP-MS experiments rely on affinity purification of a protein of interest or "bait". Purification can involve immunoprecipitation using antibodies against the native protein or expressing bait fusions with an affinity tag (e.g., FLAG, HA, His, Strep, etc.). While purifying native proteins using specific antibodies preserves relevant biology (e.g., expression patterns), affinity tagging provides an attractive level of modularity for large-scale studies. Regardless of approach, the bait protein is purified from cell lysates using antibody-binding or affinity-tag-binding beads. LC-MS then identifies the interacting proteins, or "prey". The flexibility and modularity of AP-MS make it a valuable approach, and many groups have applied AP-MS to make major PPI discoveries across a broad range of biology, including ion channel function in the mammalian brain,[17] chromatin remodeling in plants,[18] and influenza A virus replication.[19] When conducted on a large scale, they can be used to map PPIs across an entire proteome[20] and infer stoichiometries of protein complexes.[21] Despite the advantages of AP-MS, there are several caveats to the approach. The sample lysis and purification can introduce background protein binding through the mixing of cellular compartments that are usually isolated in space and/or time. Detergents in lysis buffers can also disrupt transient and/or weakly binding proteins and protein complexes and may bias the data to PPIs with higher binding affinities.

One extension of the AP-MS approach that deals with many caveats of classic AP-MS is proximity labeling. This approach takes advantage of bait fusion with an enzyme that can label neighboring proteins with biotin moieties *in situ*. Subsequent AP-MS of the biotinylated proteins can define protein neighborhoods or complexes. Proximity labeling can help maintain low-affinity or transient interactions that would be lost by AP-MS alone. By purification of proteins labeled *in situ* instead of from lysates, proximity labeling avoids mixing proteins from different compartments or cell states that could give rise to spurious interactions. However, proximity labeling is also subject to off-target background labeling, and careful controls must be included to account for this.[22,23]

While many proximity labeling tools have been developed, we highlight technologies developed by Alice Ting's group and applied to identify PPIs and their dynamics.[23−27] APEX[23] (ascorbic acid peroxidase) and the related APEX2[25] catalyze the formation of biotin phenoxyl-radicals that spontaneously react with (primarily) tyrosine residues on timescales that limit diffusion to less than 20 nm. TurboID catalyzes direct biotinylation of target proteins through its biotin ligase activity.[24] Compared to its BioID precursor,[28] the engineered TurboID reaction kinetics are improved by orders of magnitude and are much closer to APEX2 labeling kinetics. Ting and collaborators have also created variations on these tools, including split versions to directly assay PPI-dependent labeling,[26] and smaller enzymes that may be better tolerated as fusions.[24] Though not as mature as traditional AP-MS, proximity labeling combined with AP-MS has already been used for large-scale PPI mapping efforts.[29] Taken together, AP-MS is a powerful MS-based technique to identify PPIs because it is straightforward, accessible, and modular, although technological innovations continue.

In CF-MS, samples undergo separation using size exclusion chromatography or other techniques.[30−32] CF-MS identifies proteins in each discrete fraction and links them together as members of a larger complex. Proteins can often take part in multiple complexes, or a complex can have different modifying members to give it unique functions. CF-MS enables efficient identification of these ensembles compared to AP-MS. However, samples are still lysed in solution prior to fractionation, which can bias detection toward complexes stable under specific lysis conditions. Multiple complexes may also cofractionate together, making data validation important. AP can also be added as an intermediate step to further reduce complexity or identify complexes with specific members.[31] Andrea Fossati and colleagues recently applied CF-MS to identify jumbophage-bacteria PPIs.[33] Jumbophages encode more than 300 proteins, and identifying PPIs using AP-MS is not necessarily practical. By using CF-MS, the authors were able to compare jumbophage-bacteria PPIs across two jumbophages and found evidence of shared phage predation mechanisms between the two viruses. While CF-MS is still relatively new, it holds promise for many systems.

TPCA was developed by Chris Tan and colleagues as a high-throughput method to separate interacting proteins based on denaturation temperature.[34] This technique is a unique application stemming from the cellular thermal shift assay (CETSA) developed in 2013 by Martinez Molina et al.[35] and thermal proteome profiling (TPP) developed in 2014 by Savitski et al.,[36] originally used to characterize melting point


https://doi.org/10.1021/acs.jproteome.3c00892
J. Proteome Res. 2024, 23, 2723−2732

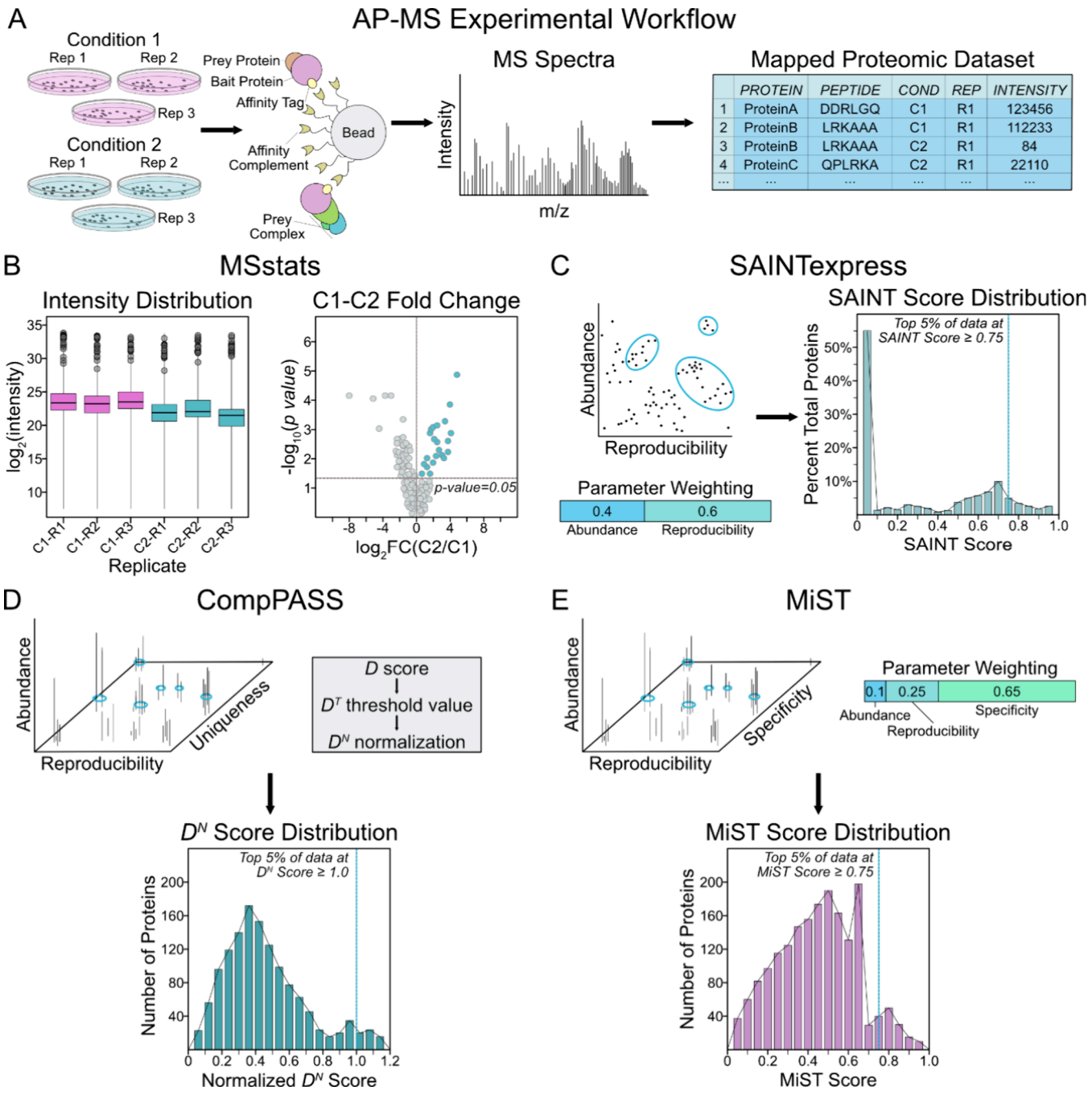



**Figure 1.** Brief explanations of four methods used when evaluating affinity purification−mass spectrometry (AP-MS) datasets. (A) A general experimental AP-MS workflow. Two conditions (C1 and C2) each with three replicates (R1, R2, and R3) are depicted. Samples undergo purification and mass spectrometry, where the peptide data are recorded and mapped to known proteins. (B) An example of MSstats outputs for an AP-MS dataset. MSstats can provide relative statistical quantitation for data in the form of a box plot showing data distribution across all conditions and replicates and a volcano plot showing fold change for proteins among the two conditions. Blue dots indicate proteins with a positive log-2 fold change and a *p*-value below 0.05 as an example of a high-confidence PPI. (C) Example of the SAINT and SAINTexpress output. Parameters for abundance and reproducibility are weighed against the dataset to provide a score associated with the presence of prey proteins in each replicate. (D) Example of the CompPASS output. The algorithm evaluates prey abundance, reproducibility, and uniqueness through all replicates and calculates a *D* score for each prey based on these criteria. A normalized *D* score plot is shown. (E) Example of MiST output for virus−host PPI data. MiST generates scores using the same three criteria as CompPASS, but the weighting for each criterion is set by the user.

shifts in proteins and protein complexes during ligand or drug binding.[37] TPCA uses MS and multiplexed quantification of the CETSA to generate melt curves for protein complexes. Since protein complexes will denature at a similar temperature and coaggregate, TPCA can identify proteins with similar melt curves and solubility behaviors to assign them to protein complexes. In an elegant application of TPCA, Joshua Justice and colleagues identify global changes in PPIs caused by herpesvirus infection.[38] The authors show that DNA sensor IFI16 (interferon gamma inducible protein 16) recruits DNA-PK (DNA-dependent protein kinase) early during infection. DNA-PK phosphorylates IFI16 on tyrosine 149 to initiate a


https://doi.org/10.1021/acs.jproteome.3c00892
J. Proteome Res. 2024, 23, 2723−2732

cytokine response. Although TPCA can be a powerful tool, cell lysis and other manipulations of the sample can change the proteins' environments and alter their thermal stabilities. Each system of interest may require new calibrated melt curves based on the experimental conditions. The computational analysis is also not trivial. Consequently, this method has not been applied as broadly as the others. Nonetheless, TPCA is far more scalable in identifying PPIs across the entire proteome and for many different conditions.[39]

XL-MS[40] uses a crosslinking reagent to create a covalent bond between adjacent peptides and can be combined with AP-MS (XL-AP-MS) to identify PPIs. When crosslinking occurs between neighboring proteins within a complex, this can stabilize protein complexes and even provide structural information about points of contact between the proteins. Reversible crosslinking mainly serves to stabilize transient interactions. Irreversible crosslinking will alter the spectra produced since the peptides remain covalently bonded for LC-MS analysis and can be used to identify which peptides are in contact. However, spectra derived from samples without reverse crosslinking must be mapped using specialized databases and algorithms, and the FDR calculation is not as straightforward. A recent large-scale study by Swantje Lenz and colleagues applied XL-MS to *Escherichia coli* lysates to recover 590 PPIs. The authors also demonstrated the power for structural discovery by successfully mapping the interaction domain between an uncharacterized protein YacL and RNA polymerase using the crosslinking data.[41] A more focused study by Sara Ayala Mariscal and colleagues used XL-MS to study the huntingtin protein (HTT), whose expanded polyglutamine (polyQ) tracts cause the neurodegenerative Huntington's Disease through HTT aggregation. The authors mapped a specific binding motif between HTT and chaperone DNAJB1 (DnaJ heat shock protein family member B1) using XL-MS. These results were used to target mutations to this binding motif and disrupt the interaction.[42] Given the advances in protein structure prediction and cryo-electron microscopy, XL-MS is an incredibly promising method to rapidly go from large-scale datasets down to amino acid resolution of PPIs.[43]

## YOU'VE GOT ME FOR YOUR PREY: PPI DATA CURATION AND SCORING TOOLS

After completion of the peptide mapping process (Figure 1A), the output dataset may contain tens or thousands of prey, not all of which indicate real or biologically relevant interactions with the bait(s). Combine this with the background contaminants that find their way into samples (e.g., hair keratins and bacterial proteins), and it becomes obvious that researchers could easily be stuck in a PPI web that mass spectrometry is spinning. Consequently, they need an unbiased method to remove these extraneous or irrelevant prey from their datasets. Some groups keep personal databases of known background and contaminant proteins while others use freely available tools such as CRAPome (Contaminant Repository for Affinity Purification).[44] Either way, elimination of extraneous proteins can reduce a dataset's size and provide a starting point for confident PPI identification. This helps to make both data analysis and complex inferences more tractable.

High-confidence interactions do not stem from a single datapoint, but instead arise from holistic analysis of the entire dataset using an unbiased and systematic scoring method. These scoring methods can be used to derive a threshold or minimum performance value that a protein must reach to be considered an interactor for a certain bait. Enrichment of specific proteins under multiple conditions can be established with statistical quantification tools such as MSstats,[45−47] a package created by Meena Choi and Olga Vitek (Figure 1B). MSstats normalizes data across runs to account for differences in total protein content and then performs direct comparisons between each condition and replicate to calculate enrichment (fold changes) and significance ($p$-values) for each protein. Statistical tools such as MSstats can perform direct comparisons and impute protein intensity or spectral counts for missing values. The user must define cutoffs for enrichment and significance to threshold their data.

Since AP-MS is still the most broadly used approach to identify PPIs, many data analysis tools have been developed to handle these datasets specifically. Algorithms and tools such as SAINT (Significance Analysis of INTeractome)[48] or SAINTexpress,[49] CompPASS (Comparative Proteomic Analysis Software Suite),[50] and MiST (Mass Spectrometry Interaction STatistics)[51] can take an AP-MS dataset and provide a numerical score for each PPI. These tools take a given dataset and review it for two or three criteria: the abundance of the protein, the reproducibility or presence of a PPI across multiple replicates, and the specificity or uniqueness of an interaction (for MiST and CompPASS only).

The SAINT/SAINTexpress[48,49] algorithms perform quantitative data analysis and probabilistic scoring of a proteomic dataset as a means of evaluating the likelihood of a bait−prey pair interaction (Figure 1C). In its original form, SAINT used only the quantitative data associated with each prey's total spectral counts recorded in a bait condition and compared them against all baits and their replicates. Negative controls were not required if sufficient independent bait conditions were profiled together, permitting analysis of both small and large datasets. Using statistical modeling, a prey's counts in a single bait are modeled from a Poisson distribution representing either a true interaction or a false interaction. Distributions are calculated for each bait−prey pair throughout the dataset in a matrix, resulting in a probability associated with each PPI. The probabilities help to determine a Bayesian FDR and a threshold probability needed to achieve it. SAINTexpress expanded upon SAINT with the ability to perform calculations using protein intensity values in addition to spectral counts. It also contains information on existing PPI data, allowing it to supplement the main score with a separate topology-based score, the topology-aware average probability score (TopoAvgP), which can improve the identification of copurifying prey−prey protein complexes.

CompPASS[50] operates similarly to SAINT and SAINTexpress in that both will automatically output PPI probability values for each bait−prey pair in a dataset (Figure 1D). The CompPASS algorithm calculates two scores for the bait−prey pairs: a $Z$ score to normalize and center the data and a $D$ score based on the adjustment of total spectral counts (TSCs) across all bait conditions and replicates. The $D$ score incorporates prey uniqueness under bait conditions, protein abundance (as TSC), and reproducibility across replicates to create a representative score for the three attributes. CompPASS then calculates a threshold $D$ score ($D^{T}$) to ensure 95% (or other configurable amount) of the data falls below this threshold. Raw $D$ ($D^{R}$) scores are normalized against $D^{T}$, producing normalized $D$ scores ($D^{N}$) that can be plotted. Preys with a $D^{N} \geq 1.0$ are considered high-confidence PPIs. If preservation of a


https://doi.org/10.1021/acs.jproteome.3c00892
J. Proteome Res. 2024, 23, 2723−2732

low-scoring PPI is desired, the *Z* score can provide evidence for further validation.

MiST[51] uses the mapped proteomic data to generate scores for specificity, reproducibility, and abundance (Figure 1E). Each of these scores is then combined in a linear combination to produce an overall MiST score. The weighting parameters for each of the three components sum to 1.0 and are typically weighted more toward specificity and reproducibility (specificity ∼65%, reproducibility ∼25%, abundance ∼10%). The emphasis on specificity emerges from the fact that MiST was originally developed to score virus−host PPIs for RNA viruses. RNA viruses have some of the smallest genomes (typically 10 kilobases or less). Evolving redundant interactions, in which multiple viral proteins interact with the same host protein, is inefficient. Consequently, the creators of MiST theorized that specificity of protein interactions would help recover gold standard virus−host PPIs.[51] However, scoring weights can be tailored to a user's dataset, provided there is a mechanism to evaluate the precision and recall of the tailored weights.[51,52]

Regardless of the scoring approach used, the end result will require some decision on a threshold above which PPIs are considered "high-confidence". This will vary depending on the dataset and scoring approach but is typically set to capture the top ∼5% of scored data or optimized based on precision and recall against a gold standard dataset. A dataset originally containing thousands of potential PPIs will retain only a few (tens to hundreds) high-confidence interactors. The reality of the experiment has finally shown itself: many PPIs identified through MS do not reflect actual biological relevance due to their promiscuity or chance contact that resulted in purification. If a PPI of particular interest is known to interact with a bait and complex member but does not make it through a strict scoring threshold, systematic approaches to rescuing low-scoring interactions have been established.[53] It is important to recognize that these scoring algorithms, while powerful in untangling spiderwebs, can evaluate only the data provided. Ultimately, refined high-confidence PPI networks are a starting point to generate hypotheses and require further testing. Next, we highlight studies that span this spectrum of hypothesis generation and testing.

## BREADTH OR DEPTH? ADVANTAGES OF COMPREHENSIVE AND FOCUSED PPI STUDIES

Comprehensive studies of PPIs require large investments of resources, but their size can provide advantages in internal validation. For example, a recent study by André Michaelis and colleagues provided a more complete view of the *Saccharomyces cerevisiae* protein−protein interactome.[20] The authors performed systematic purifications of nearly 4,000 green fluorescent protein (GFP)-tagged *S. cerevisiae* proteins to identify over 31,000 PPIs. This doubled the number of proteins and tripled the number of high-confidence interactions previously identified by other large-scale efforts.[54−56] Reverse or reciprocal purifications are often used to validate specific PPIs, as Edward Huttlin and colleagues did on a large scale for BioPlex, a similar effort to map the human protein−protein interactome.[57−59] The near complete nature of the study in *S. cerevisiae* by Michaelis and colleagues meant that almost all of these reverse purifications were already part of the dataset and could be incorporated into the approach as one of three criteria used for scoring. While not all proteins encoded by *S. cerevisiae* could be successfully tagged with GFP for purification in this study, the authors leveraged their success with other proteins to fill in the gaps presented by biochemically challenging proteins. For example, proteins from the chaperonin-containing T-complex (CCT) cannot be tagged because the tag will interfere with complex formation and function. However, the authors were able to infer CCT interactions between the eight subunits and identify novel interactions through purification of proteins that interact with one or more subunits of CCT.

As the study by Michaelis and colleagues demonstrates, the data generated by large-scale studies often drive innovation in proteomic analysis. Many of the MS scoring algorithms discussed earlier were in fact developed to systematically deal with the large PPI datasets generated for first-of-their-kind comprehensive studies. SAINT was originally designed to score yeast kinase protein interactions[48] and has been broadly applied to PPI studies. CompPASS was originally developed to score human PPI data for a comprehensive ubiquitin ligase interaction network[50] and has since been applied to numerous PPI studies, including virus−host PPIs. MiST was created to score virus−host PPI data for human immunodeficiency virus 1[51] and has been applied to many other virus−host studies since.[60−63]

Other more recent studies have developed novel comparative scoring methods to identify disease-related PPIs, since the simple overlap of networks may not be sufficient to establish the presence or absence of a PPI in a specific disease-related condition. We used holistic enrichment scoring at the pathway and complex level to identify similarities and differences across flavivirus−host PPI networks.[52] This approach highlighted a protein interaction between viral protein NS4A and the Sec61 translocon conserved across two flaviviruses and two hosts and mediated fundamental aspects of virus replication. David Gordon and colleagues developed a differential interaction score to compare three coronavirus−host PPI networks at the protein level.[53] This approach benefits from capturing differences that fall below a strict cutoff. Another study by Danielle Swaney and colleagues took comparative protein interaction mapping even further by identifying interactions for proteins implicated in head and neck squamous cell carcinoma.[64] In total, the authors identified protein interactions for 31 wild type and 23 mutant proteins in two head and neck cancer cell lines and one non-tumor esophageal cell line. By including a non-cancer cell line derived from a similar tissue and applying differential interaction scoring, the authors identified interactions relevant to cancer biology. For example, an interaction between Cyclin D1 (CCND1) and members of the PI3K complex was lost in cancer cells, while an interaction between fibroblast growth factor receptor 3 (FGFR3) and Daple (CCDC88C) was gained in cancer cells and activated cell migratory proteins.

Focused studies on single proteins and their interaction networks can take advantage of more complex and elegant experimental systems and proteomic analyses. Todd Greco and colleagues performed one such study on the HTT protein.[5] The authors expressed HTT with a normal (20) or highly expanded (140) polyQ tract in mice and affinity purified the protein from young (2 months old) and aged (10 months old) brains. In this way, the authors identified 278 HTT-interacting proteins, including some that were dependent on age or polyQ expansions. They analyzed the stability of 72% of PPIs using an isotope labeling approach in tissues and found that interactions became more stable with aging. Thus, in addition to increases or decreases of specific interactions driving disease, interaction


https://doi.org/10.1021/acs.jproteome.3c00892
J. Proteome Res. 2024, 23, 2723−2732

dynamics likely contribute to the dysfunction of HTT during aging. They validated 22 of these interactions by a luciferase two-hybrid, though some may interact indirectly. Finally, an elegant functional assay in a fruit fly model of Huntington's Disease demonstrated that many of these HTT-interacting proteins modulate HTT-induced neuronal dysfunction.

While both are impressive in their own ways, the goals of a resource creation study are fundamentally different from that of a more focused mechanistic study. The scoring and follow-up validation will also differ based on these goals. In creating a comprehensive dataset that spans an entire set of proteins from a specific biological process, disease, or proteome, authors can take advantage of this scale in their validation. As stated earlier, reciprocal purifications are often built in and can even be incorporated into novel scoring methods. Additionally, scoring cutoffs for larger studies can be optimized through precision and recall analysis of validated, gold standard interactions.[51,52,65] This approach can better identify and eliminate the proteomic background. On the other hand, smaller studies focused on a single protein can more easily sample PPIs in biologically relevant systems. However, they will have a limited ability to identify and remove proteomic background, and validation of specific interactions becomes much more important as a result.

## YOU THINK THAT WE CONNECT: USING PROTEIN STRUCTURE AND INTERACTION PREDICTIONS TO SCORE PPI DATA

With the emergence of AI-based structural prediction systems like AlphaFold 2 (AlphaFold)[66] and RoseTTAFold,[67] we can now predict protein structures and interactions with more accuracy and speed than ever before.[68] Another algorithm, Protein Structure Transformer (PeSTo),[69] predicts protein binding interfaces. Requiring only the protein's structure from either experimental or predicted sources, PeSTo has shown marked accuracy when tested on a benchmark dataset. While many groups have used AlphaFold to identify potential binding sites and drive more mechanistic studies,[20,65] there is a potential to reverse the workflow. Instead, protein interaction and binding interface scores from tools like AlphaFold, RoseTTAFold, or PeSTo could be used as an input for AP-MS scoring. For example, structure prediction could be used to refine scoring thresholds or rescue specific interactions that fall below a high-confidence threshold.

More generally, the protein structure predictions themselves could be valuable for PPI scoring. Until recently, it was computationally limiting to quickly perform structural alignment on multiple proteins. Previous sequence and structural alignment methods such as Dali,[70] FATCAT,[71] and TM-align (now US-align)[72,73] performed with reliable sensitivity and accuracy but could not scale to many comparisons, limiting their application in a larger workflow. Protein structural alignment tools now incorporate neural network models to perform searches against entire structural databases or proteomes at speeds faster than ever before. In 2023, van Kempen et al. released Foldseek, which performs structural alignment for a protein against any combination of protein structures with high accuracy.[74] Foldseek achieves this speed and accuracy in part by converting a protein's tertiary structure into a simplified character sequence from Foldseek's 3D interaction (3Di) alphabet. Each pair of residues closest to one another is described by one of 20 characters in this 3Di alphabet. This simplifies the alignment from a full structural alignment while adding rigor when compared with backbone-based structural alphabets or amino acid sequence alignments. When combined with the predictive power of AlphaFold and its protein structure prediction databases, Foldseek can search a query against entire proteomes and identify possible structurally homologous proteins. This could be especially valuable for PPI studies in non-model organisms, for which databases compiling experimentally resolved structures and validated protein complexes are limited. In this case, protein structure homology could be used to efficiently identify conserved protein complexes and add confidence to PPI scores when the gold standard reference interactions are unavailable.

## CONCLUSIONS

MS-based identification of PPIs is an exciting area of discovery science, but it can often feel like walking into spiderwebs. These webs can be unraveled into tractable threads using a variety of scoring approaches. Studies focused on large interactomes with many bait proteins can leverage existing PPI databases and built-in internal validation to guide scoring and thresholding. Targeted studies can employ complex experimental systems and involve more follow-up validation and mechanistic studies. Regardless of study size, new innovations in analyzing PPI datasets will be possible through incorporation of protein structure and interaction prediction tools.

## AUTHOR INFORMATION


### Corresponding Author

**Priya S. Shah** − *Department of Chemical Engineering, University of California, Davis 95616 California, United States; Department of Microbiology and Molecular Genetics, University of California, Davis 95616 California, United States;* orcid.org/0000-0001-7518-2839; Email: prsshah@ucdavis.edu

### Author

**Chase L. S. Skawinski** − *Department of Chemical Engineering, University of California, Davis 95616 California, United States;* orcid.org/0009-0008-3081-3981

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.3c00892


### Author Contributions

P.S.S. and C.L.S.S. conceived, wrote, and edited the manuscript together. P.S.S. secured funding for the manuscript.


### Funding

P.S.S. was supported by both NIH R01AI170857 and the Laboratory Directed Research and Development Program at Sandia National Laboratories under Project 233120. C.L.S.S. was supported by the UC Davis Dean's Distinguished Graduate Fellowship.


### Notes

To specifically identify women in proteomics who contributed work featured in this article, we have listed their references by the number in the text and whether they are a first author, corresponding author, or both. First authors: refs 6, 9, 10, 17−19, 21, 23, 31, 41, 45, 51, 52, 54, 63−65, and 67. Corresponding authors: refs 2, 3, 5, 7, 26, 27, 29, 38, and 43.



https://doi.org/10.1021/acs.jproteome.3c00892
J. Proteome Res. 2024, 23, 2723−2732

First and corresponding authors: refs 4, 22, 24, 25, 33, 37, 42, and 60−62.

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We acknowledge No Doubt for inspiration while writing this manuscript. We also acknowledge the women proteomics researchers whose work paved the way for this review, especially the many we could not cite due to space constraints. We thank Nicholas J. Lopez for critical feedback on the manuscript and Matthew W. Kenaston for critical feedback on the graphical abstract.

## ■ REFERENCES

(1) Gong, Y.; Kakihara, Y.; Krogan, N.; Greenblatt, J.; Emili, A.; Zhang, Z.; Houry, W. A. An Atlas of Chaperone-Protein Interactions in Saccharomyces Cerevisiae: Implications to Protein Folding Pathways in the Cell. *Mol. Syst. Biol.* **2009**, *5*, 275.

(2) Gestaut, D.; Roh, S. H.; Ma, B.; Pintilie, G.; Joachimiak, L. A.; Leitner, A.; Walzthoeni, T.; Aebersold, R.; Chiu, W.; Frydman, J. The Chaperonin TRiC/CCT Associates with Prefoldin through a Conserved Electrostatic Interface Essential for Cellular Proteostasis. *Cell* **2019**, *177* (3), 751−765.

(3) Daly, L. A.; Brownridge, P. J.; Batie, M.; Rocha, S.; Sée, V.; Eyers, C. E. Oxygen-Dependent Changes in Binding Partners and Post-Translational Modifications Regulate the Abundance and Activity of HIF-1$\alpha$/2$\alpha$. *Sci. Signal.* **2021**, *14* (692), eabf6685.

(4) Fang, J.; Pietzsch, C.; Tsaprailis, G.; Crynen, G.; Cho, K. F.; Ting, A. Y.; Bukreyev, A.; de la Torre, J. C.; Saphire, E. O. Functional Interactomes of the Ebola Virus Polymerase Identified by Proximity Proteomics in the Context of Viral Replication. *Cell Rep.* **2022**, *38* (12), 110544.

(5) Greco, T. M.; Secker, C.; Silva Ramos, E.; Federspiel, J. D.; Liu, J.-P.; Perez, A. M.; Al-Ramahi, I.; Cantle, J. P.; Carroll, J. B.; Botas, J.; Zeitlin, S. O.; Wanker, E. E.; Cristea, I. M. Dynamics of Huntingtin Protein Interactions in the Striatum Identifies Candidate Modifiers of Huntington Disease. *Cell Syst.* **2022**, *13* (4), 304−320.

(6) Plubell, D. L.; Käll, L.; Webb-Robertson, B.-J.; Bramer, L. M.; Ives, A.; Kelleher, N. L.; Smith, L. M.; Montine, T. J.; Wu, C. C.; MacCoss, M. J. Putting Humpty Dumpty Back Together Again: What Does Protein Quantification Mean in Bottom-Up Proteomics? *J. Proteome Res.* **2022**, *21* (4), 891−898.

(7) Po, A.; Eyers, C. E. Top-Down Proteomics and the Challenges of True Proteoform Characterization. *J. Proteome Res.* **2023**, *22* (12), 3663−3675.

(8) Shuken, S. R. An Introduction to Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2023**, *22* (7), 2151−2171.

(9) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data-Independent Acquisition-Based SWATH-MS for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2018**, *14* (8), e8126.

(10) Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020**, *20* (17-18), e1900276.

(11) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2020**, *17* (1), 41−44.

(12) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry-Based Proteomics. *Nat. Methods* **2017**, *14* (5), 513−520.

(13) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinforma. Oxf. Engl.* **2010**, *26* (7), 966−968.

(14) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367−1372.

(15) Chalkley, R. J.; Baker, P. R.; Huang, L.; Hansen, K. C.; Allen, N. P.; Rexach, M.; Burlingame, A. L. Comprehensive Analysis of a Multidimensional Liquid Chromatography Mass Spectrometry Dataset Acquired on a Quadrupole Selecting, Quadrupole Collision Cell, Time-of-Flight Mass Spectrometer: II. New Developments in Protein Prospector Allow for Reliable and Comprehensive Automatic Analysis of Large Datasets. *Mol. Cell. Proteomics MCP* **2005**, *4* (8), 1194−1204.

(16) Spectronaut Software Page. https://biognosys.com/software/spectronaut/ (accessed 2023-12-10).

(17) Kollewe, A.; Chubanov, V.; Tseung, F. T.; Correia, L.; Schmidt, E.; Rössig, A.; Zierler, S.; Haupt, A.; Müller, C. S.; Bildl, W.; Schulte, U.; Nicke, A.; Fakler, B.; Gudermann, T. The Molecular Appearance of Native TRPM7 Channel Complexes Identified by High-Resolution Proteomics. *eLife* **2021**, *10*, e68544.

(18) Guo, J.; Cai, G.; Li, Y.-Q.; Zhang, Y.-X.; Su, Y.-N.; Yuan, D.-Y.; Zhang, Z.-C.; Liu, Z.-Z.; Cai, X.-W.; Guo, J.; Li, L.; Chen, S.; He, X.-J. Comprehensive Characterization of Three Classes of Arabidopsis SWI/SNF Chromatin Remodelling Complexes. *Nat. Plants* **2022**, *8* (12), 1423−1439.

(19) Haas, K. M.; McGregor, M. J.; Bouhaddou, M.; Polacco, B. J.; Kim, E.-Y.; Nguyen, T. T.; Newton, B. W.; Urbanowski, M.; Kim, H.; Williams, M. A. P.; Rezelj, V. V.; Hardy, A.; Fossati, A.; Stevenson, E. J.; Sukerman, E.; Kim, T.; Penugonda, S.; Moreno, E.; Braberg, H.; Zhou, Y.; Metreveli, G.; Harjai, B.; Tummino, T. A.; Melnyk, J. E.; Soucheray, M.; Batra, J.; Pache, L.; Martin-Sancho, L.; Carlson-Stevermer, J.; Jureka, A. S.; Basler, C. F.; Shokat, K. M.; Shoichet, B. K.; Shriver, L. P.; Johnson, J. R.; Shaw, M. L.; Chanda, S. K.; Roden, D. M.; Carter, T. C.; Kottyan, L. C.; Chisholm, R. L.; Pacheco, J. A.; Smith, M. E.; Schrodi, S. J.; Albrecht, R. A.; Vignuzzi, M.; Zuliani-Alvarez, L.; Swaney, D. L.; Eckhardt, M.; Wolinsky, S. M.; White, K. M.; Hultquist, J. F.; Kaake, R. M.; García-Sastre, A.; Krogan, N. J. Proteomic and Genetic Analyses of Influenza A Viruses Identify Pan-Viral Host Targets. *Nat. Commun.* **2023**, *14* (1), 6030.

(20) Michaelis, A. C.; Brunner, A.-D.; Zwiebel, M.; Meier, F.; Strauss, M. T.; Bludau, I.; Mann, M. The Social and Structural Architecture of the Yeast Protein Interactome. *Nature* **2023**, *624* (7990), 192−200.

(21) Hein, M. Y.; Hubner, N. C.; Poser, I.; Cox, J.; Nagaraj, N.; Toyoda, Y.; Gak, I. A.; Weisswange, I.; Mansfeld, J.; Buchholz, F.; Hyman, A. A.; Mann, M. A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **2015**, *163* (3), 712−723.

(22) Zhong, X.; Li, Q.; Polacco, B. J.; Patil, T.; DiBerto, J. F.; Vartak, R.; Xu, J.; Marley, A.; Foussard, H.; Roth, B. L.; Eckhardt, M.; Von Zastrow, M.; Krogan, N. J.; Hüttenhain, R. An Automated Proximity Proteomics Pipeline for Subcellular Proteome and Protein Interaction Mapping. *BioRxiv Prepr. Serv. Biol.* **2023**.

(23) Lobingier, B. T.; Hüttenhain, R.; Eichel, K.; Miller, K. B.; Ting, A. Y.; von Zastrow, M.; Krogan, N. J. An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell* **2017**, *169* (2), 350−360.

(24) Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y. Efficient Proximity Labeling in Living Cells and Organisms with TurboID. *Nat. Biotechnol.* **2018**, *36* (9), 880−887.

(25) Lam, S. S.; Martell, J. D.; Kamer, K. J.; Deerinck, T. J.; Ellisman, M. H.; Mootha, V. K.; Ting, A. Y. Directed Evolution of APEX2 for Electron Microscopy and Proximity Labeling. *Nat. Methods* **2015**, *12* (1), 51−54.

(26) Cho, K. F.; Branon, T. C.; Rajeev, S.; Svinkina, T.; Udeshi, N. D.; Thoudam, T.; Kwak, C.; Rhee, H.-W.; Lee, I.-K.; Carr, S. A.; Ting, A. Y. Split-TurboID Enables Contact-Dependent Proximity Labeling in Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (22), 12143−12154.



https://doi.org/10.1021/acs.jproteome.3c00892
J. Proteome Res. 2024, 23, 2723−2732

(27) Martell, J. D.; Yamagata, M.; Deerinck, T. J.; Phan, S.; Kwa, C. G.; Ellisman, M. H.; Sanes, J. R.; Ting, A. Y. A Split Horseradish Peroxidase for the Detection of Intercellular Protein-Protein Interactions and Sensitive Visualization of Synapses. *Nat. Biotechnol.* **2016**, *34* (7), 774−780.
(28) Roux, K. J.; Kim, D. I.; Burke, B. BioID: A Screen for Protein-Protein Interactions. *Curr. Protoc. Protein Sci.* **2013**, *74*, 19.23.1−19.23.14.
(29) Go, C. D.; Knight, J. D. R.; Rajasekharan, A.; Rathod, B.; Hesketh, G. G.; Abe, K. T.; Youn, J.-Y.; Samavarchi-Tehrani, P.; Zhang, H.; Zhu, L. Y.; Popiel, E.; Lambert, J.-P.; Coyaud, É.; Cheung, S. W. T.; Rajendran, D.; Wong, C. J.; Antonicka, H.; Pelletier, L.; Palazzo, A. F.; Shoubridge, E. A.; Raught, B.; Gingras, A.-C. A Proximity-Dependent Biotinylation Map of a Human Cell. *Nature* **2021**, *595* (7865), 120−124.
(30) Drew, K.; Müller, C. L.; Bonneau, R.; Marcotte, E. M. Identifying Direct Contacts between Protein Complex Subunits from Their Conditional Dependence in Proteomics Datasets. *PLoS Comput. Biol.* **2017**, *13* (10), e1005625.
(31) McWhite, C. D.; Papoulas, O.; Drew, K.; Dang, V.; Leggere, J. C.; Sae-Lee, W.; Marcotte, E. M. Co-Fractionation/Mass Spectrometry to Identify Protein Complexes. *STAR Protoc.* **2021**, *2* (1), 100370.
(32) Havugimana, P. C.; Goel, R. K.; Phanse, S.; Youssef, A.; Padhorny, D.; Kotelnikov, S.; Kozakov, D.; Emili, A. Scalable Multiplex Co-Fractionation/Mass Spectrometry Platform for Accelerated Protein Interactome Discovery. *Nat. Commun.* **2022**, *13* (1), 4043.
(33) Fossati, A.; Mozumdar, D.; Kokontis, C.; Mèndez-Moran, M.; Nieweglowska, E.; Pelin, A.; Li, Y.; Guo, B.; Krogan, N. J.; Agard, D. A.; Bondy-Denomy, J.; Swaney, D. L. Next-Generation Proteomics for Quantitative Jumbophage-Bacteria Interaction Mapping. *Nat. Commun.* **2023**, *14* (1), 5156.
(34) Tan, C. S. H.; Go, K. D.; Bisteau, X.; Dai, L.; Yong, C. H.; Prabhu, N.; Ozturk, M. B.; Lim, Y. T.; Sreekumar, L.; Lengqvist, J.; Tergaonkar, V.; Kaldis, P.; Sobota, R. M.; Nordlund, P. Thermal Proximity Coaggregation for System-Wide Profiling of Protein Complex Dynamics in Cells. *Science* **2018**, *359* (6380), 1170−1177.
(35) Martinez Molina, D.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science* **2013**, *341* (6141), 84−87.
(36) Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Martinez Molina, D.; Jafari, R.; Dovega, R. B.; Klaeger, S.; Kuster, B.; Nordlund, P.; Bantscheff, M.; Drewes, G. Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome. *Science* **2014**, *346* (6205), 1255784.
(37) George, A. L.; Sidgwick, F. R.; Watt, J. E.; Martin, M. P.; Trost, M.; Marín-Rubio, J. L.; Dueñas, M. E. Comparison of Quantitative Mass Spectrometric Methods for Drug Target Identification by Thermal Proteome Profiling. *J. Proteome Res.* **2023**, *22* (8), 2629−2640.
(38) Justice, J. L.; Kennedy, M. A.; Hutton, J. E.; Liu, D.; Song, B.; Phelan, B.; Cristea, I. M. Systematic Profiling of Protein Complex Dynamics Reveals DNA-PK Phosphorylation of IFI16 En Route to Herpesvirus Immunity. *Sci. Adv.* **2021**, *7* (25), eabg6680.
(39) Dai, L.; Zhao, T.; Bisteau, X.; Sun, W.; Prabhu, N.; Lim, Y. T.; Sobota, R. M.; Kaldis, P.; Nordlund, P. Modulation of Protein-Interaction States through the Cell Cycle. *Cell* **2018**, *173* (6), 1481−1494.
(40) Chavez, J. D.; Mohr, J. P.; Mathay, M.; Zhong, X.; Keller, A.; Bruce, J. E. Systems Structural Biology Measurements by in Vivo Cross-Linking with Mass Spectrometry. *Nat. Protoc.* **2019**, *14* (8), 2318−2343.
(41) Lenz, S.; Sinn, L. R.; O'Reilly, F. J.; Fischer, L.; Wegner, F.; Rappsilber, J. Reliable Identification of Protein-Protein Interactions by Crosslinking Mass Spectrometry. *Nat. Commun.* **2021**, *12* (1), 3564.
(42) Ayala Mariscal, S. M.; Pigazzini, M. L.; Richter, Y.; Özel, M.; Grothaus, I. L.; Protze, J.; Ziege, K.; Kulke, M.; ElBediwi, M.; Vermaas, J. V.; Colombi Ciacchi, L.; Köppen, S.; Liu, F.; Kirstein, J. Identification of a HTT-Specific Binding Motif in DNAJB1 Essential for Suppression and Disaggregation of HTT. *Nat. Commun.* **2022**, *13* (1), 4692.
(43) Yu, C.; Huang, L. New Advances in Cross-Linking Mass Spectrometry toward Structural Systems Biology. *Curr. Opin. Chem. Biol.* **2023**, *76*, 102357.
(44) Mellacheruvu, D.; Wright, Z.; Couzens, A. L.; Lambert, J.-P.; St-Denis, N.; Li, T.; Miteva, Y. V.; Hauri, S.; Sardiu, M. E.; Low, T. Y.; Halim, V. A.; Bagshaw, R. D.; Hubner, N. C.; al-Hakim, A.; Bouchard, A.; Faubert, D.; Fermin, D.; Dunham, W. H.; Goudreault, M.; Lin, Z.-Y.; Badillo, B. G.; Pawson, T.; Durocher, D.; Coulombe, B.; Aebersold, R.; Superti-Furga, G.; Colinge, J.; Heck, A. J. R.; Choi, H.; Gstaiger, M.; Mohammed, S.; Cristea, I. M.; Bennett, K. L.; Washburn, M. P.; Raught, B.; Ewing, R. M.; Gingras, A.-C.; Nesvizhskii, A. I. The CRAPome: A Contaminant Repository for Affinity Purification Mass Spectrometry Data. *Nat. Methods* **2013**, *10* (8), 730−736.
(45) Choi, M.; Chang, C.-Y.; Clough, T.; Broudy, D.; Killeen, T.; MacLean, B.; Vitek, O. MSstats: An R Package for Statistical Analysis of Quantitative Mass Spectrometry-Based Proteomic Experiments. *Bioinforma. Oxf. Engl.* **2014**, *30* (17), 2524−2526.
(46) Kohler, D.; Staniak, M.; Tsai, T.-H.; Huang, T.; Shulman, N.; Bernhardt, O. M.; MacLean, B. X.; Nesvizhskii, A. I.; Reiter, L.; Sabido, E.; Choi, M.; Vitek, O. MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale. *J. Proteome Res.* **2023**, *22* (5), 1466−1482.
(47) R: A Language and Environment for Statistical Computing. R Core Team, 2021. https://www.R-project.org/.
(48) Choi, H.; Larsen, B.; Lin, Z.-Y.; Breitkreutz, A.; Mellacheruvu, D.; Fermin, D.; Qin, Z. S.; Tyers, M.; Gingras, A.-C.; Nesvizhskii, A. I. SAINT: Probabilistic Scoring of Affinity Purification-Mass Spectrometry Data. *Nat. Methods* **2011**, *8* (1), 70−73.
(49) Teo, G.; Liu, G.; Zhang, J.; Nesvizhskii, A. I.; Gingras, A.-C.; Choi, H. SAINTexpress: Improvements and Additional Features in Significance Analysis of INTeractome Software. *J. Proteomics* **2014**, *100*, 37−43.
(50) Sowa, M. E.; Bennett, E. J.; Gygi, S. P.; Harper, J. W. Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell* **2009**, *138* (2), 389−403.
(51) Jäger, S.; Cimermancic, P.; Gulbahce, N.; Johnson, J. R.; McGovern, K. E.; Clarke, S. C.; Shales, M.; Mercenne, G.; Pache, L.; Li, K.; Hernandez, H.; Jang, G. M.; Roth, S. L.; Akiva, E.; Marlett, J.; Stephens, M.; D'Orso, I.; Fernandes, J.; Fahey, M.; Mahon, C.; O'Donoghue, A. J.; Todorovic, A.; Morris, J. H.; Maltby, D. A.; Alber, T.; Cagney, G.; Bushman, F. D.; Young, J. A.; Chanda, S. K.; Sundquist, W. I.; Kortemme, T.; Hernandez, R. D.; Craik, C. S.; Burlingame, A.; Sali, A.; Frankel, A. D.; Krogan, N. J. Global Landscape of HIV-Human Protein Complexes. *Nature* **2011**, *481* (7381), 365−370.
(52) Shah, P. S.; Link, N.; Jang, G. M.; Sharp, P. P.; Zhu, T.; Swaney, D. L.; Johnson, J. R.; Von Dollen, J.; Ramage, H. R.; Satkamp, L.; Newton, B.; Hüttenhain, R.; Petit, M. J.; Baum, T.; Everitt, A.; Laufman, O.; Tassetto, M.; Shales, M.; Stevenson, E.; Iglesias, G. N.; Shokat, L.; Tripathi, S.; Balasubramaniam, V.; Webb, L. G.; Aguirre, S.; Willsey, A. J.; Garcia-Sastre, A.; Pollard, K. S.; Cherry, S.; Gamarnik, A. V.; Marazzi, I.; Taunton, J.; Fernandez-Sesma, A.; Bellen, H. J.; Andino, R.; Krogan, N. J. Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* **2018**, *175* (7), 1931−1945.
(53) Gordon, D. E.; Hiatt, J.; Bouhaddou, M.; Rezelj, V. V.; Ulferts, S.; Braberg, H.; Jureka, A. S.; Obernier, K.; Guo, J. Z.; Batra, J.; Kaake, R. M.; Weckstein, A. R.; Owens, T. W.; Gupta, M.; Pourmal, S.; Titus, E. W.; Cakir, M.; Soucheray, M.; McGregor, M.; Cakir, Z.; Jang, G.; O'Meara, M. J.; Tummino, T. A.; Zhang, Z.; Foussard, H.; Rojc, A.; Zhou, Y.; Kuchenov, D.; Hüttenhain, R.; Xu, J.; Eckhardt, M.;


https://doi.org/10.1021/acs.jproteome.3c00892
J. Proteome Res. 2024, 23, 2723−2732

Swaney, D. L.; Fabius, J. M.; Ummadi, M.; Tutuncuoglu, B.; Rathore, U.; Modak, M.; Haas, P.; Haas, K. M.; Naing, Z. Z. C.; Pulido, E. H.; Shi, Y.; Barrio-Hernandez, I.; Memon, D.; Petsalaki, E.; Dunham, A.; Marrero, M. C.; Burke, D.; Koh, C.; Vallet, T.; Silvas, J. A.; Azumaya, C. M.; Billesbølle, C.; Brilot, A. F.; Campbell, M. G.; Diallo, A.; Dickinson, M. S.; Diwanji, D.; Herrera, N.; Hoppe, N.; Kratochvil, H. T.; Liu, Y.; Merz, G. E.; Moritz, M.; Nguyen, H. C.; Nowotny, C.; Puchades, C.; Rizo, A. N.; Schulze-Gahmen, U.; Smith, A. M.; Sun, M.; Young, I. D.; Zhao, J.; Asarnow, D.; Biel, J.; Bowen, A.; Braxton, J. R.; Chen, J.; Chio, C. M.; Chio, U. S.; Deshpande, I.; Doan, L.; Faust, B.; Flores, S.; Jin, M.; Kim, K.; Lam, V. L.; Li, F.; Li, J.; Li, Y.-L.; Li, Y.; Liu, X.; Lo, M.; Lopez, K. E.; Melo, A. A.; Moss, F. R.; Nguyen, P.; Paulino, J.; Pawar, K. I.; Peters, J. K.; Pospiech, T. H.; Safari, M.; Sangwan, S.; Schaefer, K.; Thomas, P. V.; Thwin, A. C.; Trenker, R.; Tse, E.; Tsui, T. K. M.; Wang, F.; Whitis, N.; Yu, Z.; Zhang, K.; Zhang, Y.; Zhou, F.; Saltzberg, D.; QCRG Structural Biology Consortium; Hodder, A. J.; Shun-Shion, A. S.; Williams, D. M.; White, K. M.; Rosales, R.; Kehrer, T.; Miorin, L.; Moreno, E.; Patel, A. H.; Rihn, S.; Khalid, M. M.; Vallejo-Gracia, A.; Fozouni, P.; Simoneau, C. R.; Roth, T. L.; Wu, D.; Karim, M. A.; Ghoussaini, M.; Dunham, I.; Berardi, F.; Weigang, S.; Chazal, M.; Park, J.; Logue, J.; McGrath, M.; Weston, S.; Haupt, R.; Hastie, C. J.; Elliott, M.; Brown, F.; Burness, K. A.; Reid, E.; Dorward, M.; Johnson, C.; Wilkinson, S. G.; Geyer, A.; Giesel, D. M.; Baillie, C.; Raggett, S.; Leech, H.; Toth, R.; Goodman, N.; Keough, K. C.; Lind, A. L.; Zoonomia Consortium; Klesh, R. J.; Hemphill, K. R.; Carlson-Stevermer, J.; Oki, J.; Holden, K.; Maures, T.; Pollard, K. S.; Sali, A.; Agard, D. A.; Cheng, Y.; Fraser, J. S.; Frost, A.; Jura, N.; Kortemme, T.; Manglik, A.; Southworth, D. R.; Stroud, R. M.; Alessi, D. R.; Davies, P.; Frieman, M. B.; Ideker, T.; Abate, C.; Jouvenet, N.; Kochs, G.; Shoichet, B.; Ott, M.; Palmarini, M.; Shokat, K. M.; García-Sastre, A.; Rassen, J. A.; Grosse, R.; Rosenberg, O. S.; Verba, K. A.; Basler, C. F.; Vignuzzi, M.; Peden, A. A.; Beltrao, P.; Krogan, N. J.; et al. Comparative Host-Coronavirus Protein Interaction Networks Reveal Pan-Viral Disease Mechanisms. *Science* **2020**, *370* (6521), eabe9403.

(54) Gavin, A.-C.; Bösche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A.-M.; Cruciat, C.-M.; Remor, M.; Höfert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M.-A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. Functional Organization of the Yeast Proteome by Systematic Analysis of Protein Complexes. *Nature* **2002**, *415* (6868), 141−147.

(55) Collins, S. R.; Miller, K. M.; Maas, N. L.; Roguev, A.; Fillingham, J.; Chu, C. S.; Schuldiner, M.; Gebbia, M.; Recht, J.; Shales, M.; Ding, H.; Xu, H.; Han, J.; Ingvarsdottir, K.; Cheng, B.; Andrews, B.; Boone, C.; Berger, S. L.; Hieter, P.; Zhang, Z.; Brown, G. W.; Ingles, C. J.; Emili, A.; Allis, C. D.; Toczyski, D. P.; Weissman, J. S.; Greenblatt, J. F.; Krogan, N. J. Functional Dissection of Protein Complexes Involved in Yeast Chromosome Biology Using a Genetic Interaction Map. *Nature* **2007**, *446* (7137), 806−810.

(56) Krogan, N. J.; Cagney, G.; Yu, H.; Zhong, G.; Guo, X.; Ignatchenko, A.; Li, J.; Pu, S.; Datta, N.; Tikuisis, A. P.; Punna, T.; Peregrín-Alvarez, J. M.; Shales, M.; Zhang, X.; Davey, M.; Robinson, M. D.; Paccanaro, A.; Bray, J. E.; Sheung, A.; Beattie, B.; Richards, D. P.; Canadien, V.; Lalev, A.; Mena, F.; Wong, P.; Starostine, A.; Canete, M. M.; Vlasblom, J.; Wu, S.; Orsi, C.; Collins, S. R.; Chandran, S.; Haw, R.; Rilstone, J. J.; Gandi, K.; Thompson, N. J.; Musso, G.; St Onge, P.; Ghanny, S.; Lam, M. H. Y.; Butland, G.; Altaf-Ul, A. M.; Kanaya, S.; Shilatifard, A.; O'Shea, E.; Weissman, J. S.; Ingles, C. J.; Hughes, T. R.; Parkinson, J.; Gerstein, M.; Wodak, S. J.; Emili, A.; Greenblatt, J. F. Global Landscape of Protein Complexes in the Yeast Saccharomyces Cerevisiae. *Nature* **2006**, *440* (7084), 637−643.

(57) Huttlin, E. L.; Ting, L.; Bruckner, R. J.; Gebreab, F.; Gygi, M. P.; Szpyt, J.; Tam, S.; Zarraga, G.; Colby, G.; Baltier, K.; Dong, R.; Guarani, V.; Vaites, L. P.; Ordureau, A.; Rad, R.; Erickson, B. K.; Wühr, M.; Chick, J.; Zhai, B.; Kolippakkam, D.; Mintseris, J.; Obar, R. A.; Harris, T.; Artavanis-Tsakonas, S.; Sowa, M. E.; De Camilli, P.; Paulo, J. A.; Harper, J. W.; Gygi, S. P. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **2015**, *162* (2), 425−440.

(58) Schweppe, D. K.; Huttlin, E. L.; Harper, J. W.; Gygi, S. P. BioPlex Display: An Interactive Suite for Large-Scale AP-MS Protein-Protein Interaction Data. *J. Proteome Res.* **2018**, *17* (1), 722−726.

(59) Huttlin, E. L.; Bruckner, R. J.; Navarrete-Perea, J.; Cannon, J. R.; Baltier, K.; Gebreab, F.; Gygi, M. P.; Thornock, A.; Zarraga, G.; Tam, S.; Szpyt, J.; Gassaway, B. M.; Panov, A.; Parzen, H.; Fu, S.; Golbazi, A.; Maenpaa, E.; Stricker, K.; Guha Thakurta, S.; Zhang, T.; Rad, R.; Pan, J.; Nusinow, D. P.; Paulo, J. A.; Schweppe, D. K.; Vaites, L. P.; Harper, J. W.; Gygi, S. P. Dual Proteome-Scale Networks Reveal Cell-Specific Remodeling of the Human Interactome. *Cell* **2021**, *184* (11), 3022−3040.

(60) Ramage, H. R.; Kumar, G. R.; Verschueren, E.; Johnson, J. R.; Von Dollen, J.; Johnson, T.; Newton, B.; Shah, P.; Horner, J.; Krogan, N. J.; Ott, M. A Combined Proteomics/Genomics Approach Links Hepatitis C Virus Infection with Nonsense-Mediated mRNA Decay. *Mol. Cell* **2015**, *57* (2), 329−340.

(61) Li, M.; Johnson, J. R.; Truong, B.; Kim, G.; Weinbren, N.; Dittmar, M.; Shah, P. S.; Von Dollen, J.; Newton, B. W.; Jang, G. M.; Krogan, N. J.; Cherry, S.; Ramage, H. Identification of Antiviral Roles for the Exon-Junction Complex and Nonsense-Mediated Decay in Flaviviral Infection. *Nat. Microbiol.* **2019**, *4* (6), 985−995.

(62) Davis, Z. H.; Verschueren, E.; Jang, G. M.; Kleffman, K.; Johnson, J. R.; Park, J.; Von Dollen, J.; Maher, M. C.; Johnson, T.; Newton, W.; Jäger, S.; Shales, M.; Horner, J.; Hernandez, R. D.; Krogan, N. J.; Glaunsinger, B. A. Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes. *Mol. Cell* **2015**, *57* (2), 349−360.

(63) Batra, J.; Hultquist, J. F.; Liu, D.; Shtanko, O.; Von Dollen, J.; Satkamp, L.; Jang, G. M.; Luthra, P.; Schwarz, T. M.; Small, G. I.; Arnett, E.; Anantpadma, M.; Reyes, A.; Leung, D. W.; Kaake, R.; Haas, P.; Schmidt, C. B.; Schlesinger, L. S.; LaCount, D. J.; Davey, R. A.; Amarasinghe, G. K.; Basler, C. F.; Krogan, N. J. Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication. *Cell* **2018**, *175* (7), 1917−1930.

(64) Swaney, D. L.; Ramms, D. J.; Wang, Z.; Park, J.; Goto, Y.; Soucheray, M.; Bhola, N.; Kim, K.; Zheng, F.; Zeng, Y.; McGregor, M.; Herrington, K. A.; O'Keefe, R.; Jin, N.; VanLandingham, N. K.; Foussard, H.; Von Dollen, J.; Bouhaddou, M.; Jimenez-Morales, D.; Obernier, K.; Kreisberg, J. F.; Kim, M.; Johnson, D. E.; Jura, N.; Grandis, J. R.; Gutkind, J. S.; Ideker, T.; Krogan, N. J. A Protein Network Map of Head and Neck Cancer Reveals PIK3CA Mutant Drug Sensitivity. *Science* **2021**, *374* (6563), eabf2911.

(65) Wang, B.; Vartak, R.; Zaltsman, Y.; Naing, Z. Z. C.; Hennick, K. M.; Polacco, B. J.; Bashir, A.; Eckhardt, M.; Bouhaddou, M.; Xu, J.; Sun, N.; Lasser, M.; Zhou, Y.; Guiley, K. Z.; Chan, U.; Kaye, J. A.; Khare, P.; Drake, S.; Drury, V.; Burke, D. F.; Gonzalez, S.; Alkhairy, S.; Thomas, R.; Lam, S.; Morris, M.; Bader, E.; Seyler, M.; Baum, T.; Krasnoff, R.; Wang, S.; Pham, P.; Arbalaez, J.; Pratt, D.; Chag, S.; Rolland, T.; Bourgeron, T.; Finkbeiner, S.; Bandyopadhay, S.; Ideker, T.; Beltrao, P.; Moohmood, N.; Willsey, H. R.; Obernier, K.; Nowakowski, T. J.; Hüttenhain, R.; State, M. W.; Willsey, A. J.; Krogan, N. J. A Foundational Atlas of Autism Protein Interactions Reveals Molecular Convergence. *bioRxiv* **2023**, n/a.

(66) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.

(67) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.;



https://doi.org/10.1021/acs.jproteome.3c00892
J. Proteome Res. 2024, 23, 2723−2732

Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373* (6557), 871−876.

(68) Akdel, M.; Pires, D. E. V.; Pardo, E. P.; Jänes, J.; Zalevsky, A. O.; Mészáros, B.; Bryant, P.; Good, L. L.; Laskowski, R. A.; Pozzati, G.; Shenoy, A.; Zhu, W.; Kundrotas, P.; Serra, V. R.; Rodrigues, C. H. M.; Dunham, A. S.; Burke, D.; Borkakoti, N.; Velankar, S.; Frost, A.; Basquin, J.; Lindorff-Larsen, K.; Bateman, A.; Kajava, A. V.; Valencia, A.; Ovchinnikov, S.; Durairaj, J.; Ascher, D. B.; Thornton, J. M.; Davey, N. E.; Stein, A.; Elofsson, A.; Croll, T. I.; Beltrao, P. A Structural Biology Community Assessment of AlphaFold2 Applications. *Nat. Struct. Mol. Biol.* **2022**, *29* (11), 1056−1067.

(69) Krapp, L. F.; Abriata, L. A.; Cortés Rodriguez, F.; Dal Peraro, M. PeSTo: Parameter-Free Geometric Deep Learning for Accurate Prediction of Protein Binding Interfaces. *Nat. Commun.* **2023**, *14* (1), 2175.

(70) Holm, L.; Sander, C. Dali: A Network Tool for Protein Structure Comparison. *Trends Biochem. Sci.* **1995**, *20* (11), 478−480.

(71) Ye, Y.; Godzik, A. Flexible Structure Alignment by Chaining Aligned Fragment Pairs Allowing Twists. *Bioinforma. Oxf. Engl.* **2003**, *19* (Suppl. 2), ii246−ii255.

(72) Zhang, Y. TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score. *Nucleic Acids Res.* **2005**, *33* (7), 2302−2309.

(73) Zhang, C.; Shine, M.; Pyle, A. M.; Zhang, Y. US-Align: Universal Structure Alignments of Proteins, Nucleic Acids, and Macromolecular Complexes. *Nat. Methods* **2022**, *19* (9), 1109−1115.

(74) van Kempen, M.; Kim, S. S.; Tumescheit, C.; Mirdita, M.; Lee, J.; Gilchrist, C. L. M.; Söding, J.; Steinegger, M. Fast and Accurate Protein Structure Search with Foldseek. *Nat. Biotechnol* **2023**, n/a.



https://doi.org/10.1021/acs.jproteome.3c00892
*J. Proteome Res.* 2024, 23, 2723−2732