

Comprehensive Overview of Bottom-Up Proteomics Using Mass Spectrometry

Yuming Jiang, Devasahayam Arokia Balaya Rex, Dina Schuster, Benjamin A. Neely, Germán L. Rosano, Norbert Volkmar, Amanda Momenzadeh, Trenton M. Peters-Clarke, Susan B. Egbert, Simion Kreimer, Emma H. Doud, Oliver M. Crook, Amit Kumar Yadav, Muralidharan Vanuopadath, Adrian D. Hegeman, Martín L. Mayta, Anna G. Duboff, Nicholas M. Riley, Robert L. Moritz, and Jesse G. Meyer*



Cite This: *ACS Meas. Sci. Au* 2024, 4, 338–417



Read Online

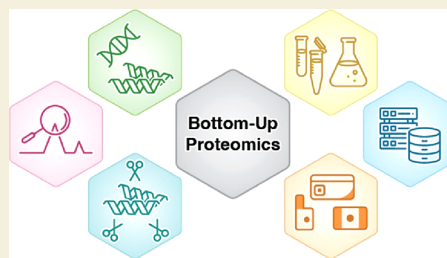
ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Proteomics is the large scale study of protein structure and function from biological systems through protein identification and quantification. “Shotgun proteomics” or “bottom-up proteomics” is the prevailing strategy, in which proteins are hydrolyzed into peptides that are analyzed by mass spectrometry. Proteomics studies can be applied to diverse studies ranging from simple protein identification to studies of proteoforms, protein-protein interactions, protein structural alterations, absolute and relative protein quantification, post-translational modifications, and protein stability. To enable this range of different experiments, there are diverse strategies for proteome analysis. The nuances of how proteomic workflows differ may be challenging to understand for new practitioners. Here, we provide a comprehensive overview of different proteomics methods. We cover from biochemistry basics and protein extraction to biological interpretation and orthogonal validation. We expect this Review will serve as a handbook for researchers who are new to the field of bottom-up proteomics.

KEYWORDS: *Proteomics, liquid chromatography, review, mass spectrometry, tutorial, proteins, peptides, LC-MS*



INTRODUCTION

Proteomics is the large-scale study of protein structure and function. The term “proteomics” is thought to have been coined by Marc R. Wilkins. Proteins are translated from messenger RNA (mRNA) transcripts that are transcribed from the complementary DNA-based genome. Although the genome encodes potential cellular functions and states, the study of proteins in all their forms is necessary to truly understand biology.

Currently, proteomics can be performed with various methods. Mass spectrometry has emerged within the past few decades as the premier tool for comprehensive proteome analysis. The ability of mass spectrometry (MS) to detect charged chemicals enables the identification of peptide sequences and modifications for diverse biological investigations. Alternative (commercial) methods based on affinity interactions of antibodies or DNA aptamers have been developed, namely Olink and SomaScan. There are also nascent methods that are either recently commercialized or still under development and not yet applicable to whole proteomes, such as motif scanning using antibodies, variants of N-terminal degradation, and nanopores.^{1–4} Another approach uses parallel immobilization of peptides with total internal reflection microscopy and sequential Edman degradation.⁵ However, by far the most common method for proteomics is

based on mass spectrometry coupled to liquid chromatography (LC).

Modern proteomics had its roots in the early 1980s with the analysis of peptides by mass spectrometry and low efficiency ion sources. One pioneer in the field was Don Hunt, who described sequencing of peptides using tandem mass spectrometry after chemical ionization with isobutane in 1981.⁶ Another pioneer was Klaus Biemann, who for example worked with Brad Gibson to report peptide identification from fast atom bombardment.⁷ Progress started ramping up around the year 1990 with the introduction of soft ionization methods that enabled, for the first time, efficient transfer of large biomolecules into the gas phase without destroying them.^{8,9} Shortly afterward, the first computer algorithm for matching peptides to a database was introduced.¹⁰ Another major milestone that allowed identification of over 1,000 proteins were improvements to chromatography upstream of MS

Received: November 15, 2023

Revised: May 3, 2024

Accepted: May 3, 2024

Published: June 4, 2024



analysis.¹¹ As the volume of data exploded, methods for statistical analysis transitioned from the wild west of ad hoc empirical analysis to modern informatics based on statistical models¹² and false discovery rate.¹³

Two strategies of mass spectrometry-based proteomics differ fundamentally by whether proteins are analyzed as a whole chain or cleaved into peptides before analysis: “top-down” versus “bottom-up”. Bottom-up proteomics (also referred to as shotgun proteomics) is defined by the intentional hydrolysis of proteins into peptide pieces using enzymes called proteases.¹⁴ Therefore, bottom-up proteomics does not actually measure proteins, but instead infers protein presence and abundance from identified peptides.¹² Sometimes, proteins are inferred from only one peptide sequence representing a small fraction of the total protein sequence predicted from the genome. In contrast, top-down proteomics attempts to measure intact proteins.^{15–18} The potential benefit of top-down proteomics is the ability to measure the many varied proteoforms.^{16,19,20} However, due to myriad analytical challenges, the depth of protein coverage that is achievable by top-down proteomics is considerably less than that of bottom-up proteomics.²¹

In this tutorial we focus on the bottom-up proteomics workflow. The most common version of this workflow is generally composed of the following steps. First, proteins in a biological sample must be extracted. Usually this is achieved by mechanically lysing cells or tissue while denaturing and solubilizing the proteins and disrupting DNA to minimize interference in analysis procedures. Next, proteins are hydrolyzed into peptides, most often using the protease trypsin, which generates peptides with basic C-terminal amino acids (arginine and lysine) to aid in fragment ion series production during tandem mass spectrometry (MS/MS). Peptides can also be generated by chemical reactions that induce residue specific hydrolysis, such as cyanogen bromide that cleaves after methionine. Peptides from proteome hydrolysis must be purified; this is often accomplished with reversed-phase liquid chromatography (RPLC) cartridges or tips to remove interfering molecules in the sample such as salts and buffers. The peptides are then almost always separated by reversed-phase LC before they are ionized and introduced into a mass spectrometer, although recent reports also describe LC-free proteomics by direct infusion.^{22–24} The mass spectrometer then collects precursor and fragment ion data from those peptides. Peptides must be identified from the tandem mass spectra, protein groups are inferred from a proteome database, and then quantitative values are assigned. Changes in protein abundances across conditions are determined with statistical tests, and results must be interpreted in the context of the relevant biology. Data interpretation is the rate limiting step; data collected in less than one week can take months or years to understand.

The diversity of experimental goals that are achievable with proteomics technology drives an expansive array of workflows. Every choice is important as every choice will affect the results, from instrument procurement to choice of data processing software and everything in between. In this tutorial, we detail all the required steps to serve as a comprehensive overview for new proteomics practitioners.

Following the Introduction, There are 17 sections in total:

1. Biochemistry Basics
2. Types of Experiments
3. Protein Extraction

4. Proteolysis
5. Peptide Quantification Methods
6. Enrichments
7. Peptide Purification
8. Liquid Chromatography
9. Peptide Ionization
10. Mass Spectrometry
11. Peptide Fragmentation (MS/MS)
12. Data Acquisition
13. Raw Data Analysis
14. Protein Databases
15. Proteomics Knowledge Bases
16. Biological Interpretation
17. Orthogonal Validation Experiments

BIOCHEMISTRY BASICS

Proteins

Proteins are large biomolecules or biopolymers made up of a backbone of amino acids which are linked by peptide bonds. They perform various functions in living organisms ranging from structural roles to functional involvement in cellular signaling and the catalysis of chemical reactions (enzymes). Proteins are made up of 20 different amino acids (not counting pyrrolysine, hydroxyproline, and selenocysteine, which only occur in specific organisms) and their sequence is encoded in their corresponding genes. The human genome encodes approximately 19,778 of the predicted canonical proteins coded in the human genome (see www.neXtProt.org).²⁵ Each protein is present at a different abundance depending on the cell type or bodily fluid. Previous studies have shown that the concentration range of proteins can span at least seven orders of magnitude to up to 20,000,000 copies per cell, and that their distribution is tissue-specific.^{26,27} Protein abundances can span more than ten orders of magnitude in human blood, while a few proteins make up most of the protein by weight in these fluids, making blood and plasma proteomics one of the most challenging matrices for mass spectrometry to analyze. Due to genetic variation, alternative splicing, and co- and post-translational modifications (PTMs), multiple different proteoforms can be produced from a single gene (Figure 1).^{16,28}

PTMs

After protein biosynthesis, enzymatic and nonenzymatic processes change the protein sequence through proteolysis or covalent chemical modification of amino acid side chains. Post-translational modifications (PTMs) are important biological regulators contributing to the diversity and function of the cellular proteome. Proteins can be post-translationally modified through enzymatic and non-enzymatic reactions *in vivo* and *in vitro*.²⁹ PTMs can be reversible or irreversible, and they change protein function in multiple ways, for example by altering substrate–enzyme interactions, subcellular localization or protein–protein interactions.^{30,31}

More than 400 biological PTMs have been discovered in both prokaryotic and eukaryotic cells. There are many more chemical artifact PTMs that occur during sample preparation, such as carbamylation. Biological modifications are crucial in controlling protein functions and signal transduction pathways.³² The most commonly studied and biologically relevant post-translational modifications include phosphorylation (Ser, Thr, Tyr, His), glycosylation (Arg, Asp, Cys, Ser, Thr, Tyr, Trp), disulfide bonds (Cys-Cys), ubiquitination (Lys, Cys, Ser,

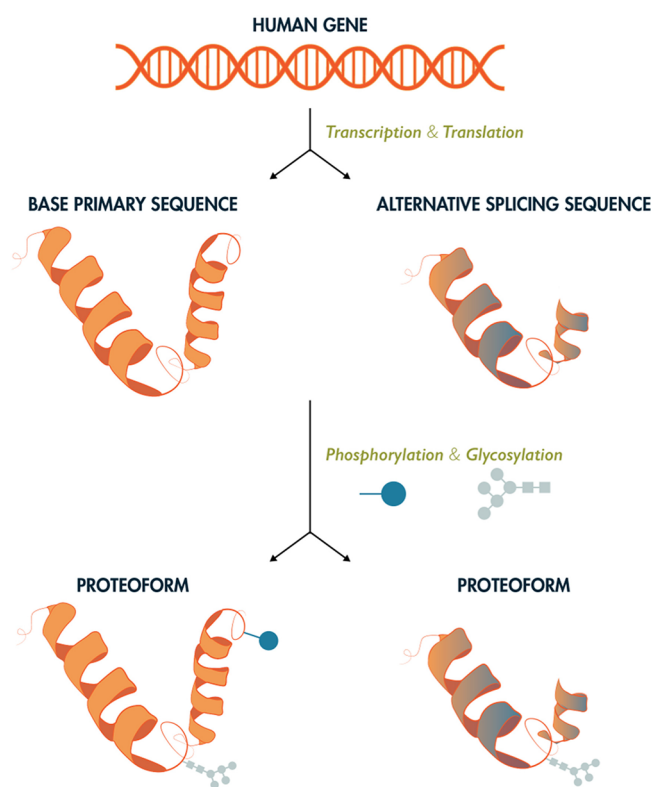


Figure 1. Proteome complexity. Each gene may be expressed in the form of multiple protein products, or proteoforms, through alternative splicing and incorporation of post-translational modifications. As such, there are many more unique proteoforms than genes. While there exist 20,000–23,000 coding genes in the human genome, upwards of 1,000,000 unique human proteoforms may exist. The study of the structure, function, and spatial and temporal regulation of these proteins is the subject of mass spectrometry-based proteomics

Thr, N-term), succinylation (Lys), methylation (Arg, Lys, His, Glu, Asn, Cys), oxidation (especially Met, Trp, His, Cys), acetylation (Lys, N-term), and lipidation.³³

Protein PTMs can alter function, activity, structure, spatiotemporal status and interaction with proteins or small molecules. PTMs alter signal transduction pathways and gene expression control³⁴ and regulation of apoptosis^{35,36} by phosphorylation. Ubiquitination generally regulates protein degradation,³⁷ SUMOylation regulates chromatin structure, DNA repair, transcription, and cell-cycle progression,^{38,39} and palmitoylation regulates the maintenance of the structural organization of exosome-like extracellular vesicle membranes.⁴⁰ Glycosylation is a ubiquitous modification that regulates various T cell functions, such as cellular migration, T cell receptor signaling, cell survival, and apoptosis.^{41,42} Deregulation of PTMs is linked to cellular stress and diseases.⁴³

Several non-MS methods exist to study PTMs, including in vitro PTM reaction tests with colorimetric assays, radioactive isotope-labeled substrates, western blot with PTM-specific antibodies and superbinders, and peptide and protein arrays.^{44–46} While effective, these approaches have many limitations, such as inefficiency and difficulty in producing pan-specific antibodies. MS-based proteomics approaches are currently the predominant tool for identifying and quantifying changes in PTMs.

■ TYPES OF EXPERIMENTS

A wide range of questions are addressable with proteomics technology, which translates to a wide range of variations of proteomics workflows. In some workflows, the identification of proteins in a given sample is desired. For other experiments, the quantification of as many proteins as possible is essential for the success of the study. Therefore, proteomic experiments can be both qualitative and quantitative. The following sections give an overview of several common proteomics experiments.

Protein Abundance Changes

A common experiment is a discovery-based, unbiased mapping of proteins along with detection of changes in their abundance across sample groups. This is achieved using methods such as label free quantification (LFQ) or isobaric tagging, which are described in more detail in subsequent sections. In these experiments, data should be collected from at least three biological replicates of each condition to estimate the variance of measuring each protein. Depending on the experiment design, different statistical tests are used to calculate changes in measured protein abundances between groups. If there are only two groups, the quantities might be compared with a *t*-test or with a Wilcoxon signed-rank test. The latter is a non-parametric version of the Student's *t*-test. If there are more than two sample groups, then Analysis of Variance (ANOVA) is used instead, followed by a post-hoc test such as Tukey's honestly significant difference test to discover pairwise differences. With either testing scheme, the *p*-values from the first set of tests must be corrected for multiple testing. A common method for *p*-value correction is the Benjamini-Hochberg method.⁴⁷ These types of experiments have revealed wide ranges of proteomic remodeling from various biological systems.

PTMs

Proteins may become decorated with various chemical modifications during or after translation,³³ or through proteolytic cleavage such as N-terminal methionine removal.⁴⁸ Several proteomics methods are available to detect and quantify each specific type of modification. See also the section on Protein/Peptide [Enrichment and Depletion](#). For a good online resource listing potential modifications, sites of attachment, and their mass differences, the website www.unimod.org is an excellent curated and freely accessible database.

Phosphoproteomics. Phosphoproteomics is the study of protein phosphorylation, wherein a phosphate group is covalently attached to a protein side-chain (most commonly serine, threonine, or tyrosine). Although western blotting can measure one phosphorylation site at a time (if using a monoclonal antibody), mass spectrometry-based proteomics can measure thousands of sites from a sample at the same time. After proteolysis of the proteome, to achieve competitive coverage of the phosphoproteome, phosphopeptides need to be enriched to be detected by mass spectrometry. Various methods of enrichment have been developed.^{49–52}

A key challenge of phosphoproteomics is the limit of detection. It is important to ensure that there is a sufficient amount of protein before conducting a phosphoproteomics project because phosphorylated proteins and peptides may represent only ~1% of the total protein. Many phosphoproteomics workflows start with at least 1 mg of total protein per sample. In addition to low stoichiometry, phosphorylation is very labile, and for this reason, great care must be taken in the

collection and storage of samples for phosphoproteomics, where proteome denaturation should be rapid and aggressive while including phosphatase inhibitors. Newer, more sensitive instrumentation is enabling detection of protein phosphosites from much less material, down to the nanogram-level of peptide loading on the LC-MS system. Despite advancement in phosphoproteomics technology, the following challenges still exist: limited sample amounts, highly complex samples, and wide dynamic range.⁵³ Additionally, phosphoproteomic analysis is often time-consuming and requires the use of expensive equipment such as enrichment kits.

See the Peptide/Protein [Enrichment and Depletion](#) section for more details.

Glycoproteomics. The importance of protein glycosylation in health and disease has been known for a long time, but do to high analytical difficulty, only recently has their large scale analysis been gaining traction. Protein glycosylation sites can be N-linked (asparagine-linked) or O-linked (serine/threonine-linked). Understanding the function of protein glycosylation will help us understand numerous biological processes since this is a universal protein modification across all domains of life, especially at the cell surface.^{54–57}

Studies of phosphorylation and glycosylation share several experimental pipeline steps including sample preparation. Protein clean-up approaches for glycoproteomics may differ from other proteomics experiments because glycopeptides are more hydrophilic than most peptides. Some approaches mentioned in the literature include: filter-aided sample preparation (FASP), suspension traps (S-traps), and protein aggregation capture (PAC).^{54,58–63} Multiple proteases may be used to increase the sequence coverage and detect more modification sites, such as: trypsin, chymotrypsin, pepsin, WaLP/MaLP,⁶⁴ GluC, AspN, pronase, proteinase K, OgpA, StcEz, BT4244, AM0627, AM1514, AM0608, Pic, ZmpC, CpaA, IMPa, PNGase F, Endo F, Endo H, and OglyZOR.⁵⁴ Mass spectrometry has improved over the past decade, and now many strategies are available for glycoprotein structure elucidation and glycosylation site quantification.⁵⁴ See also the section on “[Aminoxymethyl TMT Isobaric Mass Tags](#)” as an example quantitative glycoproteomics method.

Structural Techniques

Almost all proteins (except for intrinsically disordered proteins⁶⁵) fold into three-dimensional (3D) structures either by themselves or assisted by molecular chaperones.⁶⁶ There are four levels relevant to the folding of any protein:

- Primary structure: The protein's linear amino acid sequence, with amino acids connected through peptide bonds.
- Secondary structure: The amino acid chain's folding: α -helix, β -sheet, or turn.
- Tertiary structure: The three-dimensional structure of the protein.
- Quaternary structure: The structure of several protein molecules/subunits in one complex.

Of recent note, the development of AlphaFold, has enabled the high-accuracy three-dimensional structural prediction of all human proteins and for proteins of many other species, enabling a more thorough study of protein folding and is used to predict the relationship between fold and function.^{67,68} Several proteomics methods have been developed to reveal protein structure information for simple and complex systems.

Cross-Linking Mass Spectrometry (XL-MS). XL-MS is an emerging technology in the field of proteomics. It can be used to determine changes in protein-protein interactions and/or protein structure. XL-MS covalently locks interacting proteins together to preserve interactions and proximity during MS analysis. XL-MS is different from traditional MS in that it requires the identification of chimeric MS/MS spectra from cross-linked peptides.^{69,70} XL-MS can be used to gain structural constraints in purified protein systems or at the whole proteome scale.

The common steps in a XL-MS workflow are as follows:⁷¹

1. Generate a system with protein-protein interactions of interest (*in vitro* or *in vivo*)⁷²
2. Add a cross-linking reagent to covalently connect adjacent protein regions (such as disuccinimidyl sulfoxide, DSSO)⁷⁰
3. Proteolysis to produce peptides
4. MS/MS data collection
5. Identify cross-linked peptide pairs using special software (i.e. pLink,⁷³ Kojak,^{74,75} xQuest,⁷⁶ XlinkX⁷⁷)
6. Generate cross-link maps for structural modeling and visualization^{78,79} (optional: 7. Use detected cross-links for protein-protein docking⁸⁰)

Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS). HDX-MS works by detecting changes in peptide mass due to exchange of amide hydrogens of the protein backbone with deuterium from D₂O.⁸¹ The exchange rate depends on the protein solvent accessible surface area, dynamics, and the properties of the amino acid sequence.^{81–84} Although using D₂O to make deuterium-labeled samples is simple, HDX-MS requires several controls to ensure that experimental conditions capture the dynamics of interest.^{81,85–87} If the peptide dissociation process is tuned appropriately, residue-level quantification of changes in solvent accessibility are possible within a measured peptide.⁸⁸ HDX can produce precise protein structure measurements with high reproducibility. Masson *et al.* gave recommendations on how to prep samples, conduct data analysis, and present findings in a detailed stepwise manner.⁸¹

Radical Footprinting. This technique uses hydroxyl radical footprinting and MS to elucidate protein structures, assembly, and interactions within a large macromolecule.^{89,90} In addition to proteomics applications, various approaches to make hydroxide radicals have also been applied for footprinting studies in nucleic acid/ligand interactions.^{91–93} This chapter is very useful in learning more about this topic.⁹⁴

There are several methods of producing radicals for protein footprinting:

1. Fenton and Fenton-like Chemistry^{89,95,96}
2. Electron-Pulse Radiolysis^{89,97}
3. High-Voltage Electrical Discharge^{89,98}
4. Synchrotron X-ray Radiolysis of Water^{89,99}
5. Plasma Formation of OH Radicals^{89,100}
6. Photolysis of Hydrogen Peroxide^{89,101}

Fast Photochemical Oxidation of Proteins (FPOP) (ref 102). FPOP is an example of a radical footprinting method. In FPOP, a laser-based hydroxyl radical protein footprinting MS method that relies on the irreversible labeling of solvent-exposed amino acid side chains by hydroxyl radicals in order to understand structure of proteins. A laser produces 248 nm light that causes hydrogen peroxide to break into a pair of hydroxyl radicals.^{101,103} The flow rate of solution through the capillary and laser frequency are adjusted such that each protein molecule is irradiated only once. After they are irradiated, the

sample is collected in a tube that contains catalase and free methionine in the buffer, quenching the H_2O_2 and hydroxyl radicals and preventing secondary modification of residues that become exposed due to unfolding after the initial labeling. Control samples are made by running the sample through the flow system without any irradiation. Another experimental control involves the addition of a radical scavenger to tune the extent of protein oxidation.^{104,105} FPOP has wide application for proteins including measurements of fast protein folding and transient dynamics.

Protein Painting (refs 106, 107). Protein painting uses “molecular paints” to noncovalently coat the solvent-accessible surface of proteins. Chemically, these paints may be small aryl hydrocarbon dyes with fast on-rates with very slow off-rates.¹⁰⁶ These paint molecules will coat the protein surfaces but will not have access to the hydrophobic cores or protein-protein interface regions that solvents cannot access. If the “paint” covers free amines of lysine side chains, the “painted” parts will be protected from trypsin cleavage. After proteolysis, the peptides samples will be subjected to MS. A lack of proteolysis in a region is interpreted as solvent accessibility, which gives rough structural information about complex protein mixtures or even a whole proteome.

LiP-MS (Limited Proteolysis Mass Spectrometry) (refs 108–110). Limited proteolysis coupled to mass spectrometry (LiP-MS) is a method that tracks structural changes in complex proteomes in response to a variety of perturbations or stimuli. The underlying tenet of LiP-MS is that a stimuli-induced change in native protein structure (i.e. protein-protein interaction, introduction of a PTM, ligand/substrate binding, or changes in osmolarity or ambient temperature) can be detected by a change in accessibility of a broad-specificity protease (i.e. proteinase K) to the region(s) of the protein where the structural change occurs. For example, small molecule binding may render a disordered region protected from non-specific proteolysis by directly blocking access of the protease to the cleavage site. LiP-MS can therefore provide a somewhat unbiased view of structural changes at the proteome scale. Importantly, LiP-MS necessitates cell lysates or individual proteins be maintained in their native state prior to or during perturbation and protease treatment. LiP-MS can also be applied to membrane suspensions, to facilitate the study of membrane proteins without the need for purification or detergents.¹¹¹ For additional information about LiP-MS, please refer to the following article: ref 112.

Protein Stability and Small Molecule Binding

Cellular Thermal Shift Assay (CETSA) (refs 113, 114). CETSA obtains each detectable protein’s stability curve from the soluble fraction remaining after a range of incubation temperatures. This is done in live cells immediately before lysis, or in non-denaturing lysates. The original paper reported this method using immunoaffinity approaches for detecting changes in soluble protein. The assay is capable of detecting shifts in the thermal equilibrium of cellular proteins in response to a variety of perturbations, but most commonly in response to *in vitro* drug treatments.

Thermal Proteome Profiling (TPP) (refs 115–118). Thermal proteome profiling (TPP) follows the same principle as CETSA, but has been extended to use an unbiased mass spectrometry readout of many proteins. During a typical TPP experiment, a protein sample is first treated with a vehicle control or drug of interest to stabilize protein-ligand

interactions. The sample is then divided into multiple aliquots, which are subjected to different temperatures to induce thermal denaturation. The resulting drug-induced changes in protein stability curves are detected using mass spectrometry. By comparing protein stability curves across the temperatures between treatment conditions, TPP can provide insight into the proteins that bind a ligand.

Protein–Protein Interactions (PPIs)

Affinity Purification Coupled to Mass Spectrometry (AP-MS) (refs 119–121). AP-MS is an approach that involves enrichment of a target protein or protein complex using an antibody with specificity toward a protein of interest followed by mass spectrometry to identify the interacting proteins. If there are no good antibodies for immunoprecipitation of a protein of interest, it may be genetically tagged with an affinity epitope, such as a FLAG or hemagglutinin, which is used to selectively capture the target protein using an antibody against that epitope. In either case, the protein complex is then purified from the sample using a series of wash steps, and the interacting proteins are identified using mass spectrometry. The success of AP-MS experiments depends on many factors, including the quality of the antibody used for purification, the specificity and efficiency of the resin used for capture, and the sensitivity and resolution of the mass spectrometer. In addition, careful experimental design and data analysis are critical for accurately identifying and interpreting protein-protein interactions.

AP-MS has been used to study a wide range of biological processes, including signal transduction pathways, protein complex dynamics, and protein post-translational modifications. AP-MS has been performed on a whole proteome scale as part of the BioPlex project.^{122–124}

Despite its widespread use, AP-MS has some limitations, including non-specific interactions, the difficulty in interpreting complex data sets, and the possibility of missing important interacting partners due to constraints in sensitivity or specificity. However, with continued advances in technology and data analysis methods, AP-MS is likely to remain a valuable tool for studying protein-protein interactions.

There are other variants of this experiment where instead of an antibody against the protein of interest, the protein of interest can itself be conjugated to a solid phase by expression and purification with a his-tag or fusion to a glutathione s-transferase (GST) domain. These approaches may be useful when good antibodies for IP are not available.

The interaction of any two proteins depends both on their concentrations and their affinity for each other; two proteins could have low affinity for each other, but if present at high concentrations, they will be found together in AP-MS. Therefore, key considerations for AP-MS studies are to include negative control antibodies to help distinguish true interactions from background, and including many replicates to assess reproducibility.

APEX Peroxidase (refs 125, 126). APEX-MS is a labeling technique that utilizes a peroxidase genetically fused to a protein of interest. When biotin-phenol is transiently added in the presence of hydrogen peroxide, nearby proteins are covalently biotinylated.¹²⁷ APEX thereby enables the discovery of interacting proteins in living cells. One of the major advantages of APEX is its ability to label proteins in their native environment, allowing for the identification of interactions that occur under physiological conditions. A key

benefit of APEX is that it can detect transient or weak interactors, unlike AP-MS that detects strong and stable interactions. Despite its advantages, APEX has some limitations, including the potential for non-specific labeling, the difficulty in distinguishing between direct and indirect interactions, and the possibility of missing interactions that occur at low abundance or in regions of the cell that are not effectively labeled.

Proximity-Dependent Biotin Identification (BioID) (refs 128–131). BioID is a proximity labeling technique that allows for the identification of protein-protein interactions. BioID involves the genetic tagging of a protein of interest with a promiscuous biotin ligase in live cells, which then biotinylates proteins in close proximity to the protein of interest. One of the advantages of BioID is its ability to label proteins in their native environment, allowing for the identification of interactions that occur under physiological conditions. BioID has been used to identify a wide range of protein interactions, including receptor-ligand interactions, signaling complexes, and protein localization. BioID is a slower reaction than APEX and therefore may pick up even more transient interactions that occur on longer timescales. A newer alternative to BioID called TurboID has much higher activity, and is now more commonly used.¹³² BioID has the same limitations as APEX. For more information on BioID, please refer to ref 133.

PROTEIN EXTRACTION

Protein extraction is the initial phase of any mass spectrometry-based proteomics experiment. Protein extraction is sample dependent; a solution that is effective for plasma proteomics may not work well for plant tissue proteomics. Thought should be given to any planned downstream assays, such as specific proteolysis requirements (LiP-MS, PTM enrichments, enzymatic reactions, glycan purification or hydrogen-deuterium exchange experiments), long-term project goals (reproducibility, multiple sample types, low abundance samples), as well as to the experimental question (coverage of a specific protein, subcellular proteomics, global proteomics, protein-protein interactions or affinity enrichment of specific classes of modifications). The 2009 version of *Methods in Enzymology: guide to Protein Purification*¹³⁴ serves as a deep dive into how molecular biologists and biochemists traditionally carried out protein extraction. The *Protein Protocols handbook*¹³⁵ and the excellent review by Linn¹³⁶ are good sources of general proteomics protocols. Another excellent resource is the “*Proteins and Proteomics: A Laboratory Manual*” by Richard J. Simpson.^{137,137} This manual is 926 pages packed full of bench tested protocols and procedures for carrying out protein centric studies. Any change in extraction conditions should be expected to create potential changes in downstream results. Be sure to plan and optimize the protein extraction step first and use a protocol that works for your needs. To reproduce the results of another study, one should begin with the same extraction protocols.

Learning the fundamentals and mechanisms of how and why sample preparation steps are performed is vital because it enables flexibility to perform proteomics from a wide range of samples. For bottom-up proteomics, the overreaching goal is efficient and consistent extraction and digestion. A range of mechanical and non-mechanical extraction protocols have been developed and the choice of technique is generally dictated by sample type or assay requirements (i.e. native versus non-native extraction). Extraction can be aided by the

addition of detergents and/or chaotropes to the sample, but care should be taken that these additives do not interfere with the sample digestion step or downstream mass-spectrometry analysis.

In general, a safe and common choice for standard proteomic protein extraction would be to use 8 M urea in 100 mM Tris, pH 8.5; the pH is based on optimum trypsin activity.¹³⁸ Desalting with StageTips, Waters’ SepPaks, or similar would yield clean peptides. Triton X-100 and NP-40 should be avoided at all costs. The following sections detail the range of choices that are available.

Buffer and Denaturant Choice

General Proteomics. A common question to proteomics core facilities is, “What is the best buffer for protein extraction?” Unfortunately, there is no one correct answer. For global proteomics experiments where maximizing the number of protein or peptide identifications is a goal, 50–100 mM of a neutral pH buffer (pH 7.5–8.5) is often used with a strong denaturant. Relevant factors for buffer choice include cost, volatility, and reactivity such as primary amine containing. Volatile choices like ammonia bicarbonate are desirable because they can be removed by lyophilization. However, ammonium bicarbonate promotes methionine oxidation and we generally suggest Tris instead to minimize oxidation. Tris is desirable due to low cost but can act as a chelator and contains a primary amine, which may be incompatible with some conditions, like TMT labeling. Table 1 summarizes common

Table 1. Common Buffers Used for Proteomic Sample Preparation

Buffer	Notes
Phosphate buffered saline (PBS)	Nonvolatile, inert
Tris(hydroxymethyl)aminomethane (Tris)	Cheap, nonvolatile, primary amine containing
4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	More expensive, nonvolatile
Ammonium bicarbonate	Cheap, volatile, primary amine containing
Triethanolamine bicarbonate	Cheap, volatile, non-primary amine containing

buffers. A great online resource to help calculate buffer compositions and pH values is the website by Robert Beynon at <http://phbuffers.org>. Although there are a range of buffers that can be used to provide the correct working pH and ionic strength, not all buffers are compatible with downstream workflows.

Complete and quick denaturation of proteins in the sample is required to limit changes to protein status by endogenous proteases, kinases, phosphatases, and other enzymes. For this reason, buffers must be used in conjunction with a chaotrope or surfactant to denature and solubilize proteins.^{139,140} The choice of denaturant should be governed by compatibility with the protease (typically trypsin) and peptide cleanup steps must be considered. Table 2 lists common denaturants. Urea is an easy and a common choice because it is compatible with trypsin at <2M and it can be removed by desalting. However, urea induces carbamylation, which is made worse with sample heating.¹⁴¹ If intact protein separations are planned (based on size or isoelectric point), choose a denaturant compatible with those methods, such as sodium dodecyl sulphate (SDS).¹⁴² SDS is a strong denaturant, but it is not compatible with trypsin or reversed phase materials. Sodium deoxycholate

Table 2. Common Denaturants Used for Proteomic Sample Preparation

Denaturant	Notes
8 M Urea	Nonvolatile, chemically reactive, limit heating, must be diluted to <2 M before trypsin addition
1–5% Sodium dodecyl sulfate (SDS)	Cheap, strong denaturation and hydrophobicity, must be removed before trypsin
1–5% Sodium deoxycholate	Hydrophobic for membrane proteins, easy to remove due to precipitation with acid
<i>n</i> -Dodecyl- β -maltoside	expensive, low amounts can be used with trypsin and LC-MS
Triton X-100	Do not use this; if samples already have this or NP-40, proceed with protein precipitation

(SDC) and sodium laurate (SL) are also strong denaturants with the added benefit of compatibility with trypsin. For non-MS workflows, detergents containing poly ethylene glycol tails are common, such as triton X-100. SDS, SDC, SL, and triton X-100 are incompatible with LC-MS workflows as they can cause ion suppression and column clogging. Therefore detergents must be removed before further protein processing. Detergent removal options differ based on the chemistry of the detergent. Alternatively, mass-spectrometry-compatible detergents may be used, such as *n*-dodecyl- β -maltoside (DDM).¹⁴³

Detergent Removal. Relatively low concentrations of some detergents, such as 1% SDC, or chaotropes such as 1M urea, are compatible with proteolysis by trypsin/Lys-C. Often proteolysis-compatible concentrations of these detergents and chaotropes are achieved by diluting the sample in appropriate buffer (i.e. 100 mM ammonium bicarbonate, pH 8.5) after cell or tissue lysis in a higher concentration. However, most detergents should be removed prior to enzymatic hydrolysis. This is generally performed through precipitation of proteins. The most common types are 1) acetone, 2) trichloroacetic acid (TCA), and 3) chloroform/methanol/water (Folch).^{144,145} Proteins are generally insoluble in most pure organic solvents, so cold ethanol or methanol are sometimes used. Pellets should be washed with organic solvent for complete removal of detergents. Alternatively, solid phase based digestion methods such as S-trap (ProtiFi),¹⁴⁶ FASP,^{147,148} SP3,^{149,150} and on column/bead such as protein aggregation capture (PAC)¹⁵¹ can allow for proteins to be applied to a solid phase for detergent removal prior to proteolysis.¹⁵² Specialty detergent removal columns exist (Pierce/Thermo Fisher Scientific) but add expense and time-consuming steps to the process. SDC can be easily removed by precipitation or phase separation¹⁵³ following digestion by acidification of the sample to pH 2–3. Ethyl acetate can also remove several common detergents.¹⁵⁴ Any small-molecule removal protocol should be tested for efficiency prior to implementing in a workflow with many samples as avoiding detergent (or polymer) contamination in the LC/MS is very important.

Protein–Protein Interactions. Denaturing conditions will efficiently extract proteins, but will denature proteins and therefore disrupt most protein-protein interactions. If you are working on an antibody or affinity purification of a specific protein and expect to analyze enzymatic activity, structural features, and/or protein-protein interactions, a non-denaturing lysis buffer should be utilized.^{155,156} Check the calculated isoelectric point (pI) and hydrophobicity (e.g., try the ExPASy.org resource ProtParam) for a good idea of starting pH/conductivity, but a stability screen may be needed. In

general, a good starting point for the buffer will still be close to neutral pH with 50–250 mM NaCl, but specific proteins may require pH as low as 2 or as high as 9 for stable extraction. A low percent of mass spectrometry compatible detergent may also be used, such as *n*-dodecyl- β -maltoside. Newer mass spectrometry-compatible detergents are also useful for protein extraction and ease of downstream processing, including Rapigest (Waters), *N*-octyl- β -glucopyranoside, MS-compatible degradable surfactant (MaSDeS),¹⁵⁷ Azo,¹⁵⁸ PPS silent surfactant,¹⁵⁹ sodium laurate,¹⁶⁰ and sodium deoxycholate.¹⁶¹ Avoid using tween-20, triton-X, NP-40, and polyethylene glycols (PEGs) as these compounds are challenging to remove after digestion.¹⁶²

Optional Additives. There are several additional additives that are often found in protein extraction buffer solutions. Salts like 50–150 mM sodium chloride (NaCl) may be used to mimic physiological ionic strength. Protease, phosphatase and deubiquitinase inhibitors are optional additives in less denaturing conditions or in experiments focused on specific PTMs. For a broad range of inhibitors, a premixed tablet can be added to the lysis buffer, such as Roche cOmplete Mini Protease Inhibitor Cocktail tablets. Protease inhibitors may impact desired proteolysis from the added protease, and will need to be diluted or removed prior to protease addition. To improve extraction of DNA- or RNA-binding proteins, adding a small amount of nuclease or benzonase is useful for degradation of any bound nucleic acids and results in a more consistent digestion.¹⁶³ For non-denaturing buffer conditions, which preserve tertiary and quaternary protein structures, additional additives may still be necessary to prevent proteolysis or PTMs throughout the extraction process.

Extraction of Plant Proteins. Protein extraction from plant tissues generally more challenging due to the presence of cell walls, large vacuoles, and several different classes of interfering substances that are often present in these materials. Cell walls require vigorous disruption techniques such as grinding with or without an abrasive, use of a bead mill, or homogenizers, which, while they do release the cellular contents, also rupture and mix the contents of organelles and other subcellular compartments. As a result, isolation of proteins from organelles or other subcellular fractions of plant materials can be fairly specialized.^{164–166} Plant tissues have lower protein content compared with tissues from other organisms as only a small fraction of the tissue volume is cytoplasm with the apoplast (cell exterior and wall) and vacuole using much of the tissue volume. Depending on tissue, cell type and maturity, a plant cell's vacuole accounts for most of the cell interior space and typically contains substances that degrade or denature proteins upon tissue disruption. Isolation of functional native proteins from plants usually requires use of plant-targeted protease inhibitors,^{167,168} and strategies for preventing protein modification and precipitation by phenolic compounds and their oxidation products¹⁶⁹ in addition to the buffers, reductants, and other additives discussed previously.

Methodology for whole-tissue protein extraction of plants has been extensively reviewed.^{170–173} These procedures avoid the sample degradation by protease or phenol oxidase activities that can plague native plant protein purification by using extraction at low temperature followed by protein denaturation and removal of contaminating compound classes using precipitation strategies. Protein is extracted under denaturing conditions and precipitated using combinations of trichloroacetic acid (TCA), ammonium acetate, or acetone (or other

solvent) precipitation. Initial protein extraction and/or resolubilization of protein precipitates is accomplished using detergents, phenol or other chaotropic agents. Extraction protocols have been shown to influence proteomic results,¹⁷² and the compatibility of extracts with subsequent analytical strategies can vary significantly since protocols that were initially developed for 2D-gel electrophoretic analysis often use detergents that can be problematic for peptide LC-MS/MS proteomic approaches. More recently developed strategies make use of filters (filter-aided sample preparation, FASP¹⁷⁴) or coated magnetic beads (single-pot-solid-phase-enhanced sample preparation, SP3^{175,176}) for higher throughput shotgun proteomic sample preparation. Strategies for overcoming the dynamic range limitations caused by plant-specific hyper-abundant proteins have been developed both for RuBisCO,¹⁷⁷ which makes up ~50% of the protein in green tissues of C3 plants, and also for seed storage proteins.¹⁷⁸

Mechanical or Sonic Disruption

Cell Lysis. Small mammalian cell pellets and exosomes will lyse almost instantly upon addition denaturing buffer. If non-denaturing conditions are desired, osmotic swelling and subsequent shearing or sonication can be applied.¹⁷⁹ Efficiency of extraction and degradation of nucleic acids can be improved using various sonication methods: (1) probe sonicator with ice; (2) water bath sonicator with ice or cooling; (3) bioruptor sonication device; (4) adaptive focused acoustics (AFA).¹⁸⁰ Key to these additional lysis techniques is to keep the temperature of the sample from rising significantly which can cause proteins to aggregate or degrade. Some cell types may require additional force for effective lysis (see below). For cells with cell walls (i.e. bacteria or yeast), lysozyme is often added in the lysis buffer. Any added protein will be present in downstream results, however, so excessive addition of lysozyme is to be avoided unless tagged protein purification will occur.

Tissue/Other Lysis. Although small pieces of soft tissue can often be successfully extracted with the probe and sonication methods described above, larger/harder tissues as well as plants/yeast/fungi are better extracted with some form of additional mechanical force. If proteins are to be extracted from a large amount of sample, such as soil, feces, or other diffuse input, one option is to use a dedicated blender and filter the sample, followed by centrifugation. If samples are smaller, such as tissue, tumors, etc., cryo-homogenization is recommended. The simplest form of this is grinding the sample with liquid nitrogen and a mortar and pestle. Tools such as bead beaters (i.e. FastPrep-24) are also used, where the sample is placed in a tube with appropriately sized glass or ceramics beads and shaken rapidly. Cryo-mills are chambers where liquid nitrogen is applied around a vessel and large bead or beads. Cryo-fractionators homogenize samples in special bags that are frozen in liquid nitrogen and smashed with various degrees of force.¹⁸¹ In addition, rapid bead beating mills such as the Bertin Precellys Evolution are both economical, effective and detergent compatible for many types of proteomics experiments at a scale of 96 samples per batch. Finally, pressure cycling such as the option from pressure biosciences is useful for homogenization of many small tissue pieces.¹⁸² After homogenization, samples can be sonicated by one of the methods above to fragment DNA and increase solubilization of proteins.

Measuring the Efficiency of Protein Extraction

Following protein extraction, samples should be centrifuged (10–14,000g for 10–30 min depending on sample type) to remove debris and insoluble protein prior to determining protein concentration. Protein quantification is important to assess the yield of an extraction procedure, to match the amount of protein per sample, and to adjust the scale of the downstream processing steps to match the amount of protein. For example, when purifying peptides, the amount of sorbent should match the amount of material to be bound. Protein concentration can be calculated using a number of assays or tools.^{183,184} Extraction solution components will need to be compatible with any assay chosen; alternatively, small molecule interferences may be removed (see above) prior to protein concentration calculation. Each method will have inherent bias and error.^{185,186} These methods can be divided into colorimetric or fluorometric assays as follows.

Colorimetry-Based Methods. The method includes different assays like Coomassie Blue G-250 dye binding (the Bradford assay), the Folin-Lowry assay, the bicinchoninic acid (BCA) assay and the biuret assay.¹⁸⁷ The most commonly used method is the BCA assay. In the BCA method the peptide bonds of the protein reduce cupric ions [Cu²⁺] to cuprous ions [Cu⁺] at a rate which is proportional to the amount of protein present in the sample. Subsequently, the BCA reagent binds to the cuprous ions, leading to the formation of a complex which absorbs 562 nm wavelength light. This permits a direct correlation between sample protein concentration and absorbance.^{188,189} The Bradford assay is another method for protein quantification also based on colorimetry principle. It relies on the interaction between the Coomassie brilliant blue dye and the protein based on hydrophobic and electrostatic interactions. Dye binding shifts the absorption maxima from 470 nm to 595 nm.^{190,191} Similarly, the Folin-Lowry method is a two-step colorimetric assay. Step one is the biuret reaction wherein complexes of copper with the nitrogen in the protein molecule are formed. In the second step, the complexed tyrosine and tryptophan amino acids react with Folin-Ciocalteu phenol reagent generating an intense, blue-green color absorbing light at 650–750 nm.¹⁹²

Another simple but less reliable protein quantification method of UV-Vis Absorbance at 280 nm estimates the protein concentration by measuring the absorption of the aromatic residues: phenylalanine, tyrosine, and tryptophan.¹⁹³ This is inaccurate because different complements of proteins will have different proportions of aromatic amino acids. This approach is also sensitive to small molecule interferences that may absorb a similar wavelength.

Fluorescence-Based Methods. Colorimetric assays are inexpensive and require common lab equipment, but colorimetric detection is less sensitive than fluorescence. Total protein in proteomic samples can be quantified using intrinsic fluorescence of tryptophan based on the assumption that approximately 1% of all amino acids in the proteome are tryptophan.¹⁹⁴

NanoOrange (Invitrogen) is an assay for the quantitative measurement of proteins in solution using a merocyanine dye that produces a large increase in fluorescence quantum yield when it interacts with detergent-coated proteins. Fluorescence is measured using 485-nm excitation and 590-nm emission wavelengths. The NanoOrange assay can be performed using fluorescence microplate readers, fluorometers, and laser scanners that are standard in the laboratory.¹⁸⁴

3-(4-Carboxybenzoyl)quinoline-2-carboxaldehyde (CBQCA) is a sensitive fluorogenic reagent for amine detection, which can be used for analyzing proteins in solution. As the number of accessible amines in a protein is modulated by its concentration, CBQCA has a greater sensitivity and dynamic range when measuring protein concentration.¹⁹⁵

Reduction and Alkylation

Typically, disulfide bonds in proteins are reduced and alkylated prior to proteolysis in order to disrupt structures and simplify peptide analysis. This allows better access to all residues during proteolysis and removes the crosslinked peptides created by S-S inter peptide linkages. There are a variety of reagent options for these steps. For reduction, the typical agents used are 5–15 mM concentration of tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl), dithiothreitol (DTT), or 2- β -mercaptoethanol (2BME). TCEP-HCl is an efficient reducing agent, but it also significantly lowers sample pH, which can be abated by increasing sample buffer concentration or resuspending TCEP-HCl in an appropriate buffer system (i.e. 1M HEPES pH 7.5). Following the reducing step, a slightly higher 10–20mM concentration of alkylating agent such as chloroacetamide/iodoacetamide or *n*-ethyl maleimide is used to cap the free thiols.^{196–198} In order to monitor which cysteine residues are linked or modified in a protein, it is also possible to alkylate free cysteine residues with one reagent, reduce di-sulfide bonds (or other cysteine modifications) and alkylate with a different reagent.^{199–201} Alkylation reactions are generally carried out in the dark at room temperature to avoid excessive off-target alkylation of other amino acids.

PROTEOLYSIS

Proteolysis is the defining step that differentiates bottom-up or shotgun proteomics from top-down proteomics. Hydrolysis of proteins is extremely important because it defines the population of potentially identifiable peptides. Generally, peptides between a length of 7–35 amino acids are considered useful for mass spectrometry analysis. Peptides that are too long are difficult to identify by tandem mass spectrometry or may be lost during sample preparation due to irreversible binding with solid-phase extraction sorbents. Peptides that are too short are also not useful because they may match to many proteins during protein inference. There are many choices of enzymes and chemicals that hydrolyze proteins into peptides. This section summarizes potential choices and their strengths and weaknesses.

Before we get into details of various choices for proteolysis, we must discuss terminology. While it is true that “digestion” is commonly used in proteomics, it is important to note that “hydrolysis” is a more specific word choice to describe the chemical process because it refers to breaking peptide bonds within proteins using water. Although hydrolysis may be associated with the complete chemical hydrolysis of proteins into amino acids, for example using high temperature and acid, hydrolysis reactions catalyzed by enzymes such as pepsin and trypsin are specific for certain amino acid residues. In fact, all methods of protein cleavage to shorter peptides require a water molecule for their mechanism of action. In contrast, the definition of “digestion” relates to food breakdown into subunits usable by the body or any chemical process that breaks down substances. Therefore, while “digestion” is indeed a widely used term for the conversion of the proteome to peptides, “hydrolysis” more accurately describes the specific

biochemical process that occurs. We believe that this terminology choice enhances clarity and precision in scientific communication within the field of proteomics.

Trypsin is the most common choice of protease for proteome hydrolysis.²⁰² Trypsin is favorable because of its specificity, availability, efficiency and low cost. Trypsin is a sufficient choice for most proteomics experiments. Trypsin cleaves at the C-terminus of basic amino acids, Arg and Lys, if not immediately followed by proline (although there is debate whether a small number of R/K-P sites are actually cleaved). Many of the peptides generated from trypsin are appropriate in length and hydrophobicity for chromatographic separation, MS-based peptide fragmentation and identification by database search. The main drawback of trypsin is that the majority (56%) of the tryptic peptides are ≤ 6 amino acids, and hence using trypsin alone limits the observable proteome.^{203–205} This limits the number of identifiable protein isoforms and post-translational modifications.

Although trypsin is the most common protease used for proteomics, in theory it can only cover a fraction of the proteome predicted from the genome.²⁰⁶ This is due to production of peptides that are too short to be unique, for example due to R and K immediately next to each other. Peptides below a certain length are likely to occur many times in the whole proteome, meaning that even if we identify them we cannot know their protein of origin. In protein regions devoid of R/K, trypsin may also result in very long peptides that are then lost due to irreversible binding to the solid phase extraction device, or that become difficult to identify due to complicated fragmentation patterns. Thus, parts of the true proteome sequences that are present are lost after trypsin digestion due to both production of very long and very short peptides.

Many alternative proteases are available with different specificities that complement trypsin to reveal different protein sequences,^{203,207} which can help distinguish protein isoforms²⁰⁸ (Figure 2, Table 3). The enzyme choice mostly depends on the application. In general, for a mere protein identification, trypsin is often chosen due to the aforementioned reasons. However, alternative enzymes can facilitate *de novo* assembly when the genomic data information is limited in the public database repositories.^{209–213} Use of multiple proteases for proteome digestion also can improve the sensitivity and accuracy of protein quantification.²¹⁴ Moreover, by providing an increased peptide diversity, the use of multiple proteases can expand sequence coverage and increase the probability of finding peptides which are unique to single proteins.^{64,206,215} A multi-protease approach can also improve the identification of N-Termini and signal peptides for small proteins.²¹⁶ Overall, integrating multiple-protease data can increase the number of proteins identified,^{217,218} increase the identified post-translational modifications^{64,215,219} and decrease the ambiguity of the inferred protein groups.²¹⁵

There are, however, many challenges associated with using alternative proteases. Since peptides are not cleaved after a positively charged residue (like the R/K targeted by trypsin), they may only obtain one precursor charge and be ineffectively fragmented. The lack of a c-terminal positive charge will lead to less consistent y-ion series. Other peptides may obtain too many charges and produce highly charged fragments that are not scored well by search engines. Another common issue with alternative proteases is the potential for producing “shredded” peptides where multiple peptides differ only by a few residues

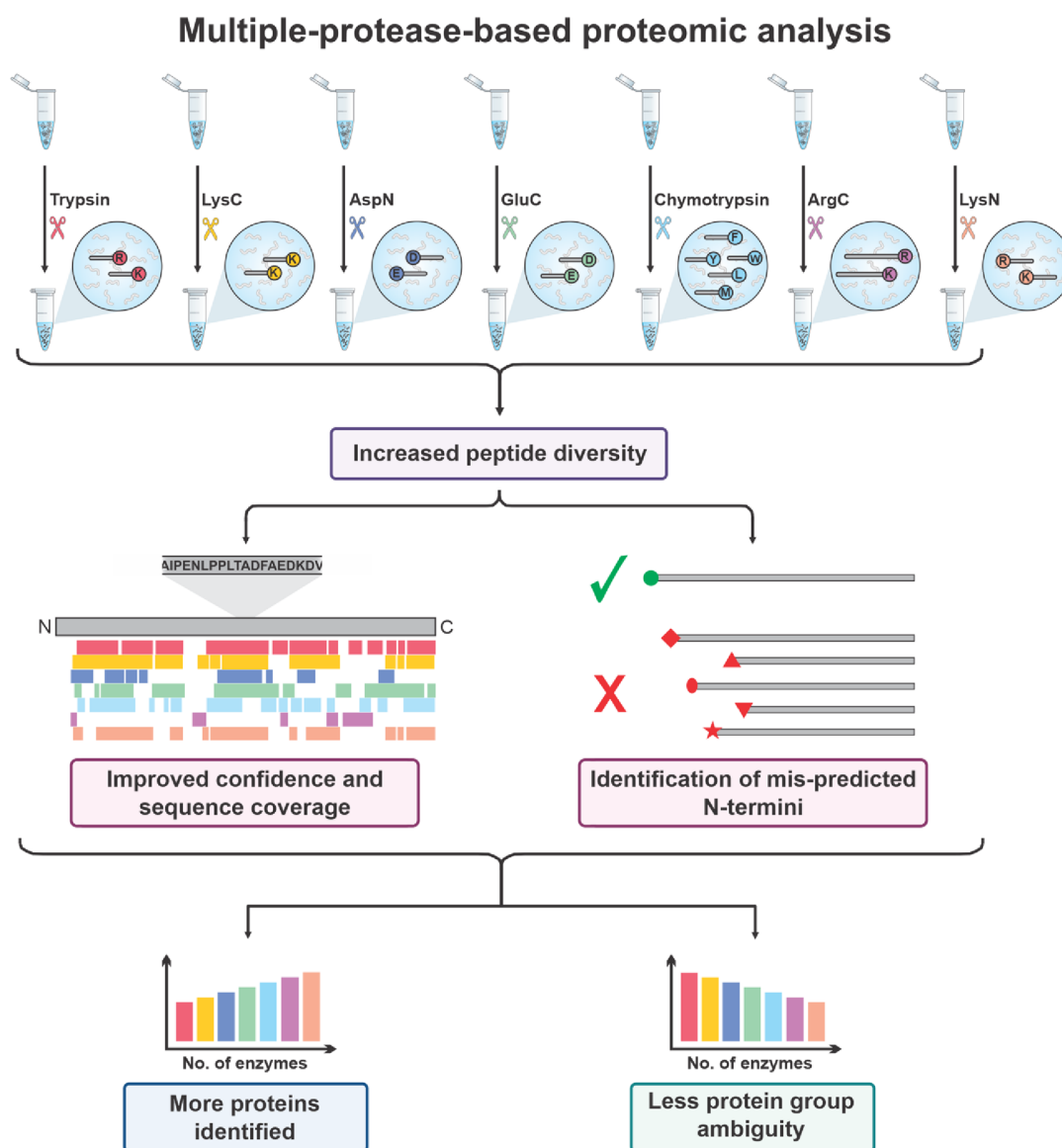


Figure 2. Multiple protease proteolysis improves protein inference. The use of other proteases beyond trypsin such as lysyl endopeptidase (Lys-C), peptidyl-Asp metalloproteinase (Asp-N), glutamyl peptidase I, (Glu-C), chymotrypsin, clostripain (Arg-C), or peptidyl-Lys metalloendopeptidase (Lys-N) can generate a greater diversity of peptides. This improves protein sequence coverage and allows for the correct identification of their N-termini. Increasing the number of complementary enzymes used will increase the number of proteins identified by single peptides and decreases the ambiguity of the assignment of protein groups. Therefore, this will allow more protein isoforms and post-translational modifications to be identified than using trypsin alone.

at either end, thus decreasing the quantity of each species and limiting sensitivity. This problem is worse with proteases that target uncharged residues, because ionic interactions are much stronger than dispersion forces used for binding aliphatic residues.

Lysyl endopeptidase (Lys-C) obtained from *Lysobacter enzymogenes* is a serine protease involved in cleaving carboxyl terminus of Lys.^{204,220} Like trypsin, the optimum pH range required for its activity is from 7 to 9. A major advantage of Lys-C is its resistance to denaturing agents, including 8 M urea - a chaotrope commonly used to denature proteins prior to digestion.²⁰⁸ Trypsin is less efficient at cleaving Lys than Arg, which could limit the quality of quantitation from tryptic peptides. Hence, to achieve complete protein digestion with minimal missed cleavages, Lys-C is often used simultaneously with trypsin digestion.²²¹

Alpha-lytic protease (aLP) is another serine protease secreted by the soil bacterial *Lysobacter enzymogenes*.²²² Wild-type aLP (WaLP) and an active site mutant of aLP, M190A (MaLP), have been used to expand proteome coverage.⁶⁴ Based on observed peptide sequences from yeast proteome digestion, WaLP showed a specificity for small aliphatic amino acids like alanine, valine, and glycine, but also threonine and serine. MaLP showed specificity for slightly larger amino acids like methionine, phenylalanine, and surprisingly, a preference for leucine over isoleucine. The specificity of WaLP for threonine enabled the first method for mapping endogenous human SUMO sites.³⁹

Glutamyl peptidase I, commonly known as Glu-C or V8 protease, is a serine protease obtained from *Staphylococcus aureus*.²²³ Glu-C cleaves at the C-terminus of glutamate, but also after aspartate.^{223,224}

Table 3. Common Proteases Used for Proteomics

Protease	source	class	specificity	optimal pH	notes
Trypsin	mammal pancreas	serine protease	c-term of R/K, not before P	7–9	most common protease
LysC	<i>Lyso bacter enzymogenesis</i>	serine protease	c-term of K	7–9	high stability
Alpha-lytic protease	<i>Lyso bacter enzymogenesis</i>	serine protease	c-term of small side chains	7–9	high stability
GluC	<i>Staphylococcus aureus</i>	serine protease	c-term of D/E	4–8	specificity for Glu depends on buffer
Asp-N	<i>Pseudomonas fragi</i>	metalloprotease	n-term of D	4–9	avoid chelators
Chymotrypsin	mammal pancreas	serine protease	c-term of larger hydrophobics	7–9	
Arg-C	<i>Clostridium histolyticum</i>	cysteine protease	c-term or R	7.2–7.8	avoid oxidation
Ulilysin	<i>Methanosarcina acetivorans</i>	metalloprotease	N-term of R/K	6–9	stable to 55 °C
Lys-N	<i>Grifola frondosa</i>	metalloprotease	N-term or K	7–9	stable to 70 °C
Pepsin A	mammal pancreas	aspartic acid protease	broad including W, F, Y, L	1–4	common for HDX
Proteinase K	<i>Tritirachium album</i>	serine protease	broadest	4–12	common for limited proteolysis

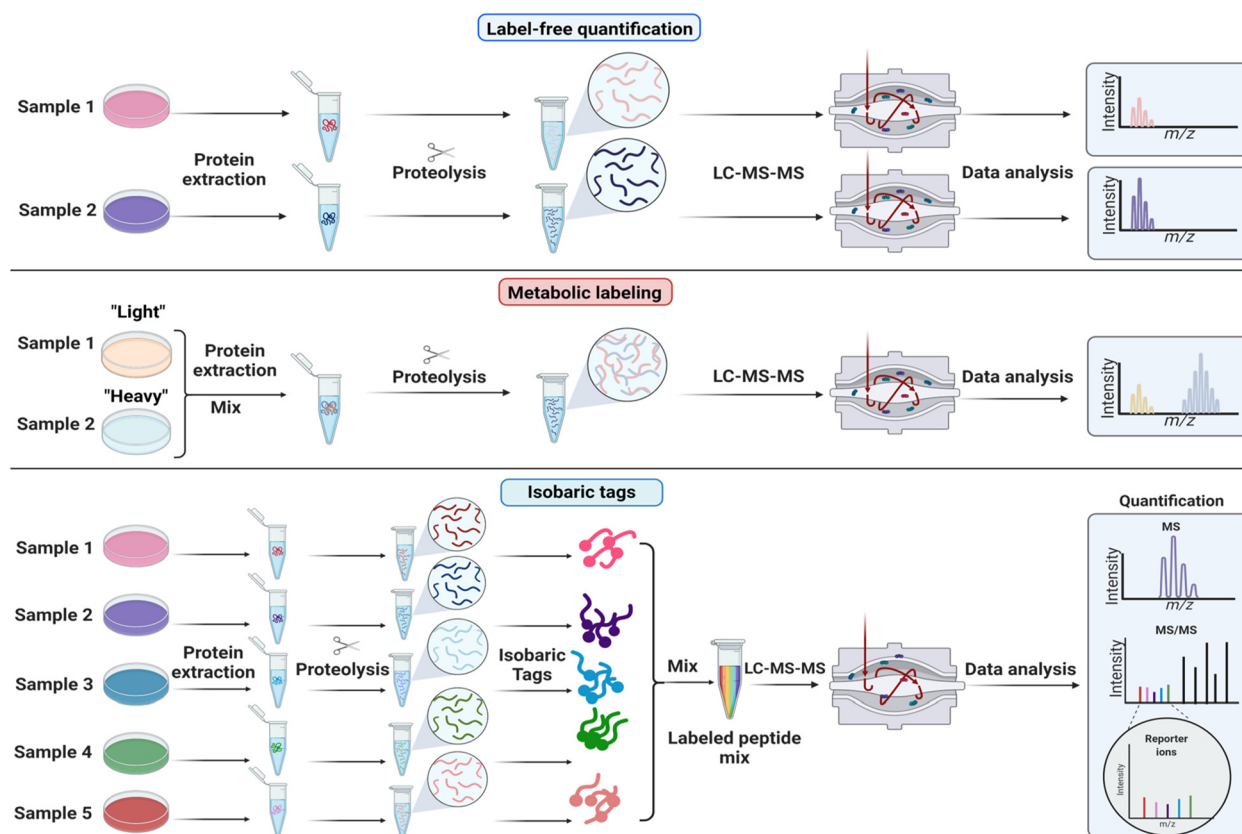


Figure 3. Quantitative strategies commonly used in proteomics. A few non-comprehensive examples of quantification methods are shown. (A) Label-free quantification. Proteins are extracted from samples, enzymatically hydrolyzed into peptides and analyzed by mass spectrometry. Extracted ion chromatograms from peptides are compared across samples that are analyzed sequentially. (B) Metabolic labeling. Stable isotope labeling by amino acids in cell culture (SILAC) is based on feeding cells stable isotope labeled amino acids ("light" or "heavy"). Samples grown with heavy or light amino acids are mixed before cell lysis. The relative intensities of the heavy and light peptide are used to compute protein differences between samples. (C) Isobaric or chemical labeling. Proteins are isolated separately from samples, enzymatically hydrolyzed into peptides, and then chemically tagged with isobaric stable isotope labels. These isobaric tags produce unique reporter mass-to-charge (m/z) signals that are produced upon fragmentation with MS/MS. Peptide fragment ions are used to identify peptides, and the relative reporter ion signals are used for quantification.

Peptidyl-Asp metalloprotease, commonly known as Asp-N, is a metalloprotease obtained from *Pseudomonas fragi*.²²⁵ Asp-N catalyzes the hydrolysis of peptide bonds at the N-terminal of aspartate residues. The optimum activity of this enzyme occurs at a pH range between 4 and 9. As with any metalloprotease, chelators like EDTA should be avoided for digestion buffers when using Asp-N. Studies also suggest that Asp-N cleaves at the amino terminus of glutamate when a

detergent is present in the proteolysis buffer.²²⁵ Asp-N often leaves many missed cleavages.²⁰⁸

Chymotrypsin or chymotrypsinogen A is a serine protease obtained from porcine or bovine pancreas with an optimum pH range from 7–9.²²⁶ It cleaves at the C-terminus of hydrophobic amino acids Phe, Trp, Tyr and barely Met and Leu residues. Since the transmembrane region of membrane proteins commonly lacks tryptic cleavage sites, this enzyme works well with membrane proteins having more hydrophobic

residues.^{208,227,228} The chymotryptic peptides generated after proteolysis will cover the proteome space orthogonal to that of tryptic peptides both in a quantitative and qualitative manner.^{228–230}

Clostripain, commonly known as Arg-C, is a cysteine protease obtained from *Clostridium histolyticum*.²³¹ It hydrolyses mostly the C-terminal Arg residues and sometimes Lys residues, but with less efficiency. The peptides generated are generally longer than that of tryptic peptides. Arg-C is often used with other proteases for improving qualitative proteome data and also for investigating PTMs.²⁰⁴

LysargiNase, also known as Ulilysin, is a recently discovered protease belonging to the metalloprotease family. It is a thermophilic protease derived from *Methanosarcina acetivorans* that specifically cleaves at the N-terminus of Lys and Arg residues.²³² Hence, it enabled discovery of C-terminal peptides that were not observed using trypsin. In addition, it can also cleave modified amino acids such as methylated or dimethylated Arg and Lys.²³²

Peptidyl-Lys metalloendopeptidase, or Lys-N, is an metalloprotease obtained from *Grifola frondosa*.²³³ It cleaves N-terminally of Lys and has an optimal activity at pH 9.0. Unlike trypsin, Lys-N is more resistant to denaturing agents and can be heated up to 70°C.²⁰⁴ Peptides generated from Lys-N digestion produce more c-type ions using ETD fragmentation.²³⁴ Hence this can be used for analysing PTMs, identification of C-terminal peptides and also for *de novo* sequencing strategies.^{234,235}

Pepsin A, commonly known as pepsin, is an aspartic protease obtained from bovine or porcine pancreas.²³⁶ Pepsin was one of several proteins crystalized by John Northrop, who shared the 1946 Nobel prize in chemistry for this work.^{237–240} Pepsin works at an optimum pH range from 1 to 4 and specifically cleaves Trp, Phe, Tyr and Leu.²⁰⁴ Since it possess high enzyme activity and broad specificity at lower pH, it is preferred over other proteases for MS-based disulphide mapping.^{241,242} Pepsin is also used extensively for structural mass spectrometry studies with hydrogen-deuterium exchange (HDX) because the rate of back exchange of the amide deuterium is minimized at low pH.^{243,244}

Proteinase K was first isolated from the mold *Tritirachium album* Limber.²⁴⁵ The epithet 'K' is derived from its ability to efficiently hydrolyze keratin.²⁴⁵ It is a member of the subtilisin family of proteases and is relatively unspecific with a preference for proteolysis at hydrophobic and aromatic amino acid residues.²⁴⁶ The optimal enzyme activity is between pH 7.5 and 12. Proteinase K is used at low concentrations for limited proteolysis (LiP) and the detection of protein structural changes in the eponymous technique LiP-MS.²⁴⁷

Peptide Quantitation Assays

After peptide production from proteomes, it may be desirable to quantify the peptide yield. Quantitation of peptide assays is not as easy as protein lysate assays. BCA protein assays perform poorly with peptide solutions and report erroneous values. A simplistic measurement is to use a nanodrop device, but absorbance measurements from a drop of solution does not report accurate values either. Especially given that low amounts of peptides are often produced for proteomics, more sensitive methods based on fluorescence are preferred. One reliable approach is to Fluorescamine based assay for peptide solutions for higher accuracy.^{248,249} This assay is based on the reaction between a labeling reagent and the N-terminal

primary amine in the peptide(s); therefore, samples must be free of amine-containing buffers (e.g., Tris-based buffer and/or amino acids). This procedure has performance similar to the Pierce Quantitative Fluorometric Peptide Assay (Cat 23290). A second option is also easy to use tryptophan fluorescence to quantify peptide yields,²⁵⁰ which is useful because it does not consume the sample because it uses intrinsic fluorescence.

■ PEPTIDE QUANTIFICATION

Label-Free Quantification (LFQ) of Peptides

LFQ of peptide precursors requires no additional steps in the protein extraction, digestion, and peptide purification workflow (Figure 3). Samples can be taken straight to the mass spectrometer and are injected one at a time, each sample necessitating their own LC-MS/MS experiment and raw file. Quantification of peptides by LFQ is routinely performed by many commercial and freely available proteomics software (see Data Analysis section below). In LFQ, peptide abundances across LC-MS/MS experiments are usually calculated by computing the area under the extracted ion chromatograms for signals that are specific to each peptide; this involves aligning windows of accurate peptide mass and retention time. LFQ can be performed using precursor MS1 signals from DDA, or using multiple fragment ion signals from DIA (see Data Acquisition section). It is important to note that due to differences in peptide ionization efficiency, LFQ only provides relative quantification, not absolute quantities.

Stable Isotope Labeling of Peptides

One approach to improve the throughput and quantitative completeness within a group of samples is sample multiplexing via stable isotope labeling. Multiplexing enables pooling of samples and parallel LC-MS/MS analysis within one run. Quantification can be achieved at the MS1- or MSn-level, dictated by the upstream labeling strategy.

Stable isotope labeling methods produce peptides that are chemically identical from each sample that differ only in their mass. Methods include stable isotope labeling by amino acids in cell culture (SILAC)²⁵¹ and chemical labeling such as amine-modifying tags for relative and absolute quantification (mTRAQ)²⁵² or dimethyl labeling.²⁵³ The latter two methods are chemical labeling processes after proteome or peptide purification. In all these approaches, the labeling of each sample imparts mass shifts (e.g. 4 Da, 8 Da) which can be detected within the MS1 full scan. The ability to label samples in cell culture has enabled impactful quantitative biology experiments.^{254,255} These approaches have nearly exclusively been performed using data-dependent acquisition (DDA) strategies. However, recent work employing faster instrumentation has shown the benefits of chemical labeling with 3-plex mTRAQ or dimethyl labels for data-independent acquisition (DIA),^{256,257} an idea originally developed nearly a decade earlier using chemical labels to quantify lysine acetylation and succinylation stoichiometry.²⁵⁸ As new tags with higher plexing become available, strategies like plexDIA and mDIA are sure to benefit.^{256,257}

Peptide Labeling with Isobaric Tags

Another approach is multiplexing via isobaric labels, a strategy which enables parallel data acquisition after pooling of samples. Commercial isobaric tags include tandem mass tags (TMT)²⁵⁹ and isobaric tags for relative and absolute quantification (iTRAQ)²⁶⁰ amongst others, and several non-commercial

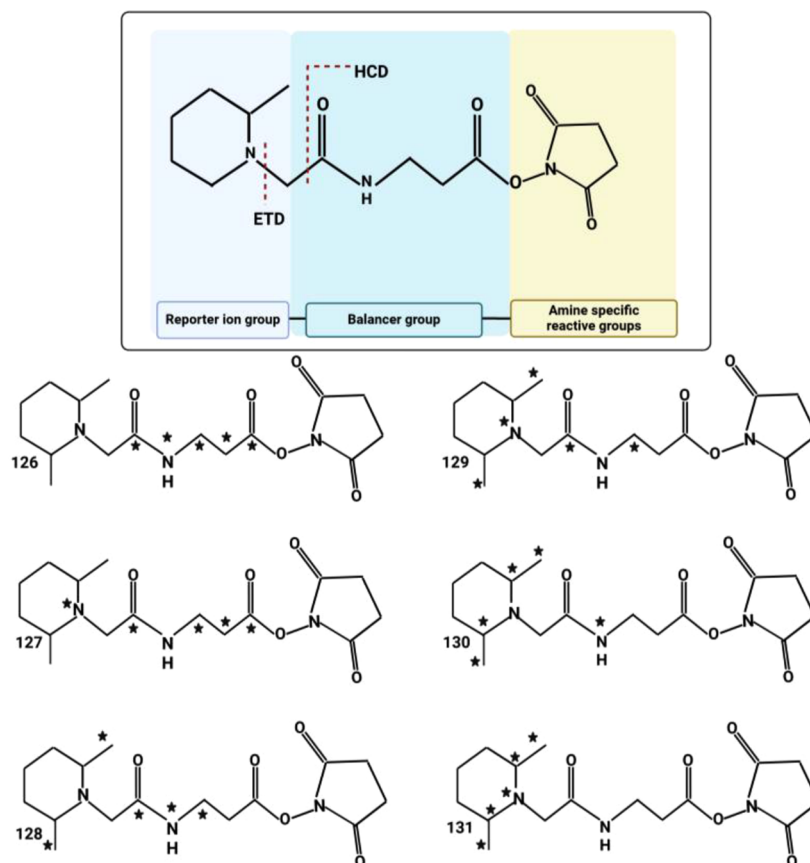


Figure 4. Chemical structure of isobaric tags. This shows the TMT 6-plex from ThermoFisher, which is an example of an isobaric tag. The structure has three elements, the reactive group (in this case *N*-hydroxysuccinimide), the balancer, and the reporter. The reactive group enables quick covalent conjugation to nucleophilic amines found at peptide *N*-terminus and lysine side chains. The balance and reporter groups together contain a total of six heavy isotopes. The stars in the structures indicate the positions of all six heavy atoms for each TMT form. For this reason, a sample labeled by the any version will have the same precursor mass. However, upon fragmentation, the balance group is lost and the reporter retains a charge. The reporter group is measured in the low mass region and is proportional to the starting amount of each sample before mixing. This ratio of reporter signals enables relative quantification.

options have also been developed.²⁶¹ Although isobaric tags enable collection of data from many samples at once, to improve depth, fractionation by high pH reversed phase is often used, which limits the benefit in throughput.

The isobaric tag labeling-based peptide quantitation strategy uses derivatization of every peptide sample with a different isotopic incorporation from a set of isobaric mass tags. All isobaric tags have a common structural theme consisting of (1) an amine-reactive groups (usually triazine ester or *N*-hydroxysuccinimide [NHS] esters) which react with peptide *N*-termini and ϵ -amino group of the lysine side chain of peptides, (2) a balancer group, and (3) a reporter ion group (Figure 4).

Peptide labeling is followed by pooling the labelled samples, which undergo MS and MS/MS analysis. Peptides labeled with these tags give a single MS peak with the same precursor *m/z* value in an MS1 scan and identical retention time of LC analysis. The modified parent ions undergo fragmentation during MS/MS analysis generating two kinds of fragment ions: (a) reporter ions and (b) peptide fragment ions. Each reporter ions' relative intensity is directly proportional to the peptide abundance in each of the starting samples that were pooled. As usual, *b*- and *y*-type fragment ion peaks are still used to identify amino acid sequences of peptides, from which proteins can be inferred. Since it is possible to label most tryptic peptides with

an isobaric mass tag at least at the *N*-termini, numerous peptides from the same protein can be detected and quantified, thus leading to an increase in the confidence in both protein identification and quantification.²⁶³

Because the size of the reporter ions is small and sometimes the mass difference between reporter ions is small (i.e., an ~ 6 mDa difference when using 13C versus 15N), these methods almost exclusively employ high-resolution mass analyzers, not classical ion traps.²⁶⁴ There are examples, however, of using isobaric tags with pulsed *q* dissociation on linear ion traps (LTQs).²⁶⁵ Suitable instruments are the Thermo Q-Exactive, Exploris, Tribrid, and Astral lines, or Q-TOFs such as the TripleTOF or timsTOF platforms.^{266,267}

The following are some of the isobaric labeling techniques:

isobaric Tags for Relative and Absolute Quantitation (iTRAQ)

The iTRAQ tagging method covalently labels the peptide *N*-terminus and side-chain primary amines with tags of different masses through the NHS-ester bond. This was the first isobaric tagging method to find widespread use, but it isn't used as much anymore. This is followed by mass spectrometry analysis.²⁶⁸ Reporter ions for an 8-plex iTRAQ are measured at nominally 113, 114, 115, 116, 117, 118, 119, and 121 *m/z*. Currently, two kinds of iTRAQ reagents are available: 4-plex and 8-plex.²⁶⁹ Using 4-plex reagents, a maximum of four

different biological conditions can be analyzed simultaneously (i.e., multiplexed), whereas using 8-plex reagents enables the simultaneous analysis of eight different biological conditions.^{270,271}

iTRAQ Hydrazide (iTRAQH)

iTRAQH is an isobaric tagging reagent for the selective labeling and relative quantification of carbonyl (CO) groups in proteins.²⁷² The reactive CO and oxygen groups which are generated as the byproducts of oxidation of lipids at the time of oxidative stress causes protein carbonylation.²⁷³ iTRAQH is produced from iTRAQ and a surplus of hydrazine. This reagent reacts with peptides which are carbonylated, thus forming a hydrazone group. iTRAQH is a novel method for analyzing carbonylation sites in proteins utilizing an isobaric tag for absolute and relative quantitation iTRAQ derivative, iTRAQH, and the analytical power of linear ion trap instruments (QqLIT). This new strategy seems to be well suited for quantifying carbonylation at large scales because it avoids time-consuming enrichment procedures.²⁷² Thus, there is no need for enriching modified peptides before LC-MS/MS analysis.

Tandem Mass Tag (TMT)

TMT labeling is based on a similar principle as that of iTRAQ. The TMT label is based on a glycine backbone and this limits the amount of sites for heavy atom incorporation. In the case of 6-plex-TMT, the masses of reporter groups are nominally 126, 127, 128, 129, 130, and 131 Da.²⁶⁴ 10- and 11-plex TMT kits were recently supplanted by proline-based TMT tags (TMTpro), originally introduced as 16-plex kits in 2019²⁷⁴ and upgraded to an 18-plex platform in 2021.²⁷⁵ Due to co-isolation of multiple precursors leading to reporter ion compression, TMT works best with MS platforms which allow quantitation at the MS3 level (e.g., Thermo Fisher Orbitrap Tribrid instruments).^{262,276} In experiments performed on Q-Orbitrap or Q-TOF platforms, MS2-based sequence identification (via b- and y-type ions) and quantitation (via low m/z reporter ion intensities) is performed. In experiments performed on Q-Orbitrap-LIT platforms, MS3-based quantitation can be performed wherein the top ~10 most abundant b- and y-type ions are synchronously co-isolated in the linear ion trap and fragmented once more before product ions are scanned out in the Orbitrap mass analyzer. Adding an additional layer of gas-phase purification limits the ratio distortion of co-isolated precursors within isobaric multiplexed quantitative proteomics.^{277,278} Infrared photoactivation of co-isolated TMT fragment ions generates more quantitative reporter ion generation and sensitivity relative to standard beam-type collisional activation.²⁷⁹ High-field asymmetric waveform spectrometry (FAIMS) also aids the accuracy of TMT-based quantitation on Tribrid systems.²⁸⁰ TMT is widely used for quantitative protein biomarker discovery. In addition, TMT labeling technique helps multiplex sample analysis enabling efficient use of instrument time. TMT labelling also controls for technical variation because after samples are mixed the ratios are locked in, and any sample loss would be equal across channels.

iodoTMT

IodoTMT reagents are isobaric reagents used for tagging cysteine residues of peptides. The commercially available IodoTMT reagents are IodoTMTzero and IodoTMT 6-

plex.^{281,282} These reagents are useful for studies of cysteine oxidation modifications because only unoxidized cysteine is modified.

aminoxylTMT Isobaric Mass Tags

Also referred to as glyco-TMTs, these reagents have chemistry similar to iTRAQH. The stable isotope-labeled glyco-TMTs are utilized for quantitating N-linked glycans. They are derived from the original TMT reagents with an addition of carbonyl-reactive groups, which involve either hydrazide or aminoxy chemistry as functional groups. These aminoxy TMTs show a better performance as compared to its iTRAQH counterparts in terms of efficiency of labeling and quantification. The glyco-TMT compounds consist of stable isotopes thus enabling (i) isobaric quantification using MS/MS spectra and (ii) quantification in MS1 spectra using heavy/light pairs. Aminoxy TMT6-128 and TMT6-131 along with the hydrazide TMT2-126 and TMT2-127 reagents can be used for isobaric quantification. In the quantification at MS1 level, the light TMT0 and the heavy TMT6 reagents have a difference in mass of 5.0105 Da which is sufficient to separate the isotopic patterns of all common N-glycans. Glycan quantification based on glyco-TMTs generates more accurate quantification in MS1 spectra over a broad dynamic range. Intact proteins or their digests obtained from biological samples are treated with PNGase F/A glycosidases to release the N-linked glycans during the process of labeling using aminoxyTMT reagents. The free glycans are then purified and labeled with the aminoxyTMT reagent at the reducing end. The labeled glycans from individual samples are subsequently pooled and then undergo analysis in MS for identification of glycoforms in the sample and quantification of relative abundance of reporter ions at MS/MS level.²⁸³

N,N-Dimethyl leucine (DiLeu)

An isobaric tag reagent was developed using N,N-Dimethyl leucine, also referred to as DiLeu.²⁸⁴ Each incorporated label produces a 145.1 Da mass shift. A maximum of 21 samples can be multiplexed with DiLeu. The labeling efficiency of DiLeu tags are similar to that of the iTRAQ tags. DiLeu-labeled peptides offer increased confidence of identification of peptides and more reliable quantification as they undergo better fragmentation, generating higher reporter ion intensities.²⁸⁴

Deuterium isobaric Amine Reactive Tag (DiART)

DiART is an isobaric tagging method used in quantitative proteomics.^{285,286} The reporter group in DiART tags is a N,N'-dimethyl leucine reporter group with a mass to charge range of 114–119. DiART reagents can label a maximum of six samples and further analyzed by MS. The isotope purity of DiART reagents is very high hence correction of isotopic impurities is not needed at the time of data analysis.²⁸⁷ The performances of DiART including the mechanism of fragmentation, the number of proteins identified and the quantification accuracy are similar to iTRAQ. Reporter ions of high-intensity are produced by DiART tags in comparison to those with iTRAQ and thus, DiART labeling can be used to quantify more peptides as well as those with lower abundance, and with reliable results.²⁸⁵ DiART serves as a cheaper alternative to TMT and iTRAQ while also having a comparable labeling efficiency. It has been observed that these tags are useful in labeling huge protein quantities from cell lysates before TiO2 enrichment in quantitative phosphoproteomics studies.²⁸⁸

Hyperplexing or Higher-Order Multiplexing

Some studies have combined metabolic labels (i.e., SILAC) with chemical tags (i.e., iTRAQ or TMT) to expand the multiplexing capacity of proteomics experiments referred to as hyperplexing^{289,290} or higher order multiplexing.^{291–293} This technique combines MS1- and MS2-based quantitative methods to achieve enhanced multiplexing by multiplying the channels used in each dimension. This allows for the quantitation of proteomes across multiple samples in a single MS run. The technique uses two types of mass encoding to label different biological samples. The labeled samples are then mixed together, which increases the MS1 peptide signal. Protein turnover rates were studied using SILAC-iTRAQ multitagging,²⁹⁴ while various combined precursor isotopic labeling and isobaric tagging (cPILOT) studies employed MS1 dimethyl labeling with iTRAQ.^{295–298} SILAC-TMT hyperplexing was used to study the temporal response to rapamycin in yeast.²⁹⁹ SILAC-iTRAQ-TAILS method was developed to study matrix metalloproteinases in the secretomes of keratinocytes and fibroblasts.³⁰⁰ TMT-SILAC hyperplexing was used to study synthesis and degradation rates in human fibroblasts.³⁰¹ Variants of SILAC-iTRAQ and BONCAT, namely, BONPlex³⁰² and MITNCAT,³⁰³ were also developed to study temporal proteome dynamics.

■ ENRICHMENT AND DEPLETION

In order to study low abundance protein modifications, or to study rare proteins in complex mixtures, various methods have been developed to enrich or deplete specific proteins or peptides.

Peptide Enrichment

Phosphoproteomics. Protein phosphorylation, a hallmark of protein regulation, dictates protein interactions, signaling, and cellular viability. This post-translational modification (PTM) involves the installation of a negatively charged phosphate moiety (PO 4-) onto the hydroxyl side-chain of serine or threonine residues on target proteins. Additionally, while less commonly modified than serine and threonine, histidine,^{304–306} arginine,³⁰⁷ and tyrosine^{308–310} phosphorylation also represent important cell signaling biology. Protein kinases catalyze the transfer of PO 4- group from ATP to the nucleophile (OH) group of serine, threonine, and tyrosine residues, while protein phosphatases catalyze the removal of PO4-. Phosphorylation changes the charge of a protein, often altering protein conformation and therefore function.³¹¹ Protein phosphorylation is one of the major PTMs that alters the stability, subcellular location, enzymatic activity complex formation, degradation of protein, and cell signaling of protein with a diverse role in cells.^{312–314} Phosphorylation can regulate almost all cellular processes, including metabolism, growth, division, differentiation, apoptosis, and signal transduction pathways.³⁴ Rapid changes in protein phosphorylation are associated with several diseases.³¹⁵

Several methods are used to characterize phosphorylation using modification-specific enrichment techniques combined with advanced MS/MS methods and computational data analysis.³¹⁶ There are many challenges with studying phosphorylation.³¹⁷ For example, many phosphopeptides are low stoichiometry compared to non-phosphorylated peptides, which makes them difficult to identify. Phosphopeptides also exhibit low ionization efficiency.³¹⁸ To overcome these challenges, it is important to reduce sample complexity to

detect large numbers of phosphorylation sites. This is accomplished using enrichment the modified proteins and/or peptides.^{319–321}

As with any proteomics experiment, phosphoproteomics studies require protein extraction, proteolytic enzyme digestion, phosphopeptide enrichment, peptide fractionation, LC-MS/MS, bioinformatics data analysis, and biological function inference. Special consideration is required during protein extraction where the cell lysis buffer should include phosphatase inhibitors such as sodium orthovanadate, sodium pyrophosphate and beta-glycerophosphate.³²²

Enrichment can be done at the protein level before proteolysis. Phosphoprotein enrichment typically involves the use of immobilized metal-affinity chromatography (IMAC) to selectively capture phosphorylated proteins based on their high-affinity binding to metal ions such as Ga(III), Fe(III), Zn(II) and Al(III).^{323–327}

Enrichment is more commonly performed at the peptide level because there are several advantages over phosphoprotein-level enrichment. First, peptides have simpler three-dimensional structures than proteins, which makes them easier to separate and analyze. Second, phosphopeptide enrichment is not hindered by small, lipophilic, and very acidic or alkaline proteins.³²¹ Third, prefractionation techniques such as strong anion exchange chromatography (SAX), strong cation exchange chromatography (SCX) and hydrophilic interaction chromatography (HILIC) are easier to use for peptide separation than they are for protein separation, and they are more sensitive than 2D-gel electrophoresis that is often used for intact proteins.^{328,329} As a result, phosphopeptide enrichment has yielded more experimental data than phosphoprotein enrichment.³²⁶ Phosphopeptide enrichment is typically done after any isobaric labeling strategy, although several have investigated the importance of order at these stages.

Phosphopeptide enrichment often uses titanium dioxide (TiO₂)³³⁰ and/or IMAC such as Fe³⁺ coupled to solid-phase materials.^{322,325,331} The most common cost-effective beads for phosphopeptide extraction with Ti are ReSyn and GL Sciences, and CubeBio for Fe-based beads. Often organic acids such as glutamic acid, lactic acid, glycolic acid are added to compete with acidic non-phosphopeptides for binding to the metal-ions. Carr and coworkers even demonstrated phosphoproteome analysis without any enrichment.³³²

The use of Fe-IMAC column chromatography allows for the improved phosphopeptide enrichment from complex peptide mixtures.³³³ Compared to other formats like StageTips or batch incubations with TiO₂ or Ti-IMAC beads, Fe-IMAC columns do not suffer from problems with poor binding or elution of phosphopeptides, and the efficiency of enrichment increases linearly with the amount of starting material.³³⁴ Also with recent improvements to Ti based beads, the MagReSyn Ti-IMAC HP with Ti⁴⁺ attached with a flexible linker (to reduce steric hindrance) activated with phosphonate groups for Ti⁴⁺ chelation, and the MagReSyn Zr-IMAC HP, also with a flexible linker activated with phosphonate groups for Zr⁴⁺ chelation, have shown superior phosphopeptide extraction as compared to Fe-IMAC.

Multiple IMAC steps can be used in parallel or sequentially to improve phosphopeptide coverage. Lai et al. showed that the combined use of Fe³⁺-IMAC and Ti(4+)-IMAC chromatography enables complementary identification of more phosphorylation sites than either technique alone.³³⁵ A novel phosphopeptide enrichment technique using sequential

enrichment with magnetic Fe₃O₄ and TiO₂ particles was developed to detect mono- and multi-phosphorylated peptides.³³⁶

More recently, the use of Src Homology 2 (SH2) domains as specific affinity reagents for phosphotyrosine is an emerging technology allowing an expanded fractionation of tyrosine phosphopeptides. Here, short chain protein domains have been constructed and affinity enhanced through yeast two hybrid screening to arrive at high affinity matrices capable of outperforming traditional IMAC approaches.^{337,338}

Tips for Studying Phosphorylation.

- Cell lysates should always be prepared using phosphatase inhibitors and samples should be placed on the ice during sonication for protein extraction.
- Increase the amount of starting material of your sample for phosphoenrichment to at least 1 mg of protein or more for optimal results.
- If using anti-phosphorylation antibodies, ensure their specificity is confirmed with other methods.
- Make sure to select a suitable method for the phosphoenrichment that fits the experiment goals.
- TiO₂-based phosphopeptide enrichment methods have different enrichment specificities; selecting non-phosphopeptide excluders such as glutamic acid, lactic acid, glycolic acid, and dihydroxybenzoic acid are the key part of the study.³³⁹
- Do not use milk as a blocking agent when western blotting for phosphorylation because milk contains the phosphoprotein casein and can lead to a higher background due to non-specific binding.

Glycosylation. Mass spectrometry-based analysis of protein glycosylation has emerged as the premier technology to characterize such a universal and diverse class of biomolecules. Glycosylation is a heterogeneous post-translational modification that decorates many proteins within the proteome, conferring broad changes in protein activity.^{56,340} This PTM can take many forms. The covalent linkage of mono- or oligosaccharides to polypeptide backbones through a nitrogen atom of asparagine (N) or an oxygen atom of serine (S) or threonine (T) side-chains creates N- and O-glycans, respectively. The heterogeneity of proteoglycans is not directly tied to the genome, and thus cannot be inferred. Rather, the abundance and activity of protein glycosylation is governed by glycosyltransferases and glycosidases which add and remove glycans, respectively. The fields of glycobiology and bio-analytical chemistry are intricately intertwined with mass spectrometry at the center thanks in part to its power of detecting any modification that imparts a mass shift.

Due to the myriad glycan structures and proteins which harbor them, the enrichment of glycoproteins or glycopeptides is not as streamlined as that of other PTMs.³⁴¹ The enrichment of glycoproteome from the greater proteome inherently introduces bias prior to the LC-MS/MS analysis. One must take into account which class or classes of glycopeptides they are interested in analyzing before enrichment for optimal LC-MS/MS results. Glycopeptides can be enriched via glycan affinity, for example to glycan-binding proteins, chemical properties like charge or hydrophilicity, chemical coupling of glycans to stationary phases, and by bioorthogonal, chemical biology approaches. Glycan affinity-based enrichment strategies include the use of lectins, antibodies, inactivated enzymes, immobilized metal affinity chromatography

(IMAC), and metal oxide affinity chromatography (MOAC). The enrichment of glycopeptides by their chemical properties, for example by biopolymer charge and hydrophobicity, include hydrophilic interaction chromatography (HILIC), electrostatic repulsion-hydrophilic interaction chromatography (ERLIC), and porous graphitic carbon (PGC). One variation of ERLIC that combines strong anion exchange, electrostatic repulsion, and hydrophilic interaction chromatography (SAX-ERLIC) has risen in popularity thanks to robustness and commercially available enrichment kits.^{342,343}

Chemical coupling methods most often used to enrich the glycoproteome employ hydrazide chemistry for sialylated glycopeptides. Glycan are cleaved from the stationary phase by PNGase F. The dependence of chemical coupling methods on PNGase F biases their output toward N-glycopeptides. Alkoxyamine compounds and boronic acid-based methods have also shown utility. We direct readers to several reviews on glycopeptide enrichment strategies.^{341,344–347}

Antibody Enrichments of Modifications. Western blot analysis is used to detect the PTMs in a protein through the use of antibodies.³⁴⁸ As an extension, pan-PTM antibodies have been used to isolate peptides bearing the PTM of interest.³⁴⁹ One benefit of this approach is that peptides are less likely to experience non-specific binding than proteins.³¹⁶ Initially, peptide immunoaffinity precipitation was developed to enrich phosphotyrosine-containing peptides.³⁵⁰ Peptide immunoprecipitation yielded significantly greater coverage of the phosphotyrosine proteome than global phosphorylation enrichment strategies by enriching for a subset of the phosphoproteome. Since then, peptide immunoaffinity precipitation has been used successfully to enrich for peptides with other phosphorylation motifs^{351,352} as well as peptides with other modifications such as the diglycyl-lysine residue of ubiquitin modification after trypsin proteolysis,^{353–355} acetyl-lysine,^{356–360} arginine methylation,³⁶¹ tyrosine nitration,³⁶² and tyrosine phosphorylation.^{363,364}

The O-linked β -D-N-acetylglucosamine (O-GlcNAc) is found on serine and threonine residues and is involved in involved in the progression of cancers in multiple systems throughout the body.³⁶⁵ Anti-O-GlcNAc monoclonal antibody enables enrichment from O-GlcNAcylated peptides of cells and tissues. These antibodies have high sensitivity and specificity toward O-GlcNAc-modified peptides and do not identify O-GalNAc or GlcNAc in extended glycans.³⁶⁶

Protein Depletion (Blood Samples)

Many plasma proteomics studies involve the analysis of plasma.^{367,368} However, the abundance range of proteins in the blood/plasma proteome exceeds 10 orders of magnitude. Due to this wide dynamic range, detection of proteins with medium and low abundance by proteomic analyses is difficult,³⁶⁹ and identifying protein biomarkers from biological samples such as blood is often obstructed by proteins present at higher concentrations. In fact, the top 14 most abundant proteins in human plasma constitute over 99% of the total protein mass. The removal of these high-abundant proteins enables the detection of less abundant proteins. The ability to deplete abundant proteins with specificity, reproducibility, and selectivity is extremely important in proteomic studies.³⁷⁰

The following are some of the methods used for abundant protein depletion:

Dye-Ligand Depletion. This method is used for the depletion of serum albumin based on the interaction between

albumin and dyes like Cibacron Blue (CB) through electrostatic force, hydrogen bonding and hydrophobic interactions. The method is relatively low cost, widely available, robust and has high binding capacity. However, it lacks specificity and has varying efficiency.^{371,372}

Protein-Ligand Depletion. This method is used for depletion of immunoglobulins (Ig) based on the interaction between the Fragment crystallizable (Fc) region of these Ig's³⁷³ and cell wall protein A, G, or A/G of *Staphylococcus aureus* and *Streptococcus* spp.^{374,375} It is highly selective and has high yield and purity. However, non-specific binding may occur due to co-absorption of other proteins.³⁷⁶

Immunodepletion. This method is used for depletion of proteins having high abundance in plasma or serum on the basis of the specific interaction of these proteins with their respective antibodies (antigen-antibody interaction).³⁷⁷ Immunodepletion has high specificity and commercial kits based on columns (Agilent MARS column) or loose beads (Thermo-Fisher High-Select depletions beads) both deplete the Top 14 most abundant protein in blood and are also readily available but expensive. For some protocols, non-specific binding to these immunodepletion columns or beads by proteins of interest is highly dependent on washing conditions.³⁷⁶

Combinatorial Peptide Ligand Library. This method is used for partial depletion of major proteins, i.e., those with high abundance and for relative enrichment of lower and medium abundant proteins.³⁷⁸ It is based on the interaction with an array of ligands which are essentially peptides of 6 amino acids in length. It is also used for normalization of the global protein abundance.³⁷⁹ However, the drawbacks include non-specific binding as well as loss of proteins due to incomplete elution or inefficient binding.³⁷⁶

Precipitation. This method of abundant protein depletion works by altering the solubility of proteins using a chemical reagent including inorganic salt solutions,³⁸⁰ organic solvents,³⁸¹ non-ionic polymers,³⁸² and reducing agents.³⁸³ It is extremely simple and cost-effective. However, it is less specific with a risk of protein loss, difficulty in protein resolubilization, in addition to being time consuming.³⁷⁶

New Technologies. Newer methods of highly abundant protein depletion are based on the interaction between polymers such as bacterial cellulose nanofibers,³⁸⁴ cryogels,³⁸⁵ and nanomaterials.³⁸⁶ These techniques are highly specific, relatively cheap, and very stable. They can also be reused since they have larger binding capacity and less cross-reactivity.³⁷⁶

Protein enrichment/depletion strategies which make use of protein coronas^{387,388} or extracellular vesicle enrichment³⁸⁹ are enabling researchers to probe deeper into the plasma, serum, lymph, and cerebrospinal fluid proteomes. Automated nanoparticle (NP) protein corona-based proteomics workflows are a novel approach to perform deep blood-based proteomics analysis at unprecedented protein IDs above 6000 proteins.³⁹⁰ NPs can efficiently compress the dynamic range of protein abundances into a mass spectrometry accessible detection range and allow full automation of the protein preparation process providing a platform that can rival affinity based approaches with equivalent reproducibility and sensitivity.³⁹¹

■ PEPTIDE PURIFICATION AND FRACTIONATION

Peptide Purification Methods

Before peptide analysis, interferences from sample preparation must be removed. There are several approaches to purify peptides.

Solid Phase Extraction (SPE)

Solid phase extraction (SPE) is a common MS-based proteomics technique employed during sample preparation. In this method, compound isolation is based on chemical and physical properties, which determines the distribution of compounds between a mobile phase (liquid) and a stationary phase (solid). After the molecules bind, washing of the bound compounds is performed and then molecules are made to elute from the stationary phase after replacing the mobile phase with the elution buffer. The material used for SPE is usually discarded after every sample and no gradient is applied for elution (single-step procedure of elution).³⁹² Thus, using SPE only a specific analyte group gets separated, which depends on the stationary phase. Hence, SPE is primarily used for sample clean-up and for reducing complexity of the sample. For MS-based proteomic analysis, it is largely used to get rid of salts and other contaminants that might lead to ion suppression.

The material for SPE is available in various types, including (micro-) columns, cartridges, plates, micropipette tips, and functionalized magnetic beads (MBs).^{393,394} Reversed-phase is the most widely used material for SPE in proteomic studies for the proteins and peptide fractionation and rarely, ion-exchange material. For the separation of glycosylated proteins and peptides, the preferred material is normal phase such as HILIC.^{395,396} SPE materials which are less commonly used are silica- or polystyrene-based ones.^{397,398} The other types of SPE methods are IEX, metal chelation, and affinity-based.³⁹⁹

The basic idea behind the choice of binding and wash versus elution solutions for SPE is that the binding and wash solutions should favor the interaction between the analytes of interest and the solid phase, whereas the elution solution should favor the interaction of the analyte with the liquid phase (Figure 5). For example, with reversed phase SPE, the solid phase is C18 or some other hydrophobic chemistry. Binding of peptides to this solid phase is based on the hydrophobicity of peptides, mostly due to the presence of hydrophobic amino acid side chains; leucine is the most common amino acid in human proteins. To encourage peptides to 'like' the stationary phase more than the liquid phase, the peptides are loaded in aqueous solution. This will enable washing of the hydrophilic contaminants like salts, small polar buffer molecules, and polar denaturants like urea. After washing the bound peptides, they can be eluted by switching the liquid phase to something hydrophobic, which allows the peptides to partition more into the liquid phase and elute from the solid phase.

Specific Types of Peptide Purification. There are many additional peptide purification methods that are commonly used in proteomics currently. These methods include the following:

1. StageTips, in-stagetiip (iST)^{400,401}
2. SP2 or SP3⁴⁰²
3. Suspension trapping (S-trap)⁵⁹

Peptide Fractionation Methods

The number of peptides produced from proteolysis of the whole proteome is immense. Thus, after peptides are cleaned

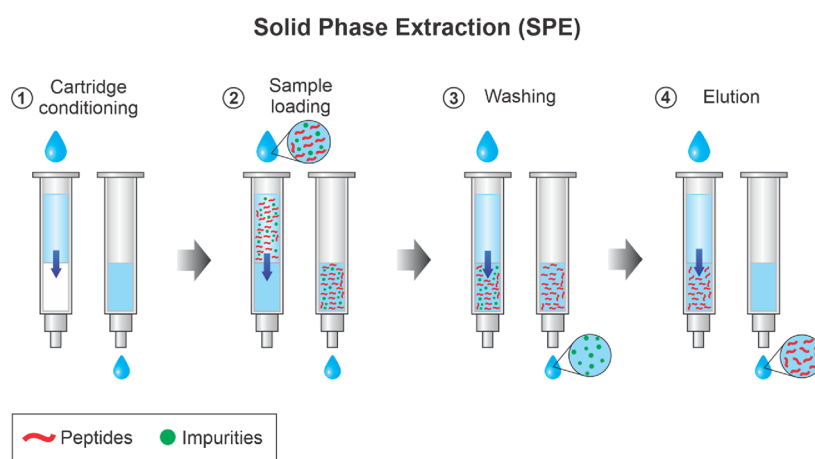


Figure 5. Solid phase extraction (SPE). SPE is a sample preparation technique that uses a solid adsorbent contained most commonly in a cartridge device to selectively adsorb certain molecules from solution. The first step is the conditioning of the cartridge which involves wetting the adsorbent to solvate its functional groups and filling the void spaces with solvent thereby removing any air in the column. This is necessary to produce a suitable environment for adsorption and thus ensure reproducible interaction with the analytes. After conditioning, the sample is loaded in the cartridge. This can be performed with the aid of positive or negative pressure to ensure a constant flow rate. In this step molecules bind the adsorbent and interferences pass through. Next, the column is washed with the mobile phase to eliminate the contaminants while ensuring the analyte remains bound. Finally, peptides are eluted in an appropriate buffer solution with polarity or charge that competes with interaction with the solid phase.

from interferences, they are often fractionated into subsets to enable increased proteome coverage. The characterization of the whole proteome is expected from higher order organisms, and with rising interest in post-translational modifications, an elaborate coverage of protein sequence is required. There are different methods for peptide fractionation as follows:

Ion-Exchange chromatography (IEC). This method involves the separation based on contrasting electric charge.⁴⁰³ In this approach, the mechanism of analyte retention is based on the principle of electrostatic attraction between the sample and the stationary phase functional groups (FGs), having opposite charges. IEC is classified into two types: cation-exchange and anion-exchange chromatography. In cation-exchange chromatography, at an acidic pH, the negatively charged functional groups such as sulfates are attracted to positively charged peptides, whereas, in anion-exchange chromatography, positively charged FGs such as quaternary ammoniums are attracted to peptides with negative charge at an alkaline pH. These techniques are further classified into: strong (cation [SCX] and anion [SAX] exchange), and weak exchangers (cation [WCX] and anion [WAX] exchange), based on the type of FG attached.⁴⁰⁴ These functional groups are most commonly supported in resins made up of silica and synthetic polymers, however, some inorganic materials are sometimes used.⁴⁰³ In the IEC method, peptide elution is performed using a mobile phase with higher ionic strength, to ensure peptide partition into the liquid phase. SCX along with a salt gradient/plug is a routinely used proteomics technique. In the SCX method, peptides are resolved according to their net charge, in which the peptide with the lowest positive charge is eluted first. Increasing the salt concentration decreases the peptide retention time due to competition with the electrostatic interactions between the peptides and the solid phase. However, SCX resolution is limited compared to reversed phase chromatography and will thus limit the suitability of this technique for complex mixtures.⁴⁰⁵

Reversed-Phase Chromatography (RPLC). Reversed-phase chromatography is the most commonly used chromatographic

technique which separates molecules in solution having neutral pH based on their hydrophobicity. The separation occurs on the basis of the partition coefficient of analytes between the mobile phase and the hydrophobic stationary phase. Highly polar peptides elute before the ones having less polarity because of the strong interaction with the hydrophobic functional groups forming a layer similar to a liquid around the silica resin.⁴⁰⁶ RPLC has been widely used in separation of peptides because of its compatibility with gradient elution and aqueous samples and its retention mechanism, which modulates separation owing to changes in the properties like pH, additives and organic modifier.⁴⁰⁷ Numerous factors influence the capacity of chromatographic peaks, such as temperature, column length, stationary phase, particle size, mobile-phase ion-pairing reagent, mobile-phase modifier and gradient slope.⁴⁰⁸ Usually online RPLC is done at acidic pH to ensure peptide ionization, but it can be paired with offline high pH RPLC and multiple fraction concatenation to produce orthogonal separation due to altered ionization of amino acids changing peptide hydrophobicity.⁴⁰⁹

Hydrophilic Interaction Liquid Chromatography (HILIC). Inverse-gradient chromatography was the forerunner to HILIC.⁴¹⁰ HILIC is similar in its principle to normal-phase chromatography where the stationary phase is polar and the initial solvent conditions are nonpolar. Gradient elution in HILIC is accomplished by increasing the polarity of the mobile phase, by decreasing the concentration of organic solvent, i.e., in the “opposite” direction compared to RPLC separations. With charged HILIC stationary phases there is also a possibility of increasing the salt or buffer concentration during a gradient to disrupt electrostatic interactions with the solute.^{411,412} Thus, the peptides with less polarity elute before the more polar peptides. It is used for the separation of hydrophilic peptides and polar analytes.⁴¹³ This separation is achieved by a stationary phase that is hydrophilic in nature, for example: cyano-, diol-, amino- bonded phases,⁴¹⁴ and an organic and hydrophobic mobile phase.⁴¹¹ HILIC can also be used for enrichment and targeted proteomic analysis of PTMs,

such as glycosylation, N-acetylation and phosphorylation, which alters the polarity of peptides and therefore also their retention on HILIC.⁴⁰⁶

Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC). ERLIC is a method based on use of a weak anion exchange column operated at low pH with high organic solvent enabling isocratic elution.⁴¹⁵ Acidic peptides are retained by electrostatic interaction, basic and neutral peptides are retained through hydrophilic interaction made favorable by high organic solvent. This improves retention of acidic peptides and reduces retention of basic peptides compared to normal HILIC.⁴¹⁶

Isoelectric Focusing (IEF). IEF is a type of high-resolution (HR) electrophoresis used for the separation as well as concentration of peptides that are amphoteric in nature on the basis of their isoelectric point (pI) using a solution without buffer consisting of either carrier ampholytes or a gel with immobilized pH gradient (IPG). After IEF separation, the separated amphoteric peptides in the liquid phase are recovered for further analysis by RPLC-MS/MS.⁴¹⁷ IEF reveals peptide pIs, which can improve validation of peptides during the database search step.⁴¹⁸ IEF is useful for label-free⁴¹⁹ and stable isotope labeling experiments.⁴¹⁸ IEF and gel-based separations have fallen out of favor in the last decade due to improvements in liquid chromatography.

■ LIQUID CHROMATOGRAPHY (LC)

Chromatography is the physical sorting of a mixture of molecular species that are dissolved in a mobile phase through the strength of binding, or affinity, to the chromatographic column's stationary phase.⁴²⁰ The mobile phase is pressure driven through the column and molecular species, or analytes, that have a strong affinity to the stationary phase are retained, or slowed, while those with a weak affinity pass through quickly. Thusly the analytes are separated by order of elution from the column. Chromatography can exploit most physical properties of the analytes, including ionic charge (anion/cation exchange chromatography), hydrogen binding (hydrophilic interaction), and size (size exclusion chromatography, capillary electrophoresis). In some chromatographic separations the mobile phase composition is adjusted by mixing two or more buffers at different ratios to influence the strength of affinity of individual analytes to the stationary phase and exquisitely regulate retention.

Mass spectrometers suffer from ion suppression, a phenomenon where the over-abundance of one or a few species within the ion population entering the mass spectrometer masks the presence of less abundant species.⁴²¹ Complex biological samples, such as tissue, cell lysate, or physiological fluids contain a wide dynamic range of molecule concentrations that span many orders of magnitude. The physical separation of analytes from biological samples by LC reduces the complexity of the ion population presented to the mass spectrometer at a given time, thus allowing the instrument to carry out the necessary fragmentation scans to identify and quantify the detectable species. Therefore, one major benefit of LC is that it allows detection of low abundant analytes in other elution windows.

The field of proteomics predominantly separates peptides using reversed phase liquid chromatography.^{422–424} Reversed stationary phase is most commonly composed of microscopic (1–3 μm) silica beads coated with covalently bound long (e.g. C18) hydrophobic alkyl chains. The hydrophobic side chains

of certain residues and the peptide backbone bind to this stationary phase through non-polar interactions. These interactions are strong in an aqueous solvent but are disrupted when the organic composition of the solvent is increased. Thus, in a reversed phase separation the proportion of non-polar, or organic, solvent in the mobile phase is gradually increased to release analytes from the stationary phase based on the strength of hydrophobic binding: weakly bound hydrophilic analytes elute with a low organic level in the mobile phase and strongly bound hydrophobic analytes only elute when the organic composition reaches a higher percentage. By far the most popular combination of solvents for peptide analysis is water and acetonitrile with dilute acid modifier (such as 0.1% formic acid or 0.5% acetic acid). The programmed rate at which the proportion of organic solvent is increased in the mobile phase is called the “gradient”, which you will often find described in the methods sections for reversed phase separations.

LC Considerations Related to Electrospray Ionization (ESI)

LC is paired to MS through ESI, and LC parameters greatly influence ESI. The analytes are eluted in a liquid mobile phase and must be released into the gas phase as charged ions for detection by mass spectrometry. This is achieved by spraying the eluent from the chromatographic separation through a narrow nozzle under a high voltage potential (1000–4000 V) between the nozzle, or emitter, and the mass spectrometer inlet. The eluent is sprayed as a mist of small charged droplets that explode into smaller droplets as the solvent evaporates and the repelling columbic force of the charged analytes increases.⁴²⁵ The droplets become progressively smaller until individual analyte molecules are ejected. The ejected analytes are ionized by the retained charge and can thus be manipulated by the electric fields in the mass spectrometer to measure their mass and perform the necessary fragmentations to elucidate structure.

The chromatographic flowrate (the volume of mobile phase driven through the chromatographic column per unit time, e.g., $\mu\text{L}/\text{min}$) dictates the efficiency of electrospray ionization (proportion of analytes eluting from the column that are ionized and into the gas phase) and is thus a key consideration for sensitivity of analysis.⁴²⁶ Reduced flowrates generate smaller droplets which degrade into ejected charged analytes rapidly, thus resulting in more detectable analytes and higher ionization efficiency. Electrospray ionization efficiency is also aided by an inert sheath gas, high temperature, and reduced pressure between the nozzle and ion lensing elements, thus decent sensitivity can still be achieved at high flowrates. For more detailed discussion of ionization, see the “Ionization” section.

Quality Attributes of Chromatographic Separation

The quality of chromatographic separation defines the number of analytes that are identified and quantified by LC-MS analysis. The theory around chromatographic separation was developed when LCs were paired with spectrophotometer detectors that only measure the combined signal intensity from all co-eluting analytes. The ability of MS to simultaneously detect the masses of individual components re-defines the significance of certain LC attributes. For those looking for mathematical descriptions of chromatographic quality, refer to the “Van Deemter equation”, which we do not cover here to maintain simplicity.⁴²⁷ The following attributes are the most important to consider in LC-MS.

Chromatographic Resolution. Chromatographic resolution is defined as the ability to fully resolve adjacent chromatographic peaks containing analytes with nearly equal affinities to the solid phase. In mass spectrometry, analytes are distinguished by mass even if they are not resolved by LC. Thus in LC-MS, the more relevant, but closely related concept is the peak width at the half maximum (FWHM). A low FWHM indicates a sharp elution peak. In a sharp peak the entirety of the analyte population is electrosprayed into the mass spectrometer in a short time thus increasing the signal. Low FWHM of high abundance species also confines their ionization suppression to narrow time windows, which means a lower number of co-eluting analytes are hidden. Conversely, high FWHM means that the analyte signal is spread out over time, thus reducing sensitivity. Furthermore, at a high FWHM, high abundance species mask analytes through ion suppression over a larger portion of the separation.

Peak Capacity. Peak capacity is defined as the maximal number of peaks that ideally can be completely resolved in a pre-established time window. A long separation in which FWHM remains low would have a large peak capacity and thus allow identification of many species. Unfortunately increasing the length of a reversed phase gradient also increases the FWHM due to an increase in diffusion, which results in a diminishing return for longer analytical methods. A longer separation provides more time and opportunities for the mass spectrometer to sample each analyte to acquire fragmentation spectra required for identification and the selection of gradient length should consider both the desired throughput and the speed of the MS data acquisition strategy.

Reproducibility and Robustness. Reproducibility is defined as the ability to repeatedly obtain the same measurement for the same analytes each time that the analysis is repeated. In liquid chromatography this means that each analyte should elute at nearly the same retention time (the time elapsed since the start of the analysis until the analyte's elution from the chromatographic column) with the same peak width. Robustness is the ability of the system to maintain reproducible performance despite nonoptimal conditions. The most typical obstacles to robustness are mechanical wear of the system components and the analytical column, fouling of the system by contaminants introduced in the samples, and clogging due to accumulation of contaminants. High flow methods tend to be more robust due to reduced impact of pump and plumbing configurations and changes in dwell volumes, and the wider bore of the components used is more resilient to clogging. However, higher flowrate comes at the cost of reduced sensitivity due to reduced ionization efficiency at higher flow rates and increases in the overall peak volume at constant sample loading, thus nanoflow (100–300 nL/min flowrate) chromatography remains a widely utilized strategy in proteomics. For applications where sample is not limited, slightly higher amounts of applied samples can take advantage of robustness of higher flow rates in the microflow range using newer optimized electrospray sources.⁴²⁸

Throughput and Instrument Utilization. Throughput is the number of samples that are analyzed in a given timeframe, for example samples per day. High throughput is required to analyze thousands of samples that truly represent biological diversity in a timely manner. Increasing throughput means less data are collected for individual samples. Furthermore, many steps in the LC process are required for sample analysis in which no useful data is collected including sample injection,

and system cleaning and equilibration, which reduce the ratio of data collected to instrument operation time, or instrument utilization. The ability to perform these steps while a different sample is analyzed, or parallelization, increases instrument utilization and the amount of data collected by several minutes which is a significant increase when several samples are analyzed per hour.

Trapping and Precolumns

Trapping and pre-columns are short chromatographic columns that are used to increase robustness of an LC-MS system. A pre-column is connected directly to the front of the analytical column and is intended to be disposable and to absorb contaminants and protect the analytical column. The trapping column is connected indirectly to the analytical column through a valve. The valve can be switched to redirect the flow through the trapping column away from the analytical column. This allows analytes to be loaded on the trapping column while analytes that are hydrophilic and poorly retained are washed away and do not contaminate the analytical column or the mass spectrometer. This process is referred to as desalting, and once it is complete, the valve configuration is changed to connect the trapping column to the analytical column, and analytes captured on the trapping column can be eluted off the trap and through the analytical column for analysis by MS. Certain trapping columns can be operated in both directions, which allows aggregates to be flushed away when the trapping column is cleaned in the reverse direction. Additionally trapping columns are shorter and have less backpressure so they can be loaded with sample quickly at a fast flowrate. Whereas loading the sample directly on the analytical column requires a slower flowrate. Two trapping columns can be used in tandem to provide parallelization, while one trapping column is cleaned and loaded with samples the second trapping column is in line with the analytical column analyzing the sample that was loaded on it in the previous run.^{429,430}

Multidimensional LC

Depth of profiling has previously been increased by combining two or more orthogonal LC separations. Orthogonal in this context means that each separation sorts the analytes into different populations.⁴³¹ For example, strong cation exchange (SCX) separate analytes based on positive charge, and when paired with reversed phase chromatography, results in higher peak capacity and more analytes identified. The first highly popular method was multidimensional protein identification technology (MudPIT), which used online separation by SCX followed by C18 reversed phase.⁴³² However, the resolution of peptide separation by SCX is low, leading to the presence of peptides in many fractions. The currently accepted most popular method for two-dimensional separation combines iterative reversed phase at different high and then low pH to sort analytes by changes in hydrophobicity due to changes in amino acid side chain ionization. Although the separations are not entirely orthogonal, multiple fraction concatenation across the high pH elution can produce entirely orthogonal peptide sets.⁴³³ In recent years the focus of proteomics has shifted from deep profiling of fewer samples to rapid profiling of large cohorts. Thus, lengthy multidimensional methods have been replaced with single shot experiments only using one dimension of high resolution reversed phase separation.⁴³⁴ Peak capacity is often regained by using ion mobility spectrometry (separation of ionized peptides in the gas phase).

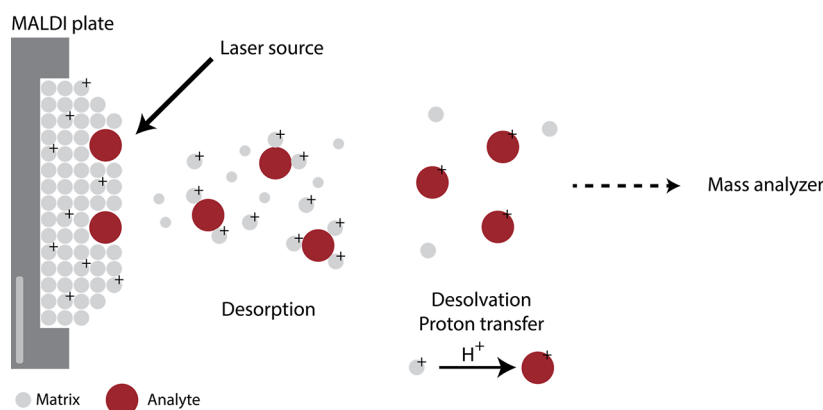


Figure 6. MALDI. The analyte-matrix mixture is irradiated by a laser source, leading to ablation. Desorption and proton transfer ionize the analyte molecules that can then be accelerated into a mass spectrometer.

■ PEPTIDE IONIZATION

As early as the late 1950s, derivitization reagents were used to make peptides volatile enough for electron impact ionization analysis.⁴³⁵ Eventually this led to GC-MS analysis of derivatized peptides for sequencing.⁴³⁶ In the early 1980s, fast atom bombardment (FAB) enabled peptide ionization and sequencing by MS/MS,⁴³⁷ but difficulty interfacing FAB with LC limited its utility.⁴³⁸ New soft ionization techniques called matrix-assisted laser desorption (MALDI) and electrospray ionization (ESI) were applied to peptides around 1990, which revolutionized the field of proteomics by making high throughput ionization of peptides easy. These two techniques were so impactful that the 2002 Nobel Prize in Chemistry was co-awarded to John Fenn (ESI) and Koichi Tanaka (MALDI) “for their development of soft desorption ionization methods for mass spectrometric analyses of biological macromolecules”.⁴³⁹

MALDI

The term “Matrix-assisted laser desorption” was coined by Hillenkamp and Karas in 1985, although this original paper only applied the technique to dipeptides.⁴⁴⁰ It was Koichi Tanaka who first applied this idea to proteins above 10,000 Daltons in size and published a paper in the Proceedings of the 2nd Japan-China Joint Symposium on Mass spectrometry in 1987 (Tanaka, K., Ido, Y., Akita, S., Yoshida, Y. and Yoshida, T. (1987) Detection of high mass molecules by laser desorption time-of-flight mass spectrometry. Proceedings of the 2nd Japan-China Joint Symposium on Mass spectrometry, 185-187), and then in a follow-up paper published in 1988.⁹ A few months later, Karas and Hillenkamp also demonstrated MALDI applied to proteins above 10kDa with MALDI.⁴⁴¹ This resulted in some controversy about who should have won the Nobel prize⁴⁴² as it was felt by the community that Hillenkamp and Karas had provided the technology several years before but it was Koichi Tanaka that was the first to apply the MALDI technology to proteins a year before Hillenkamp and Karas.

MALDI Mechanism. MALDI first requires the peptide sample to be co-crystallized with a matrix molecule, which is usually a volatile, low molecular-weight, organic aromatic compound (Figure 6). Some examples of such compounds are cyno-hydroxycinnamic acid, dihydrobenzoic acid, sinapinic acid, alpha-hydroxycinnamic acid, and ferulic acid.⁴⁴³ Subsequently, the analyte is placed in a vacuum chamber in which it is irradiated with a laser, usually at 337 nm.⁴⁴⁴ This laser energy

is absorbed by the matrix, which then transfers that energy along with its free protons to the co-crystallized peptides without significantly breaking them. The matrix and co-crystallized sample generate plumes, and the volatile matrix imparts its protons to the peptides as it gets ionized first. The weak acidic conditions used as well as the acidic nature of the matrix allows easy exchange of protons for the peptides to get ionized and fly under the electrical field in the mass spectrometer. These ionized peptides generally form the metastable ions, most of them will fragment quickly.⁴⁴⁵ However, it can take several milliseconds and the mass spectrometry analysis can be performed before this time. Peptides ionized by MALDI almost always take up a single charge and thus observed and detected as $[M+H]^+$ species.

According to PubMed, the number of publications related to MALDI peaked in 2013 and has been steadily declining. Concurrently, the usage of MALDI for bottom-up proteomics has subsided in favor of the better depth and throughput possible from using ESI. MALDI is still widely used for mass spectrometry imaging of proteins and metabolites.⁴⁴⁶

Electrospray Ionization

ESI was first applied to peptides by John Fenn and coworkers in 1989.⁸ Concepts related to ESI were published at least as early as 1882, when Lord Rayleigh described the number of charges that could assemble on the surface of a droplet.⁴²⁵ ESI is usually coupled with reverse-phase liquid-chromatography of peptides directly interfaced to a mass spectrometer. A high voltage (~ 2 kV) is applied between the spray needle and the mass spectrometer (Figure 7). As solvent exits the needle, it forms droplets that take on charge at the surface, and through a debated mechanism, those charges are imparted to peptide

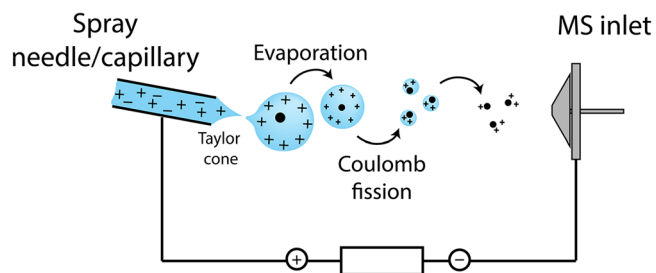


Figure 7. Electrospray ionization. Charged droplets are formed; their size is reduced due to evaporation until charge repulsion leads to Coulomb fission and results in charged analyte molecules.

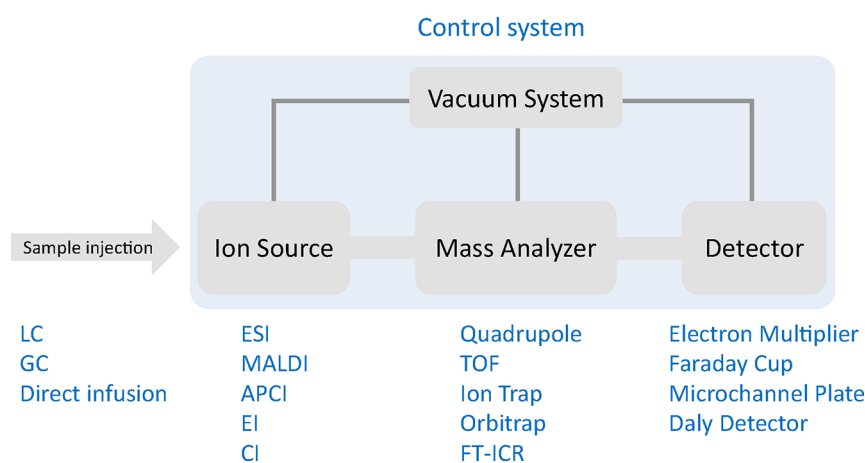


Figure 8. Diagram of typical mass spectrometer modules. Systems must have an ion source, mass analyzer, detector, vacuum system, and control system.

ions. The liquid phase is generally kept acidic to help impart protons easily to the analytes.

Tryptic peptides ionized by ESI usually carry one charge on the side chain of their C-terminal residue (Arg or Lys) and one charge at their N-terminal amine. Peptides can have more than one charge if they have a longer peptide backbone, have histidine residues, or have missed cleavages leaving extra Arg and Lys. In most cases, peptides ionized by ESI are observed at more than one charge state. Evidence suggests that the distribution of peptide charge states can be manipulated through chemical additives.⁴⁴⁷

Electrospray Mechanism. The main goal of ESI is the production of gas-phase ions from electrolyte ions in solution. During the process of ionization, the solution emerging from the electrospray needle or capillary is distorted into a Taylor cone and charged droplets are formed. The charged droplets subsequently decrease in size due to solvent evaporation. As the droplets shrink, the charge density and Coulombic repulsion increase. This process destabilizes the droplets until the repulsion between the charges is higher than the surface tension and they fission (Coulomb explosion).^{448,449} Typical bottom-up proteomics experiments make use of acidic analyte solutions which leads to the formation of positively charged analyte molecules due to an excess presence of protons.

MASS SPECTROMETERS

Mass Spectrometry

Mass spectrometry is a science of ions; mass spectrometers serve as sophisticated instruments for determining the masses of compounds and elements. Mass spectrometers can therefore be likened to an ultra-precise weigh scale that can differentiate mass variations down to a single electron, or even lighter. Since J. J. Thomson's initial exploration in 1912, the field of mass spectrometry has undergone numerous improvements, spanning from isotope assessment to the interpretation of biomacromolecules,⁴⁵⁰ all thanks to the combined efforts of diverse fields like chemistry, physics, electronic engineering, and computer science. With the rapid improvement of sensitivity, mass resolution, tandem mass spectrometry methods and ion dissociation methods, mass spectrometers have evolved as a core tool for proteomic (and metabolomic) analysis. It is precisely the widespread application of mass

spectrometry in proteomics analysis that has given rise to more instrument manufacturers and a greater diversity of mass spectrometer types. This also brings a happy annoyance to many beginners or researchers in other fields who have no background in mass spectrometry: which manufacturer and which type of mass spectrometry should I choose to analyze my samples? Here, to help new learners build a basic understanding faster, we will briefly introduce some basic concepts, common types of mass spectrometers, and their suitable application scenarios.

Mass Spectrometer Structure and Basic Principles

The fundamental principle of mass spectrometry revolves around specific physical processes that can be described by various mathematical formulas. Since this article serves as a guide for those new to the field, particularly those from a biology background, we've chosen to steer clear of delving too deeply into intricate mathematical and physical explanations. However, for those keen on a deeper understanding, we've included references pertaining to these foundational principles. Our focus lies on introducing fundamental concepts and outlining the typical workflow in mass spectrometry.

The process of mass spectrometry (MS) is to generate gas phase ions from compounds in samples by any suitable method, to separate these ions by their mass to charge (m/z) ratio, and then detect them by their respective m/z and abundance. The successful implementation and demonstration of this process requires participation of five fundamental systems (Figure 8):

The Ion Source. The ion source is where gas phase ions are generated. As discussed in the prior chapter, for proteomic analysis, soft ionization methods such as ESI and MALDI are the most widely applied techniques.^{8,9} Additional ionization methods used to generate ions for mass spectrometry of small molecules include atmospheric pressure chemical ionization (APCI), atmospheric pressure photo ionization (APPI), electron ionization (EI) and chemical ionization (CI).^{451,452}

The Mass Analyzer. The mass analyzer is where gas phase ions are separated according to their m/z ratio based on physical principles. There are several types of mass analyzers applied in mass spectrometry, including the quadrupole, linear ion trap and three-dimensional ion trap, orbitrap, Fourier transform-ion cyclotron resonance (FT-ICR), time-of-flight (TOF), and the magnetic sector analyzers,^{453,454} each with

unique advantages and applications (Table 4). For proteomic analysis, tandem mass spectrometry, which involves combining two or more stages of mass analysis, is typically used to achieve precursor selection, structural analysis, and improved sensitivity.⁴⁵⁵ The mass analyzer is the core component of a mass spectrometer, it is also the most important factor that we need to take into consideration when choosing a mass spectrometer for a specific project.

The Detector. The detector is where ions are detected and their respective m/z values and abundances are recorded, generating a mass spectrum. Common types of ion detectors are listed in Table 5, including the Electron Multiplier (EM), Photomultiplier Tube (PMT), Microchannel Plate (MP), and Faraday Cup (FC), along with a summary of their strengths and limitations. It is worth noting that Orbitrap and FT-ICR mass analyzers don't use conventional detectors as listed above. Instead, these analyzers detect an image current produced by oscillating ions.^{456–458} In both mass analyzers, the detector is essentially measuring an electrical current (or more accurately, a voltage that's proportional to the current) that's induced by the motion of the ions. This signal is then processed to extract the frequencies of oscillation and Fourier-transformed into a mass spectrum, which is quite different from other types of detectors that count individual ions or particles striking a surface. Longer transients generate higher resolution spectra.

The Vacuum System. This is designed to maintain a high-vacuum environment for ion transmission inside the instrument. The vacuum system consists of different type of pumps including roughing vacuum pumps (rotary vane pumps, scroll pumps) and high-vacuum pumps (turbo molecular pumps, diffusion pumps). Maintaining a high vacuum is essential to reduce collisions between analyte ions and inert gas molecules during their transmission from one region of the mass spectrometer to another, or during oscillations within a mass analyzer. Collisions within the vacuum chamber may lead to unstable ion trajectories, unwanted fragmentation, poorer transmission efficiency, in turn leading to lower resolving powers and poorer sensitivities. Even so, some inert gas is intentionally plumbed into the mass spectrometer either for collisional activated dissociation (CAD), typically with nitrogen, helium, or argon, or to dampen ions' energy. FT-ICR and Orbitrap mass analyzers require higher vacuum in the 10^{-9} to 10^{-11} Torr range, while TOFs require medium vacuum in the 10^{-7} to 10^{-8} Torr range, and quadrupole and ion trap instruments require a relatively low vacuum in the 10^{-5} to 10^{-6} Torr range.

The Control System. This is needed to regulate and coordinate the various parts of the mass spectrometer to ensure seamless functioning. This typically includes ion source control, mass analyzer control, detector control, data acquisition control, interfacing with auxiliary systems (such as a liquid chromatograph and gas chromatograph), and modules for instrument diagnostics and calibration.

Types of Mass Spectrometers Used for Proteomics

Typically, mass spectrometers are named based on the abbreviations of their principal or tandem mass analyzers. This naming convention stems from the fact that the mass analyzer forms the core component of a mass spectrometer, and it also dictates key performance attributes such as mass resolution, scanning speed, sensitivity, and cycle time. These performance metrics, in turn, determine what type of analysis

Table 4. Common Mass Analyzers

Type	Acronym	Principle	Characteristics
Time-of-flight	TOF	Time dispersion of a pulsed ion beam; separation by the time it takes for ions to travel a fixed distance	High-speed analysis, large mass range and good sensitivity. Suited for fast data acquisition and high-throughput applications. Modern TOF systems usually can achieve mass resolution well over 10,000 ($m/\Delta m$) or even higher.
Linear quadrupole	Q	Continuous ion beam in linear radio frequency quadrupole field; separation due to instability of ion trajectories; rods have applied alternating DC and RF	High transmission efficiency, simple design, good sensitivity, and tunable mass range; relatively low mass resolution ranges from several hundreds to a thousand; often used in tandem mass spectrometry (MS/MS) experiments
Quadrupole ion trap	QIT	Traps ions by electromagnetic fields; separation in three-dimensional radio frequency quadrupole field by resonant excitation	Efficient for fragmenting ions and structural elucidation, higher sensitivity, and relatively compact which good for benchtop instruments. Relatively a low mass resolution around 1000–3000.
Fourier transform-ion cyclotron resonance	FT-ICR	Traps ions in a strong magnetic field by Lorentz force; separation by cyclotron frequency; image current detection and Fourier transformation of transient signal	Ultimate high mass resolution (over 2,700,000 with 21 tesla magnets), making it ideal for elemental and isotopic analysis. Large size, low speed, and expensive in terms of both initial purchase cost and ongoing operation and maintenance costs.
Orbitrap	Orbitrap	Axial oscillation in inhomogeneous electric field; detection of frequency after Fourier transformation of transient signal	Extremely high resolution and accuracy (up to 1,000,000), capable of resolving complex mixtures with high sensitivity. Relatively low speed, expensive in terms of both initial purchase cost and ongoing operation and maintenance costs. Need high vacuum.

Table 5. Common Detectors

Type	Principle	Characteristics
Electron Multiplier	Amplifies signals by utilizing a sequence of dynodes that emit secondary electrons when struck by an incident electron, creating a cascading effect. This results in an amplified output current at the final anode, proportional to the intensity of the initial signal.	Very good signal amplification to even one electron (may cause more noise dependent on gain), high sensitivity, need high vacuum and high voltage, expensive. Limited dynamic range, finite lifespan and need to be replaced periodically.
Faraday Cup	Charged particles, such as ions or electrons, enter the cup and transfer their charge to it, causing a change in electric potential that can be measured over time to infer the number of particles.	Suitable for particles and charge state detection. Simplicity and robustness. Wide dynamic range, no need for high voltage and high vacuum. Lower sensitivity. Sensitive to Secondary Emission directional sensitivity (direction of incoming particles).
Microchannel Plate	Similar to electron multiplier, a two-dimensional matrix or "plate" of many tiny, parallel, hollow channels made from a type of glass that can generate secondary electron emissions upon incident particles striking the channel walls. These secondary emissions create an electron avalanche down the channels and amplifies the original signal.	Signal Amplification (Not as good as electron multiplier, but lower noise), Spatial Resolution ability, shorter life expectancy due to channel aging and depletion of the secondary emission material, smaller and cheaper than electron multiplier.
Daly Detector	Directing ions onto a surface (Doorknob) to trigger the emission of electrons, which are then accelerated towards a phosphor screen to produce photons, that are subsequently detected and amplified by a photomultiplier tube, thereby converting the ion signal into a measurable electrical signal.	High gain, ruggedness, wide dynamic range, suitable for high mass and high energy ions. Limited mass resolution, larger size and need high voltage, finite lifespan.

we can conduct, its speed and its accuracy. Next, we will focus on introducing several classic tandem mass spectrometry types commonly used in proteomics.

Triple Quadrupole (QqQ). Triple quadrupole mass spectrometer (often abbreviated as QqQ, QQQ, TQ, or TQMS) is a type of tandem mass spectrometer where three quadrupole mass analyzers are combined in series (Figure 9).

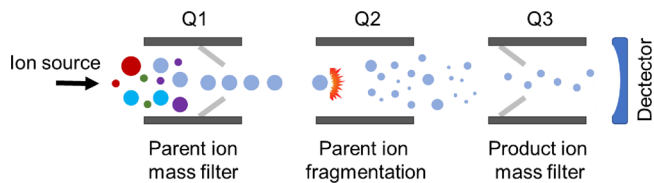


Figure 9. Schematic diagram of typical QqQ system. Three quadrupoles enable precursor selection, fragmentation, and fragment ion selection.

Each quadrupole is essentially a set of four parallel metal rods to which radio frequency (RF) and direct current (DC) voltages are applied to each opposing pair of rods. The QqQ operates in a synchronized manner to isolate ions of interest (according to the Mathieu function) in the first quadrupole, induce fragmentation with inert gas in the second, and then detect the resulting product ions in the third quadrupole. Specifically, the first quadrupole (Q1) is a mass filter, where ions of a specific m/z are selected from the incoming ion beam. This is achieved by adjusting the voltage applied to the pair of rods within the quadrupole, allowing ions with a particular m/z value to pass through while deflecting others. The second quadrupole (Q2), also known as the collision cell, is where selected ions from Q1 are fragmented into product ions. This fragmentation happens due to the collisions between inert gas molecules (nitrogen, argon, or helium) and ions, which causes the ions to break up (fragment) into smaller pieces (fragment ions). For more detail about peptide fragmentation, see the [Tandem Mass Spectrometry section](#). This process is known as collision-induced dissociation (CID).^{459,460} The Q2 is usually only subjected to RF potential and does not filter ions; instead, it transmits the product ions to the third quadrupole. In some tandem mass spectrometry, hexapoles or octupoles are also used to replace a quadrupole as the collision cell. Lastly, the third quadrupole (Q3) acts as a secondary mass filter, similar to Q1, but with the purpose of selecting specific fragment ions produced in the collision cell while excluding other ions. The chosen ions are then directed to the detector, where their abundance is measured (Figure 9). This process, involving precursor ion selection, precursor ion fragmentation, and product ion detection, is a general operating principle in tandem mass spectrometry and determines what kind of scan mode you can utilize. While discovery-based proteomics approaches can be performed on triple-quadrupole systems, the data produced would be inferior to competing high resolution options. Instead, QQQ instruments are widely used for targeted proteomics by operating in selected reaction monitoring (SRM) mode, which is also referred to as multiple reaction monitoring (MRM).⁴⁶¹ QQQ instruments are available from all major vendors, including the QTRAP (Sciex), TSQ (Thermo), Xevo TQ-XS (Waters), LCMS-8050 (Shimadzu), 6475 (Agilent), and EVOQ LC-TQ (Bruker). A key characteristic and advantage of QqQ is the

flexibility of choosing various scan modes,^{460,462,463} including the following.

Product Ion Scan. Q1 is set to filter a specific precursor ion, which is then fragmented in Q2. Q3 scans the full range of product ion masses. This mode is usually used to identify the structure of a particular compound.

Precursor Ion Scan. Q3 is set to filter a specific product ion. Q1 scans the full range of precursor ions, that when fragmented in Q2, yield the selected product ion. This mode is used to find compounds that yield a specific fragment ion, which can be particularly useful when looking for compounds with a common structural motif.

Neutral Loss Scan. Both Q1 and Q3 scan the full range of ions, but with a mass difference equal to a specific “neutral loss”. This mode is used to identify compounds that, when fragmented, lose a specific neutral molecule.

Multiple/Selected Reaction Monitoring (M/SRM). Both Q1 and Q3 are set to filter specific ions (precursor and product, respectively). This highly selective mode is used for quantitative analysis of specific compounds, offering excellent sensitivity and specificity.^{464,465}

The triple quadrupole mass spectrometer is a highly versatile instrument, capable of both qualitative and quantitative analysis. Enke and Yost at Michigan State University developed the first working triple-quadrupole mass spectrometer in the late 1970s.⁴⁶⁶ QqQ is particularly well-suited for targeted quantitative analysis due to its high sensitivity, selectivity, and dynamic range, which has made it a go-to instrument in areas such as drug metabolism studies, environmental monitoring, food safety analysis, pharmaceuticals, and clinical diagnostics.^{467–470}

However, quadrupoles suffer from inherent limitations in mass resolution due to the constraints of principles and precision in mechanical manufacturing. Consequently, QQQ instruments face difficulties in accurately identifying unknown molecules within complex mixtures and thus not appropriate for applications like structure analysis and biomarker discovery.

Q-TOF. Even though quadrupoles face difficulties in accurately identifying unknown peptides within complex mixtures due to its mass resolution, they serve effectively as mass filters, making them an excellent choice for combining with other high-resolution mass analyzers to form tandem mass spectrometry systems. One commonly used approach is Quadrupole-Time-of-Flight Mass Spectrometer (Q-TOF-MS), a ‘hybrid’ device, integrating quadrupole techniques with a time-of-flight mass analyzer. W. E. Stephens constructed and published the design of the first time-of-flight (TOF) analyzer in 1946.^{471,472} The principle of TOF is quite straightforward: ions of different m/z are imparted with the same initial kinetic energy ($E = Uq = \frac{1}{2}mv^2$) and then separated over time as they travel along a field-free drift path of known length. If all ions begin their flight simultaneously, or at least within a short enough time span, the lighter ions will reach the detector before the heavier ones due to their faster velocity (V).⁴⁷³ Based on this principle, the m/z of different ions can be calculated according to the order in which they reach the detector. Similarly, we can easily conclude that the longer the drift path, the higher of the mass resolution can reach if keep the response time of detector the same. In fact, in pursuit of higher mass resolution, researchers have indeed built time-of-flight (TOF) drift tubes that are tens of meters long. However, apparently, this is not practical for wide application in a regular lab place. An alternative way to expand drift length

and achieve higher resolution is to apply reflector (often called a reflectron). The principles and advantages of using a reflector can be summarized as follows.

Under ideal circumstances within a TOF mass spectrometer, ions sharing the same m/z would reach the detector concurrently post-acceleration, thus generating a sharp peak on the mass spectrum. However, the inherent oscillation path variability of ions within the mass spectrometer makes it challenging to maintain uniform initial kinetic energy amongst all ions, leading to peak broadening and a substantial reduction in mass resolution. The reflector is designed to rectify this issue. Comprising a series of electrodes that set to different voltages, the reflector generates a retarding electric field that reverses ion trajectories back through the flight tube. Notably, the reflector is engineered such that ions carrying lower kinetic energy delve less into the reflector and have a reduced flight path, while those with higher kinetic energy permeate more deeply and follow a longer flight path. This equalizes the variances in initial kinetic energy, enabling ions of the same m/z to hit the detector almost simultaneously, thereby enhancing the resolution of TOF.

Furthermore, the usage of reflector effectively expands the flight path length within the same physical confines, resulting in superior ion separation and consequently, higher resolution. This reflection comes at the cost of some ion loss, and therefore some sensitivity loss. As such, reflecting TOFs are the basis of most commercial instruments currently in use.

The construction of a Q-TOF bears significant resemblance to a triple-quadrupole mass spectrometer, with the critical distinction that the third quadrupole has been replaced by a time-of-flight tube. Figure 10 delineates the schematic of a typical Quadrupole-Time-of-Flight (Q-TOF) mass spectrometer, which comprises three fundamental components:

Quadrupole Mass Analyzer (Q). This part of the instrument is basically the same to the Q1 in QqQ, which selects specific m/z values to pass through by applying a combination of DC and RF voltages across the rods.

Collision cell. Here, selected ions undergo collision-induced dissociation (CID) by interacting with a neutral gas, leading to their fragmentation into smaller constituents. This process yields structural information about the original molecules. Usually, quadrupole, hexapole, or even octopoles are used as the collision cell for better focusing and transporting.

Time-of-Flight (TOF) Mass Analyzer. Upon exiting the collision cell, the fragmented ions are reaccelerated into the ion modulator region of the time-of-flight analyzer. There, they undergo pulsing by a strong electric field (typically 20 kV or higher) and get accelerated to a field free drift tube, and then reflected to the detector.

TOFs generally offer mass resolutions surpassing 50,000, rendering it a reliable instrument for identifying unknown compounds. Moreover, the rapid travel time of ions in the vacuum tube (at the nanosecond level) confers the Q-TOF with distinctive benefits in short gradient and high-throughput analyses.^{474–476} Another advantage of TOF is its broad mass range, which allows for the detection of large proteins, nanoclusters, and even large particles.^{477–479} However, it should be noted that due to ion numbers and detector limitations, mass resolution is typically difficult to maintain over a wide mass range.

Presently, Q-TOF related instruments are available from all leading instrument manufacturers, and the main models are listed below: Sciex: “TripleTOF 6600+”, “TripleTOF 5600+”

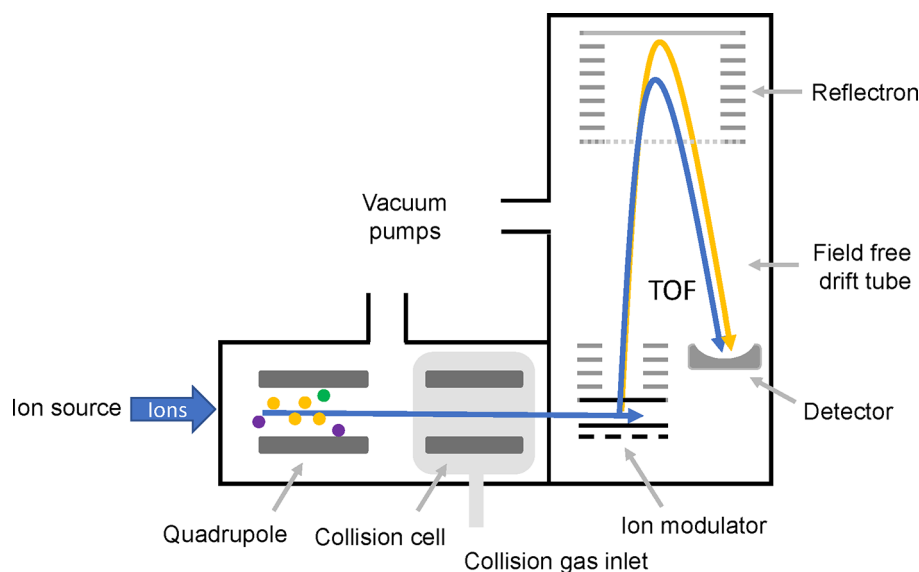


Figure 10. Schematic diagram of a typical quadrupole time-of-flight mass spectrometer. Like a QQQ, a Q-TOF will have two quadrupoles for selection and fragmentation followed by the TOF for the final higher resolution separation and detection.

System, and “X500R QTOF” System. Bruker Corporation: “Impact II”, “timsTOF” series, “microTOF-Q III”, “ultrafleXtreme-MALDI-TOF/TOF”, and “maXis II”. Agilent Technologies: “Agilent 6530 Accurate-Mass Q-TOF”, “Agilent 6545 Accurate-Mass Q-TOF”, and “Agilent 6550 iFunnel Q-TOF”. Waters Corporation: “SYNAPT G2-Si HDMS”, “Xevo G2-XS QToF”, and “SYNAPT XS”.

Q-Orbitrap

The orbitrap is a critical pillar in the field of proteomics. In the late 20th century, Russian scientist Alexander Makarov invented the Orbitrap,⁴⁸² which is a novel mass analyzer that operates based on the principle of electrodynamic ion trapping and Fourier Transform. The orbitrap consists of two main components: an inner spindle-like electrode and a coaxial outer barrel-like electrode (Figure 11A). The Orbitrap was based on the “Kingdon trap” invented in the 1920s⁴⁸⁰ and the Knight trap.⁴⁸¹ The ions are trapped in an orbit around the spindle electrode due to the electrostatic attraction. Once inside, the ions begin oscillating along the central axis of the device, or “orbiting”, due to the electric field formed by the inner and outer electrodes. The oscillation frequency of an ion is inversely proportional to the square root of its mass-to-charge ratio. The frequency at which each ion oscillates induces an image current on the detector, which can be measured and transformed into a mass spectrum using Fourier transform.

The biggest difference between Orbitrap and other mass spectrometers (TOF, Q) is that it does not use ions to hit an induction device like an electron multiplier. One of the main advantages of the Orbitrap is its ultra-high mass resolution, often exceeding 240,000 or even higher. This gives the Orbitrap a significant superiority in the identification of unknown molecules such as peptides and metabolites.^{454,483} Moreover, Orbitrap spectrometers are also appreciated for their compact structure, small size, robustness, and reliability. Just like the Q-TOF, the Orbitrap is also usually used for tandem mass spectrometry. It is important to note that many Orbitraps are sold with a linear ion trap as a complementary detector, called a “tribrid” because these models contain a quadrupole, linear ion trap, and an orbitrap. Figure 11B

demonstrates a typical 2D schematic diagram of Q-Orbitrap. Ions first pass through an ion optics module, which consists of a high-capacity ion transfer tube (HCTT), an electrodynamic ion funnel (EDIF), and an advanced active beam guide (AABG). These are designed to capture ions, reduce ion losses, prevent neutrals and high-velocity clusters from entering the quadrupole, and increase sensitivity. The ions are then segmented by the quadrupole for precursor ion selection, and the selected ions are trapped by the ion-routing multipole for higher energy collisional dissociation. Finally, the fragmented ions are captured once again by the C-trap and injected into the Orbitrap batch-by-batch for accurate mass-to-charge analysis. Overall, this process still follows the logical sequence of precursor ion selection, precursor ion fragmentation, and fragment ion detection.

Compared to a TOF, one disadvantage of the Orbitrap is its longer cycle time (AGC pre-scan, ion injection, ion isolation, ion activation and mass analysis, usually >100ms), which is a negative factor for the currently favored short gradient, high-throughput analysis. Another minor flaw of Orbitrap is the challenge encountered when trying to pair it with MALDI. This primarily stems from the fact that MALDI uses a pulsed ionization technique, whereas the Orbitrap operates continuously. This mismatch can lead to inefficiencies and challenges in coupling the two techniques. At present, Orbitraps are widely used in almost all aspects of proteomics including biomarker discovery,⁴⁸⁴ post-translational modification (PTM) analysis,^{29,485} quantitative proteomics (LFQ, TMT, iTRAQ),^{22,23,486} protein–protein interaction studies,⁴⁸⁷ and structural proteomics.^{488,489} It can perform both top-down and bottom-up analyses owing to its broad mass range, and is suitable for both Data-Dependent Acquisition (DDA) and Data-Independent Acquisition (DIA) methods. Right now, the Orbitrap is still under patent protection and only one company, ThermoFisher, is allowed to manufacture related products. Classic models from ThermoFisher include Orbitrap Astral, Ascend Tribrid, Eclipse Tribrid, Fusion, Lumos, Exploris series (120, 240, 480) and Q Exactive series.

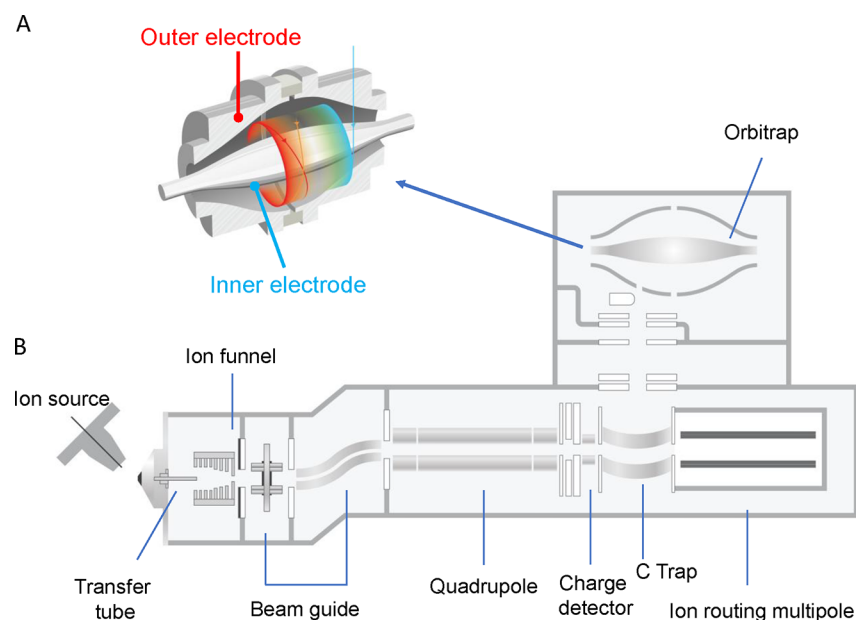


Figure 11. Schematic diagram of orbitrap. (A) Close up of an Orbitrap. (B) General schematic of complete Q-Orbitrap system.

Quadrupole Fourier Transform Ion Cyclotron Resonance (Q-FT-ICR)

The Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer is a type of mass spectrometry that uses magnetic fields to separate ions based on their mass-to-charge ratio. FT-ICR was first invented in 1974 by Alan G. Marshall and Melvin B. Comisarow from the University of British Columbia⁴⁹⁰ and is widely recognized for its high mass resolution and precision, making it a highly valuable tool in many scientific fields including proteomics, metabolomics, petroleum analysis, and environmental science. The central feature of an FT-ICR mass spectrometer is a superconducting magnet coupled with an ICR cell (Figure 12A). This magnet creates a strong and homogeneous magnetic field in which ions are injected. Once the ions are inside ICR cell, under the influence of the strong magnetic field, they follow a circular path with a very small orbital radius at a specific frequency directly proportional to their mass-to-charge ratio. At this point, no detectable image current signal is generated by detector plates located inside the ICR cell. To improve the signal, a voltage is applied by excitation plates and resonance occurs when the frequency of the strong magnetic field matches the cyclotron frequency of the ions. The ions absorb radio frequency energy, which increases the radius of their circular path, and consequently, the excited ions move closer to the detector plates and generate a current. The resulting signal is an oscillating pattern or a time-domain signal.

Similar to Orbitraps, this time-domain signal is then transformed into a frequency-domain signal using Fourier transform, hence the name Fourier Transform ion cyclotron resonance (ICR). The Fourier transformed data forms a mass spectrum where each peak corresponds to a specific ion present in the sample. One of the most important advantages of FT-ICR mass spectrometry is its exceptionally high mass resolution and mass accuracy, even for large and complex molecules. This enables precise identification and characterization of a wide range of compounds in complex mixtures.^{491,492} Moreover, FT-ICR mass spectrometry can be used for multiple stages of mass analysis (MSn), including

tandem mass spectrometry (MS/MS), providing detailed information about the structure of ions. Another significant benefit of FT-ICR is its broad mass range, making it possible to identify macromolecules like proteins for top-down proteomics.^{21,493}

Despite its advantages, FT-ICR mass spectrometry is not without challenges. The technique requires high-performance superconducting magnets, which are expensive for both initial purchase and further maintenance. This is because FT-ICR requires liquid nitrogen and liquid helium cooling systems to keep the magnet at a sufficiently low temperature to maintain its superconducting state. Moreover, the device demands high vacuum conditions and careful temperature control to maintain the stability of the magnetic field and the ion trajectories. A schematic representation of a Q-FT-ICR system is shown in Figure 12B. In congruence with the tandem mass spectrometers elucidated earlier, ions pass through an array of ion optics modules which designed for ion focusing and purification. Following this, the ions are selectively filtered by the first quadrupole. After this filtration, precursor ions undergo fragmentation in the collision cell, which can be a quadrupole, hexapole, or octopole. The fragmented ions are subsequently re-concentrated by the ensuing focusing lens. Ultimately, these fragmented ions are trapped, excited, and detected within the ICR cell. At present, commercial FT-ICR mass spectrometers are available in both Thermo Fisher Scientific ("LTQ FT Ultra" and "LTQ FT Ultra Hybrid" systems) and Bruker Daltonics ("solariX" and "apex" series).

Ion Mobility. In the context of omics research, a fundamental task is the separation, identification, and quantification of molecules in complex mixtures. Mass spectrometry alone can only provide two-dimensional data including mass-to-charge ratio and their intensity. Liquid chromatography contributes to the separation of compounds and further provides the third dimension of information, retention time (RT), which make LC-MS the "golden standard" for proteomic analysis.^{494,495} Despite the substantial improvements in mass spectrometry resolution and liquid chromatography consistency, accurately identifying extremely

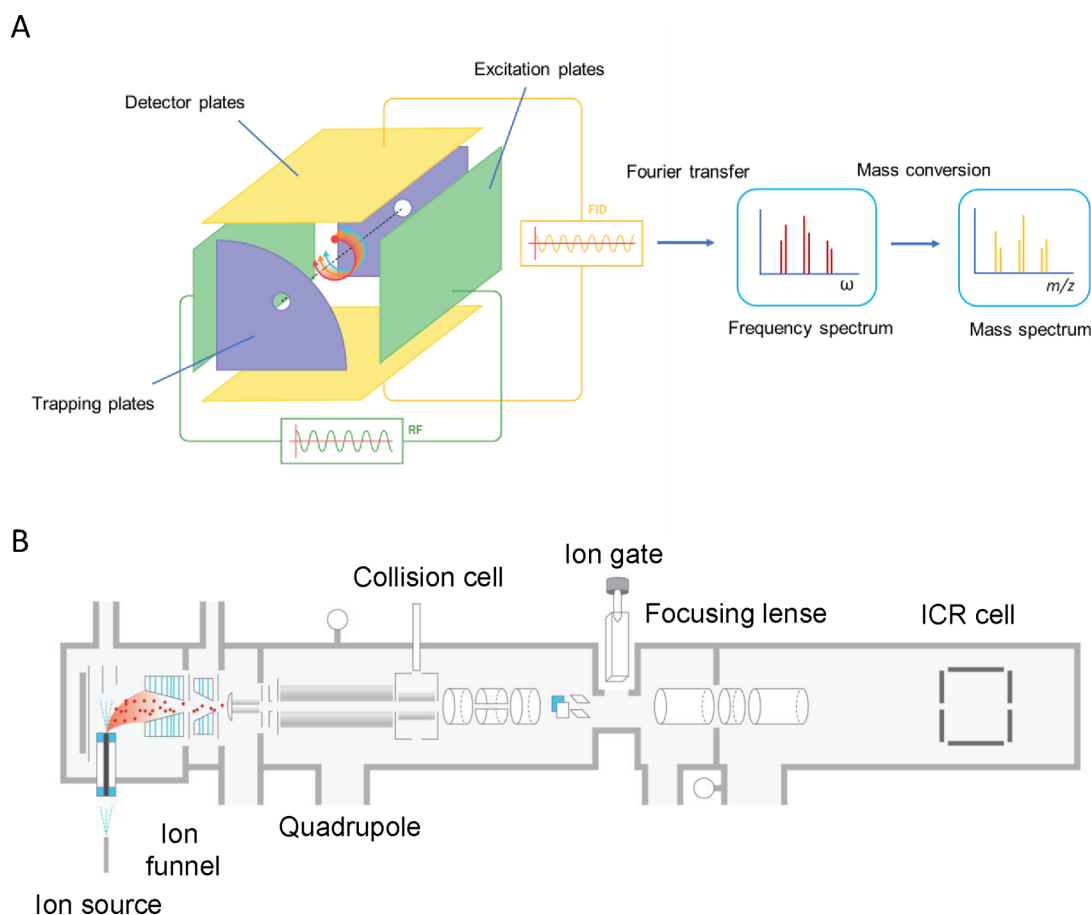


Figure 12. Schematic of FT-ICR. (A) Typical FT-ICR cell. (B) Example of complete FT-ICR system.

similar molecules such as isomers with LC-MS remains a challenge. Ion mobility mass spectrometry (IM-MS), a technique that utilizes electric fields to transport analytes through a buffer gas, is beneficial for separating and identifying ions based on their size, shape, and charge state. This technique provides the fourth dimension of information, collision cross section (CCS), which allows for more comprehensive characterization of molecules.⁴⁹⁶ Apparently, multi-dimensional data is always beneficial for us to understand things comprehensively and accurately, thus getting closer to the truth.

In terms of mass spectrometry based proteomic analysis, adding CCS data can help us better separate, identify, and quantify peptides.

The core principle of ion mobility spectrometry is to separate ions in an inert gas under the influence of an electric field (E), and then measure the amount of time it takes for each ion to pass through drift tube, which is defined to be the steady-state drift velocity (V_d) correlated to the specific analyte's mobility (K), as shown in eq 1.

$$V_d = KE \quad (1)$$

While the primary measurement in IMS analyses is the mobility (K), for many analytical applications, it has become routine to convert K into the calculated collision cross-section value (CCS or Ω) using Mason-Schamp equation (eq 2).⁴⁹⁷

$$\Omega = (3/16[(2\pi/\mu K_b T)]^{(1/2)ze})/N_0 K_0 \quad (2)$$

The components of the equation are defined as follows: e , charge of an electron; z , ion charge; N_0 , buffer gas density; μ , reduced mass of the collision partners; K_b , Boltzmann's constant; and T , the drift region temperature. Although the Mason-Schamp equation isn't universally embraced, it is currently the primary formula the community uses to compute CCS. In basic terms, the CCS serves as a standard metric for the size in the gas phase, generally expressed in units of square Angströms (\AA^2). However, according to the eq 2, parameters including gas composition, working pressure, temperature within the mobility region, path of analyte movement, and the strength of the applied field can influence the final CCS value and may differ for each specific IMS platform. Hence, direct comparison of CCS value between different platforms often requires calibration.

Generally, ion mobility techniques can be categorized into three separation concepts: (1) temporally dispersive, (2) spatially dispersive, and (3) ion confinement (trapping) and selective release (Figure 13A).⁴⁹⁴ Temporally dispersive methods produce an arrival time spectrum based on differences in the time it takes for ions to traverse a similar gas-filled drift region under the influence of an electric field. Time-dispersive technique inherently provides an extensive examination of all signals detected during a given observation window. However, a fundamental limitation of this wide-ranging analysis is the diminished sensitivity linked to a single time dispersion occurrence, which usually requires many (10–100) events to be aggregated to achieve statistically significant ion mobility measurements. In contrast, spatially dispersive methods

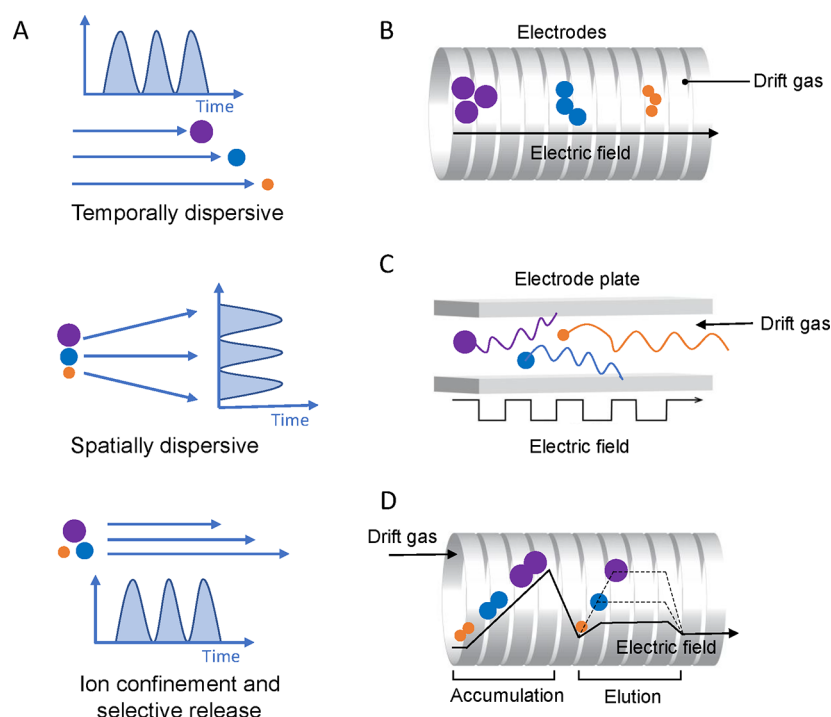


Figure 13. Ion mobility. (A) Conceptual diagram of three types of ion mobility strategies. (B) Schematic of drift tube ion mobility spectrometry. (C) Schematic of high field asymmetric waveform ion mobility spectrometry (FAIMS). (D) Schematic of trapped ion mobility spectrometry (TIMS).

separate ions based on mobility differences (charge, shape and size), leading them on distinct drift paths or trajectories, but without significant time differences. A characteristic of spatially dispersive techniques is the scanning of voltage to obtain a broad-band ion mobility spectrum. Types of spatially dispersive ion mobility include High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS), uniform-field differential mobility analyzers (DMA), and the newly introduced scanned frequency ion mobility filter called transverse modulation ion mobility spectrometry (TMIMS). Ion confinement and release strategies are recently developed techniques which trap ions in a pressurized drift cell by electric field, and then release them based on mobility distinctions. This technique relies on the ability to control the position of ions under elevated pressure conditions using precisely adjustable electrodynamic fields. It requires a precise fabrication craft and more complicated control system. While it has only been perfected recently, typical products like trapped ion mobility spectrometry (TIMS)^{498,499} and traveling wave IMS have become commercially available.⁵⁰⁰ Table 6 summarizes typical ion mobility separation techniques, their separation concept, electric field direction, gas flow direction, strengths, and drawbacks. Also, for three categories of ion mobility techniques, we have selected a typical technique from each for brief introduction.

Drift Tube Ion Mobility Spectrometry (DTIMS). The principle of Drift Tube Ion Mobility Spectrometry (DTIMS) is based on the differential migration (time) of ions through a neutral buffer gas (commonly helium or nitrogen) under the influence of a weak uniform electric field (typically tens of V/cm). The mobility (K) of an ion is proportional to its drift velocity (V) and inversely proportional to the strength of the applied electric field (E). For ions with same charge states, the drift velocities are primarily determined by their collisional

interactions with a buffer gas, namely, mainly affected by their shape and size. To illustrate this process, imagine two objects with identical mass: a solid metal ball and a feather. Due to its lower density, the feather should have a larger volume than the ball. When both are dropped from the same height, the solid ball reaches the ground before the feather because of air resistance. This observation doesn't contradict Newton's law of universal gravitation, as we have accounted for air resistance. In the context of DTIMS, the buffer gas in the drift tube acts as the "air resistance", while the uniform electric field represents the "gravity". Hence, ions with the same mass-to-charge ratio are separated based on their shape and size. This capability allows DTIMS to distinguish between isomeric compounds with identical masses but different structural configurations, given that these isomers might have distinct interactions with the drift gas. Also, following the intuition of the free fall example, in DTIMS, smaller ions will move faster and hit the detector earlier than larger ions in DTIMS (Figure 13B). DTIMS possesses strengths including high resolving power and allows for straightforward measurement of an ion's CCS from first principles.^{501,502} However, DTIMS also suffers from disadvantages including: (1) separation time is too long for all ions passing through the drift tube, relative to the accumulation time, which decreases the duty cycle. (2) A longer drift tube or higher pressure is needed for greater resolving power. However, this inevitably increases ion diffusion and ion losses unless ion focusing techniques are employed. (3) Segmentation and collision between ions and gas molecules during the traveling process in drift tube reduces the sensitivity. Continued advancements in DTIMS design and the application of ion focusing has pushed the resolution of these platforms to between 100–250 ($t/\Delta t$) or even higher.

High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS). High Field Asymmetric Waveform Ion

Table 6. Typical Ion Mobility Separation Techniques^a

Separation concept	Ion mobility techniques	Ion movement direction	Electric field direction	Drift Gas direction	Characteristics
Temporally Dispersive	drift tube IMS (DTIMS)	↑	→	#	High mobility resolution (need long drift tube), direct measurement of CCS. Low speed, large size, low sensitivity.
Temporally Dispersive	traveling wave IMS (TWIMS)	↑	→→→	#	High mobility resolution, faster than DTIMS. Low sensitivity, large size, low sensitivity, traveling electric field waves.
Spatially Dispersive	high-field asymmetric IMS (FAIMS)	↑	↑↓	→	Good as mass filter, Fast. No CCS measurement (Compensation voltage instead), low mobility resolution.
Spatially Dispersive	transverse modulation IMS (TMIMS)	↑	→ and ↑↓	#	Transverse Modulation, compact instrumentation, orthogonal Separation, fast and high resolution.
Confinement and Selective Release	trapped ion mobility spectrometry (TIMS)	↑	←	→	High mobility resolution, compact instrumentation, high sensitivity, high speed, high ion utilization rate.
Confinement and Selective Release	multi-pass cyclic traveling wave IMS	↑	→→→	#	High mobility resolution, improved Signal-to-Noise ratio and sensitivity, versatility (from small molecules to large biomolecules) and adjustability (number of passes can often be adjusted).

^a# means stationary drift gas; →, ←, ↑↓ indicate drift gas direction or electric force direction; →→→ represents a wave and gradient electric field.

Mobility Spectrometry (FAIMS) represents a distinct version of spatially dispersive ion mobility spectrometry. This technique differentiates ions utilizing a pronounced asymmetric oscillating electric field combined with a moving gas. The principle of FAIMS is based on the different trajectories of ions as they move through a high asymmetric electric field, which are determined by their physical structure and charge states.^{503–505} In FAIMS, gas-phase ions are carried by a flow of carrier gas between two electrodes in a direction orthogonal to the direction of asymmetric electric field (E). The asymmetric waveform electric field is typically characterized by a short, high-voltage pulse of one polarity followed by a longer, lower-voltage pulse of the opposite polarity. An ion's mobility within such an electric field is determined by its charge state, its physical structure, and the properties of the surrounding gas it moves through. Once the ions are subjected to an asymmetric electric field, the ions will alternate between travelling toward one electrode or the other as the field oscillates in polarity, resulting in a curved trajectories between the electrodes. Some ions move more in the high field relative to the low field, and vice versa (Figure 13C). To differentiate between ions, a so-called “compensation voltage” (CV), which is a DC offset voltage that compensates for the differential ion movement in the high and low fields, is applied.⁵⁰⁶ In this case, only ions with a specific response to the changing electric field and those that match the applied compensation voltage (CV) will have a zero net movement and are able to traverse the drift region to the detector, while others hit the electrode plate and are neutralized. By scanning or modulating the CV, different ion species can be selectively transmitted through the FAIMS device. In contrast to drift tube IMS in which the ion stream is sampled in discrete packets and all ions reach the detector, FAIMS is a continuous filtration technique that allows uninterrupted sampling of the ion stream, but only for a selected subset of the ion population. One of the primary advantages of this continuous collection technique is to greatly increase the signal-to-noise ratio for the ion(s) of interest by removing unwanted chemical noise, which make FAIMS more similar to a m/z filter than other ion mobility spectrometry tools. FAIMS also has the advantage of operating at atmospheric pressure. Drawbacks of FAIMS, however, are that it does not produce any CCS values and it has relatively low resolution separations. Commercial FAIMS products from vendors including Thermo Fisher and Waters are available now.

Trapped Ion Mobility Spectrometry (TIMS). Trapped ion mobility spectrometry (TIMS) is a common type of ion mobility which uses ion refinement and release strategy.⁵⁰⁷ The basic idea behind TIMS is a combination of traditional ion mobility spectrometry and ion trapping techniques. Instead of driving ions through a drift tube filled with stationary gas, TIMS holds the ions stationary in a drift cell under a moving buffer gas and then releases them by adjusting electric fields (voltages on electrodes). This process was realized by applying two different electric fields as follows.

1. Radially confining pseudopotential. An RF (radio frequency) voltage is applied to the electrodes of the TIMS analyzer to generate a radially confining pseudopotential, with essentially no axial component; this is only used for focusing ions to the central region of the TIMS tube, preventing them from diffusion into electrodes.

2. Axial electric field. An axial electric field gradient, produced by superimposing DC potentials on tunnel electrodes, is applied for “trapping” ions based on the equilibrium between the force of drift gas and the opposing force from the electric field gradient, which is stronger at the entrance and becomes progressively weaker moving deeper into the tunnel.

Once ions enter the device, lower mobility ions are trapped at positions where the magnitude of the axial electric field is larger, while higher mobility ions are confined to deeper positions of the tunnel where the axial field is lower. Then, after enough ions have been accumulated in the TIMS tunnel, additional ions are prevented from entering the tunnel region and residing ions are trapped for a short time (usually few milliseconds) which can be defined by users. Finally, the magnitude of axially electric field is decreased at a user defined rate so that ions are eluted as an order of mobilities value (K) from high to low (Figure 13D). The axially electric field gradient is set by a resistor divider. Importantly, like other ion mobility strategies, the resolving power of TIMS is highly dependent on the length of the gas column through which the ions traverse. In TIMS, ions are trapped in a specific location while buffer gas continuously flows past them. Thus, the resolving power achieved by TIMS depends on the “quantity” of gas, specifically the length of the gas column, that passes by the ions during the separation time. This offers the direct benefit of allowing the analyzer to maintain a compact physical size (around 5 cm) and achieve a high resolving power ($R \sim 300$), while the analytical gas column—the portion that flows during an analysis—can be extensive (up to 10 m) and tailored to the user’s needs. Moreover, by leveraging the “trapping” capability (trapping time) of TIMS and the high scanning speeds of TOF, platforms such as TIMS-Q-TOF can implement a full duty cycle acquisition protocol known as Parallel Accumulation-Serial Fragmentation (PASEF).^{498,508} This is particularly meaningful for identifying more peptides within a given time frame, such as capturing more precursors from co-eluted peptides in the same liquid chromatography peak. Currently, Bruker is the primary provider of commercial mass spectrometers that utilize TIMS-tof technology. (TIMS-tof pro, TIMS-tof pro2, SCP. etc.).

Structures for Lossless Ion Manipulation (SLIM). A final type of ion mobility spectrometry discussed here is Structures for Lossless Ion Manipulation (SLIM), invented by Richard Smith and colleagues at Pacific Northwest National Labs.⁵⁰⁹ SLIM uses printed circuit boards to confine ions in long path lengths for high resolution ion mobility. Ions can be passed through the board multiple times to achieve path lengths of several meters to over 1 km for high-resolution IMS separation.⁵¹⁰ This technology is currently under commercial development by Mobilion, in a platform named “Mobie”.⁵¹¹

■ TANDEM MASS SPECTROMETRY AND PEPTIDE FRAGMENTATION

Tandem Mass Spectrometry

Tandem MS, where precursor ions are selected and fragmented to generate an MS/MS spectrum containing peptide-derived product ions, is a fundamental process in modern proteomics.^{512,513} This is largely because intact peptide mass alone cannot unambiguously provide a peptide’s sequence;⁵¹⁴ however, MS/MS spectra provide more information due to predictable fragmentation behavior of peptide ions

to generate sequence-informative fragments.^{512,515} Some more advanced proteomic acquisition methods use MS1-only feature detection in combination with retention time to maximize information used for downstream quantitation.⁵¹⁶ In most of these, identifications are fundamentally based on MS/MS spectra, either acquired as part of a specific LC-MS/MS analysis that contains the MS/MS spectra themselves or on a spectral library of MS/MS spectra acquired previously.^{517,518} True MS1-only methods that use only accurate mass and retention time for identification have been discussed, but these have yet to be widely adopted.⁵¹⁶

The value of MS/MS spectra for peptide identification comes from predictable fragmentation behavior of peptide ions to generate sequence-informative fragments.^{512,515} Multiple dissociation methods exist to generate product ions in MS/MS spectra through various mechanisms (Figure 14). In non-modified peptides, the most labile bonds are typically peptide bonds (i.e., amide bonds) between amino acids. Depending on where peptides dissociate along the peptide backbone, the fragments are assigned different ion types (Figure 14A). Fragment ion nomenclature was first developed by Roepstorff and Fohlman in 1984⁵¹⁹ and then refined by Biemann in 1990.⁵²⁰ The main ion types are the fragments that contain the original peptide N-terminus (i.e., a-, b-, and c-type ions), or the original peptide C-terminus (i.e., x-, y-, and z-type ions). The number associated with each fragment ion indicates how many amino acids from each terminus are included.

One of the earliest and most ubiquitous peptide fragmentation methods is collision-induced dissociation (CID, also called collisionally-activated dissociation, CAD)⁴⁵⁹ (Figure 14B). Here, collisions with inert gas molecules are used to increase the internal energy of peptide ions to reach bond dissociation energies that fragment them into products. Various inert gases can be used; helium, nitrogen, and argon are the most common. Preferences for which gas is used is often a function of how much energy per collision is desired. Two main versions of CID are used in proteomics, with the most common being beam-type CID (beamCID, sometimes called higher-energy collisional dissociation, HCD).^{521,522} BeamCID typically uses nitrogen or argon as a collision gas, and peptide ions are accelerated into a collision cell filled with several mTorr of bath gas. The kinetic energy used to accelerate precursor ions (often generated using direct current voltage differentials between the source of the ions and the collision cell) determines the energy imparted through collisions with the bath gas, which in turn governs their fragmentation behavior.

Since in non-modified peptides the most labile bonds are typically peptide bonds (i.e., amide bonds) between amino acids, the increase in internal energy from beamCID generates b- and y-type ions that represent this peptide bond cleavage, as shown in Biemann fragment ion nomenclature (Figure 14A). b-type ions provide sequence information for fragments that have an intact N-terminus, while y-type ions denote fragment ions with an intact C-terminus. Collisions in beamCID cause near instantaneous generation of primary fragment ions. Because the increase in internal energy happens rapidly before energy can be redistributed, beamCID can generate fragments that are not necessarily derived from cleavage of the most labile bonds (e.g., PTM-modified peptides, discussed below), but spectra are often dominated by b/y-type ions from amide bond cleavage (Figure 14B). BeamCID can also generate secondary fragments, such as immonium ions from side chain losses⁵²³ or

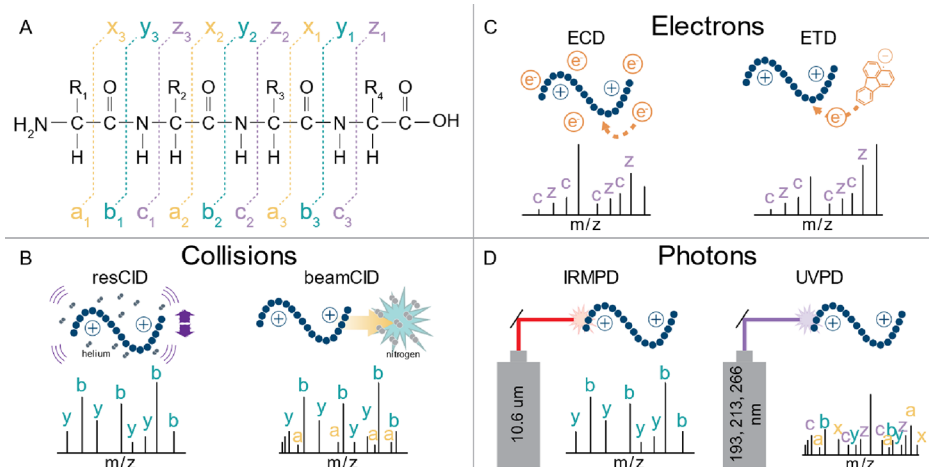


Figure 14. Peptide fragmentation methods. (A) Sequence-informative fragment ions are termed a/x-, b/y-, and c/z-type fragments depending on which bond along the peptide backbone breaks. Fragments that explain the intact N-terminus of the peptide are a-, b-, and c-type ions, while x-, y-, and z-type ions explain the intact C-terminus of the peptide. Other panels show common dissociation methods, including collision, electron, and photon-based fragmentation. (B) Resonant collision-induced dissociation (resCID) and beam-type CID (beamCID) both produce mainly b/y-type sequencing ions through collisions with background gases like helium and nitrogen that increase the internal energy of peptide cations. (C) Electron capture and electron transfer dissociation (ECD and ETD) generate mainly c/z-type fragments through electron-mediated radical driven cleavage of the peptide backbone. (D) Infrared multi-photon dissociation (IRMPD) is a slow heating method similar in dissociation mechanism to resCID, but very different in implementation due to the IR lasers required (often with lower energy 10.6 micron photons). Ultraviolet photodissociation (UVPD) can use a range of wavelengths (popular options shown) to introduce higher energy photons to peptide cations, causing vibrational and electronic excitation that can generate all major fragment ion types depending on wavelength used.

a-type fragment ions that come from water loss from b-type ions due to multiple collision events (note: a-type ions can form as primary fragmentation products in other dissociation methods). The simplicity of beamCID, which requires an rf-only collision cell, has made it widely implemented on most instrument platforms used in modern proteomics.

A second form of CID is called resonant CID (resCID), where the internal energy of peptide ions is slowly increased through multiple low-energy collisions. Here, helium gas is most often used, as it imparts less energy per collision, and activation typically happens in ion trap devices where supplemental frequencies can be used to excite ions. In other words, ions are trapped using axial rf-frequencies, and an additional rf-frequency is applied to the electrodes of the ion trap.⁵²⁴ This supplemental rf is selected to have a frequency resonant with the fundamental frequency of the ions to be fragmented, as determined by the Mathieu equations, which excites the ions of interest so that they have increased kinetic energy as they move in the ion trap.^{525,526} The increased kinetic energy creates more collisions with the background helium gas to slowly build up the internal energy of the precursor ions until the dissociation energy of the most labile bond is reached, causing fragmentation. Once ions dissociate, the fragments have different *m/z* values than the precursor ions, meaning they fall out of resonance with the supplemental rf and are no longer activated. Thus, resCID typically fragments only the most labile bonds in precursor ions and does not have secondary fragmentation behavior. As above, for non-modified peptide ions, this typically generates sequence-informative b- and y-type product ions. For modified peptides where the bonds connecting the modification to an amino acid are more labile than peptide bonds (e.g., phosphopeptides and glycopeptides), resCID MS/MS spectra can be dominated by product ions only of the PTM-loss rather than sequence-informative fragment ions, although many factors govern this behavior.^{527,528} Because of this, and because this method

requires an ion trap device with the ability to apply supplemental rfs, resCID is less prevalent than beamCID. For both beamCID and resCID, the mobile proton model has been widely accepted to explain fragmentation behavior,⁵²⁹ and this largely predictable behavior has greatly helped in manual and algorithm-assisted spectral interpretation.

Despite the utility and broad adoption of CID, alternative dissociation methods have been explored for a variety of uses, including applications where CID is inadequate for the experimental question.^{530–532} The most popular of these alternative dissociation methods are electron-based dissociation (ExD) approaches, which include electron capture dissociation (ECD) and electron transfer dissociation (ETD). In both of these, peptide cations capture thermal electrons (ECD⁵³³) or abstract an electron from a reagent anion (ETD⁵³⁴) to generate radical-driven dissociation of the N-Cα bond that predominantly generates sequence-informative c- and z-type product ions (Figure 14C). The mechanisms of ExD methods have been widely explored,^{535,536} and the preferential cleavage of N-Cα bonds along the peptide backbone have been particularly useful for PTM-modified species because the modifications remain largely intact even during peptide backbone bond fragmentation. ExD methods have shown promise for analysis of numerous PTMs, including phosphorylation, glycosylation, ADP-ribosylation, and more.^{537,538} Electron-based dissociation is also more suitable than collision-based dissociation for MS analyses of intact proteins^{539,540} and larger oligonucleotides.^{541–546}

Two fundamental challenges exist with ExD methods. First, ExD implementation requires instruments that can manipulate cations and anions (or free electrons) within the same scan sequence and can trap both simultaneously for electron capture/transfer events to occur. This has been successfully accomplished on a number of instruments, including FT-ICR systems, ion traps, ToFs with quadrupole ion traps, and hybrid Orbitrap instruments, but it is not a ubiquitous feature of all

platforms. That said, several exciting advances in recent years have made ExD methods more accessible on numerous instrument configurations.^{537,538,547–549} A second challenge is the dependence of ExD dissociation efficiency on precursor ion charge density.⁵⁵⁰ ExD methods generally produce robust fragmentation for charge dense precursor ions (i.e., those with relatively low m/z values and higher z). Alternatively, precursors with low charge density (i.e., higher m/z values) have relatively condensed secondary gas-phase structure that leads to non-covalent interactions. Even in the cases when ExD methods drive peptide backbone cleavage, product ions (i.e., c- and z-type fragments) are held together by the non-covalent interactions so that few (or no) sequence-informative product ions are produced. This process is called non-dissociative electron-capture/transfer (ECnoD/ETnoD).⁵⁵¹ Several strategies to mitigate ECnoD/ETnoD have been successfully explored, including supplemental activation of product ions with resCID (ETcaD⁵⁵²) or beamCID (EThcD^{553,554}), supplemental activation with infrared photons (AI-ECd^{555,556} and AI-ETD^{557–560}) or ultraviolet photons (ETuvPD⁵⁶¹), and use of higher energy electrons.^{549,562,563} Despite their successes, these methods still require instrumentation capable of ExD in addition to extra hardware needed for a given strategy (e.g., a CO₂ laser in AI-ETD⁵⁶⁴). As with ExD in general, recent advances in supplemental activation strategies for ExD are making these tools more accessible.^{537,538}

Photoactivation is another family of alternative dissociation strategies that has been steadily gaining popularity.^{565,566} Infrared multi-photon dissociation (IRMPD) is canonically the photodissociation method used in early proteomic applications,⁵⁶⁶ but ultraviolet photodissociation (UVPD) has been the more widely used approach in the recent decade.⁵⁶⁷ IRMPD functions similarly to resCID; it is a slow heating approach that causes vibrational excitation due to absorption of low energy photons, generally 10.6 μm photons from a CO₂ laser.^{568,569} Predominant fragments are b- and y-type fragments, although secondary fragmentation occurs because fragment ions remain in the photon path after the initial dissociation event (Figure 14D). Despite limited use in the past decade, recent work shows that IRMPD, or more generally activation with IR photons, may still have value in the proteomics toolkit.^{279,559} UVPD has been explored with a number of wavelengths, including 157 nm, 193 nm, 213 nm, 266 nm, and 355 nm.^{570–575} Higher-energy UVPD approaches, like 193 and 213 nm photons, are typically used for underivatized peptide and protein ions,⁵⁶⁷ while others, like 266 and 355 nm, can be used for directed fragmentation at specific residues with natural chromophores (e.g., tyrosine) or exogenously added chromophore tags.^{576,577} UVPD with 193 and 213 nm generate multiple fragment types, including sequence-informative a-, b-, c-, x-, y-, and z-ions in addition to other fragmentation pathways, which occur through vibrational and electronic excitation.⁵⁷⁸ UVPD has been explored for bottom-up proteomic applications, but its more impactful utility, arguably, has been realized for intact protein characterization.⁵⁷⁹ The laser needed for UVPD (i.e., the photon wavelength desired) determines much about its implementation. 193 nm photons are typically generated using an Excimer laser with ArF gas,⁵⁸⁰ while 213 nm photons can be generated with a solid-state laser that is easier to integrate into an instrument platform and maintain.^{572,581} That said, 213 nm photons tend to provide more directed, preferential cleavage

pathways compared to 193 nm photons that cleave more broadly in non-directed fashion.⁵⁸² Outside of ExD and photoactivation approaches, other alternative dissociation methods have been explored for various proteomic applications, although they are not as widely adopted at ExD and UVPD methods.⁵⁶⁵

DATA ACQUISITION

Hybrid mass spectrometers used for modern proteome analysis offer the flexibility to collect data in many different ways. Data acquisition strategies differ in the sequence of precursor scans and fragment ion scans, and in how analytes are chosen for MS/MS. Constant innovation to develop better data collection methods improves our view of the proteome, but many method options may confuse newcomers. This section provides an overview of the general classes of data collection methods.

Data acquisition strategies for proteomics fall into one of two groups.

1. Data dependent acquisition (DDA), in which the exact scan sequence in each analysis depends on the data that the mass spectrometer observes.
2. Data independent acquisition (DIA), in which the exact scan sequence in each analysis DOES NOT depend on the data; the collected scans are the same whether you inject yeast peptides, human peptides, or a solvent blank.

DDA and DIA can both be further subdivided into targeted and untargeted methods.

DDA

In most cases, the peptide masses that will be observed are not known before doing the experiment. Data collection methods must account for this. DDA was invented in the early 1990s, which enabled collecting MS/MS spectra for observed peptides as they eluting from the LC column.^{583–585}

Untargeted DDA. A common method currently used in modern proteomics is untargeted DDA. The MS collects precursor (MS1) scans iteratively until precursor mass envelopes meeting certain criteria are detected. Criteria for selection are usually specific charge states and a minimum signal intensity. When those ions meet these criteria, the MS selects those masses for fragmentation.

Because ions are selected as they are observed, repeated DDA of the same sample will produce a different set of identifications. This stochasticity is the main drawback of DDA.

Because DDA is required for quantification of proteins using isobaric tags like TMT, this stochasticity of DDA limits the ability to compare quantities across batches. For example, if you have 30 samples, you can use two sets of the 16-plex kit to label 15 samples in each set with one channel labeled by a pooled sample to enable comparison across the groups. When you collect DDA data from each of those sets, each set will have MS/MS data from an overlapping but different set of peptides. If one set has MS/MS from a peptide but the other set does not, then that peptide cannot be quantified in the whole sample group. This limits the number of quantified proteins in large TMT experiments with multiple batches.

Targeted DDA. Targeted DDA is not common in modern proteomics. In targeted DDA, in addition to general criteria like a minimum intensity and a certain charge state, the mass spectrometer looks for specific masses. These masses might be previously observed signals that were previously missed by

MS/MS.^{586,587} In these studies, the sample is first analyzed by LC-MS to detect precursor ion features with some software, and then subsequent analyses target those masses for fragmentation with inclusion lists until they are all fragmented. This was shown to increase proteome coverage.

DIA

The simplest method to operate a mass spectrometer is to have predefined scans that are collected for each sample analysis. This is data-independent acquisition (DIA); the scan sequence does not depend on the data that the instrument observes. Thus, the scan sequence is repetitive, looping through predetermined scans, most often which are m/z quadrupole selection ranges followed by fragmentation in a second quadrupole and fragment ion detection in a final MS stage. In DIA, the same scan sequence is performed if we inject air, a blank, peptides, ammonia, or anything. Like DDA, DIA can also be either targeted or untargeted.⁵⁸⁸ The two targeted DIA methods are SRM/MRM or PRM. Untargeted DIA (uDIA) is often referred to simply as “DIA” or “SWATH” (Sequential Window Acquisition of All Theoretical Mass Spectra) (Figure 15).

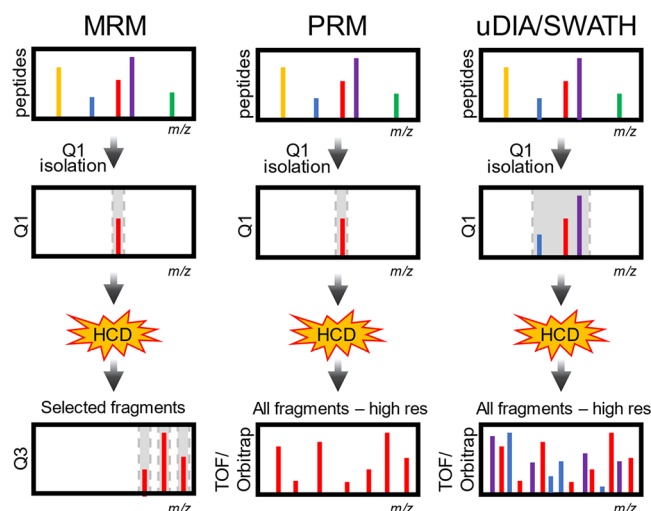


Figure 15. Types of DIA. (A) SRM/MRM. Peptides are ionized by ESI and although there are many peptides entering the mass spectrometer at any time, the first quadrupole (Q1) isolates one mass, which is then fragmented by HCD. Fragment masses from the peptide are then selected in the third quadrupole (Q3). This leads to very low noise and high sensitivity. (B) PRM. Like MRM, peptides are selected in the first quadrupole, but this analysis is done on a high-resolution instrument like an Orbitrap or TOF. Selectivity is gained by exploiting the high mass accuracy and resolution to monitor multiple fragment ions. (C) uDIA/SWATH. Like MRM and PRM, peptides are isolated with Q1, but in this case a much wider isolation window is used. This usually results in co-isolation of many peptides simultaneously. Fragments from many peptides are measured with high resolution and high mass accuracy. Special software is used to get peptide identities and quantities from the fragment ions.

Targeted DIA. The first type of targeted DIA is called SRM or MRM.⁵⁸⁹ The popularity of this method in the literature peaked in 2014, with just under 1,500 documents on PubMed that year resulting from a search for “MRM”. In this strategy, the QQQ MS is set so that the first quadrupole selects the precursor mass of the peptide(s) of interest, the second quadrupole fragments the peptide, and the third quadrupole monitors the product of specific fragments from that peptide.

This strategy is very sensitive and has the benefit of very low noise. The fragments monitored in Q3 are chosen such that it is unlikely these fragments could arise from another peptide. Usually at least a few transitions are monitored for each peptide in order to get multiple measures for that peptide.

An early example of MRM applied to quantify c-reactive protein was in 2004.⁵⁹⁰ Around the same time, SRM was combined with antibody enrichment of peptides from target proteins.⁵⁹¹ This approach was popular for analysis of plasma proteins.⁴⁶⁴ These early examples led to many more studies that used QQQ MS instruments to get accurate quantitation of many proteins in one injection.^{592,593} Scheduling MRM measurement when chromatography is stable additionally enabled better utilization of instrument duty cycle and therefore monitoring of more peptides per injection.⁵⁹⁴ Efforts even developed libraries of transitions that allow quantification of any protein in model organisms.⁵⁹⁵

Another similar targeted DIA method is called parallel reaction monitoring (PRM).⁵⁹⁶ Instead of using a QQQ instrument, PRM uses a hybrid MS with a quadrupole and a high-resolution mass analyzer, such as an Q-TOF or Q-Exactive. The idea is that instead of monitoring specific fragments in Q3, the high mass accuracy can be used to filter peptide fragments for high selectivity and accurate quantification. Studies have found that PRM and MRM/SRM have comparable dynamic range and linearity.⁵⁹⁷

Untargeted DIA. There were many implementations of uDIA over the years, starting in 2003 by Purvine et al from the Goodlett lab.⁵⁹⁸ In this first work they demonstrated uDIA using a Q-TOF with in source fragmentation and showed that extracted ion chromatograms of precursor and fragment ions matched in shape suggesting that this could be used to identify and quantify peptides. The following year, Venable et al from the Yates lab introduced uDIA with an ion trap.⁵⁹⁹ Subsequent methods include MSE,⁶⁰⁰ PACIFIC,⁶⁰¹ all ions fragmentation (AIF).⁶⁰² Computational methods were also developed to automate interpretation of this data, such as DeMux,⁶⁰² XDIA,⁶⁰³ and ETISEQ.⁶⁰⁴

The paper that is often cited for uDIA that led to widespread adoption was by Gillet et al. from the Aebersold group in 2012.⁶⁰⁵ In this paper they branded the idea as SWATH. Widespread adoption may have been facilitated by the co-marketing of this idea by ABSciex as a proteomics solution on their new 5600 Q-TOF (called “tripleTOF” despite containing only one TOF, likely a portmanteau of “triple quadrupole” and “Q-TOF”). Importantly, in the Gillet et al. paper the authors described a computational method to extract information from SWATH where peptides of interest were queried against the data. They also demonstrated the application of SWATH to measure proteomic changes that happen in diauxic shift, and showed that SWATH can reveal modified peptides, in this case a methionine oxidation.

There are also many papers describing uDIA with orbitraps. One early example described combining random isolation windows together and then demultiplexing the chimeric spectra.⁶⁰⁶ In another landmark paper, over 6000 proteins were identified from mouse tissue by at least 2 peptides.⁶⁰⁷ In 2018, the new model orbitrap at that time (HF-X) enabled identification of nearly 6000 human proteins in only 30 minutes. Currently orbitraps have all but replaced the Sciex Q-TOFs for DIA data collection.

A new direction in uDIA is the addition of ion separation by ion mobility. This has appeared in two forms. On the

timsTOF, diaPASEF makes use of the trapped ion mobility to increase speed and sensitivity of analysis.⁶⁰⁸ On the orbitrap, the combination of FAIMS and DIA has enabled the identification of over 10,000 proteins from one sample, which is a major milestone.⁴³⁴

Acquisition Methods for PTMs. Phosphopeptides. Resonant CID⁶⁰⁹ and beam-type HCD⁶¹⁰ are the most popular methods for unmodified and modified peptides due to their speed, accessibility, and efficiency. Due to the weak phosphoester bond relative to the peptide backbone, resonant CID usually produces spectra that are dominated by only the neutral loss of the phosphate. For this reason, the optimal dissociation methods for phosphopeptide identification and phosphosite localization include HCD or ExD-based methods, discussed in the previous chapter in more depth.^{611,612} ExD methods generate phosphopeptide MS/MS spectra with many c- and z⁺-type fragment ions for peptide sequencing and localization of labile phosphate modifications, typically disrupted with CID.⁵³⁴ Gas-phase phosphate rearrangement induced by collisional activation represents a glaring challenge for the field and several have explored site localization in the face of rearrangement.^{613–615}

Advanced data acquisition schemes trigger predetermined MS/MS events when a specific fragment ion or neutral loss is detected in a spectrum. Certain decision-tree strategies have arisen to increase data acquisition efficiency, including pseudo-MS3 scans which are triggered on detection of phosphate losses⁵²⁸ and the use of site-specific x-type ions.⁶¹⁶ For example, when linear ion traps were the main proteomics workhorses, resonant CID analysis of phosphopeptides would result in predominantly neutral loss of the phosphate with limited sequence ion information. To gain sequence ions in these experiments, instruments could be set to isolate a loss of 98 Daltons for MS3 activation.^{617,618} The newer collisional dissociation technique HCD, or beam-type collisional activation, significantly improves the detection of peptide fragments with the phosphorylation intact on fragment ions, and thus, this neutral loss scanning technique is no longer common.

Recently developed approaches to phosphopeptide identification include DIA-based phosphoproteomics with Spectronaut,^{619,620} “plug-and-play” high-resolution MS,⁶²¹ SureQuant for phosphotyrosine,³⁰⁹ PIQED for direct identification and quantification of phosphorylation from DIA without a prior spectral library,⁴⁸⁵ and FAIMS front-end separations which yield 15–20% more phosphosite identifications than non-FAIMS experiments.⁵⁰⁵ For quantification of phosphoproteins, Högberg et al. investigated several of the most common strategies and concluded that TMT-based MS2 strategies may be the current best approach.⁶²²

Glycopeptides. A similar product-dependent MS/MS triggering strategy was introduced for N-linked glycopeptides.⁶²³ Collisional dissociation of glycosylated peptides produces oxonium ions, for example at m/z 204.09 (HexNAc) or m/z 366.14 (HexHexNAc). If oxonium ions from the fragmented glycan are detected among the most abundant fragment ions of the HCD spectra, then an ETD scan is triggered. This ETD scan provides information about the peptide sequence, while the original HCD scan provides glycan structure information.

RAW DATA ANALYSIS

The goal of raw data analysis is to convert raw spectral data into lists of altered protein groups, which requires many steps, including checking data quality, peptide spectra matching, protein inference,^{624,625} quantification, and statistical hypothesis tests. Subsequently, many additional analyses can be performed to make biological inferences, which is covered in a subsequent section. An overview of the entire data analysis cycle is shown in (Figure 16).

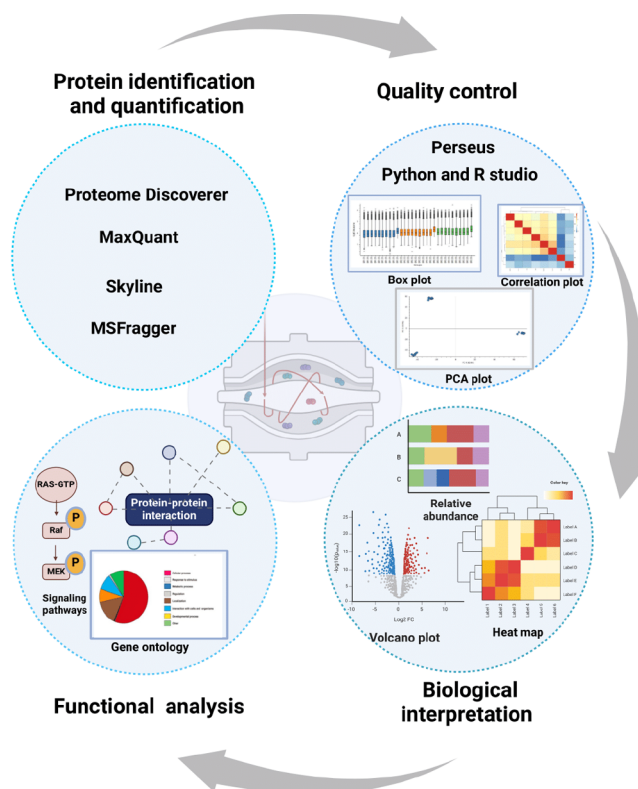


Figure 16. Proteomics data analysis and biological interpretation. The process begins with protein identification and quantification using tools such as Proteome Discoverer, Spectronaut, Spectromine, MS Fragger, MaxQuant, and Skyline. Quality control measures ensure data integrity, leading to a biological interpretation of the results. Differential expression analyses may include relative abundance charts, heat maps, and volcano plots. Functional analysis encompasses gene ontology, protein-protein interactions, and signaling pathways.

Due to the inherent differences in the data structures of DDA and DIA measurements, there exist different types of software that can facilitate the steps mentioned above. The existing software for DDA and DIA analysis can be further divided into freeware and non-freeware.

Analysis of DDA Data

DDA data analysis either directly uses the vendor proprietary data format directly with a proprietary search engine like Mascot, SEQUEST (through Proteome Discoverer), Paragon (through Protein Pilot), or it can be processed through one of the many freely available search engines or pipelines, for example, Comet, MaxQuant, MSGF+, X!Tandem, Morpheus, MS-Fragger, and OMSSA. Table 7 gives weblinks and citations for these software tools. For analysis with freeware, raw data is converted to standard open XML formats like mzML.^{626–628} The appropriate FASTA file containing proteins predicted

Table 7. DDA Data Analysis Software Examples

Name	Publication	Website	Free/ Paid
MaxQuant	Cox and Mann, 2008 ⁶³⁰	MaxQuant	
MSFragger	Kong et al., 2017 ⁶³¹	MSFragger	free
Mascot	Perkins et al., 1999 ⁶³²	Mascot	free
MS-GF+	Kim et al. ⁶³³	MS-GF+	free
X!Tandem	Craig et al. ^{634,635}	GPMDB	free
Comet	Eng et al., 2012 ⁶³⁶	Comet	free
Skyline	MacLean et al., 2010 ⁶³⁷	Skyline	free
ProteomeDiscoverer		ProteomeDiscoverer	paid
Mascot	Perkins et al., 1999 ⁶³²	Mascot	paid
Spectromine		Spectromine	paid
PEAKS	Tran et al., 2018 ⁶³⁸	PEAKS	paid

from that organism's genome is chosen as a reference database to search the experimental spectra. All search parameters like peptide and fragment mass errors (i.e., MS1 and MS2 tolerances), enzyme specificity, number of missed cleavages, chemical artifacts and potential biological modifications (variable/dynamic modifications) are specified before executing the search. The search algorithm scores each query spectrum against its possible peptide matches.⁶²⁹ A spectrum and its best scoring candidate peptide are called a peptide spectrum match (PSM). The scores reflect a *goodness-of-fit* between an experimental spectrum and a theoretical one and do not necessarily depict the correctness of the peptide assignment.

Recall that we noted the stochasticity of DDA; every injection will select different peptide precursors for fragmentation leading to different identifications from each sample. To ameliorate this issue, often strategies are used to transfer identifications between multiple sample analyses. This transfer of IDs across runs is known as "match between runs", which was originally made famous by the processing software MaxQuant.^{639,640} There are several other similar tools and strategies, including the accurate mass and time approach,⁶⁴¹ Q-MEND,⁶⁴² IDEAL-Q,⁶⁴³ and superHIRN.⁶⁴⁴ More recent work has introduced statistical assessment of MBR methods using a two-proteome model.⁶⁴⁵ Statistically controlled MBR is currently available in the IonQuant tool.⁶⁴⁶

Strategies for Analysis of DIA Data

DIA data analysis is fundamentally different from DDA data analysis because, instead of a single MS/MS spectrum for each peptide, we can observe the elution of peptide fragments for any peptide over chromatography time. Even though DIA data analysis derives peptide matches differently, the same target-decoy analysis described above is often used. There are two general approaches for peptide identification from DIA data: peptide-centric and spectrum-centric.

Peptide-centric approaches look for evidence of specific peptides that are in some assay library of MS/MS spectra. That library could be predicted spectra (e.g., using Prosit),⁶⁴⁷ or previously measured spectra (e.g., from a organism-wide knowledge base).⁶⁴⁸ Examples of software that perform peptide-centric analysis include OpenSWATH,⁶⁴⁹ Spectronaut,⁶⁰⁷ csoDIAq,⁶⁵⁰ and DIA-NN.⁶⁵¹

Spectrum-centric approaches instead ask if there is evidence for any peptide based on analysis of the observed spectra. Examples of spectrum-centric approaches include DIA-Umpire⁶⁵² and PECAN.⁶⁵³ Spectrum-centric approaches may assemble pseudo-MS/MS spectra from co-elution of fragments that can then be used with any DDA database search.⁶⁵² Spectrum-centric approaches may be less sensitive at peptide identification than peptide-centric approaches.

A non-comprehensive list of software for DIA data analysis is found in Table 8.

Table 8. DIA Data Analysis Software Examples

Name	Publication	Website	Free/ Paid
MaxDIA	Cox and Mann, 2008 ⁶³⁰	MaxQuant	free
Skyline	MacLean et al., 2010 ⁶³⁷	Skyline	free
DIA-NN	Demichev et al., 2019 ⁶⁵¹	DIA-NN	free
EncyclopeDIA	Searle et al., 2018 ⁶⁵⁴	EncyclopeDIA	free
Spectronaut	Bruderer et al., 2015 ⁶⁵⁵	Spectronaut	paid
PEAKS	Tran et al., 2018 ⁶³⁸	PEAKS	paid
Scaffold DIA		Proteome Software	paid

Deriving Statistical Significance of PSMs with the Target-Decoy Approach

For evaluating the probability that a PSM is not random, matches to a decoy database of shuffled or reversed sequences are used as the null model for peptide matching. A randomized or reversed version of target database is used as a non-parametric null model. The decoy database can be searched separate from the target database (Käll's method)⁶⁵⁶ or it can be combined with the target database before search (Elias and Gygi method).⁶⁵⁷ Using either method, an estimate of false hits can be calculated which is used to estimate the false discovery rate (FDR).⁶⁵⁸ The FDR denotes the proportion of false hits in the population accepted as true. For Käll's method: the false hits are estimated to be the number of decoys above a given threshold. It is assumed that the number of decoy hits that pass a threshold are the false hits. A similar number of target population may also be false. Therefore, the FDR is calculated as:⁶⁵⁹

$$\text{FDR} = \frac{\text{DecoyPSMs} + 1}{\text{TargetPSMs}}$$

For Elias and Gygi Method, the target population in which FDR is estimated changes. The target and decoy hits coming from a joint database compete against each other. For any spectrum, either a target or a decoy peptide can be the best hit. It is argued that the joint target-decoy population has decoy hits as confirmed false hits. However, due to the joint database search, the target database may also have equal number of false hits. Thus, the number of false hits is multiplied by two for FDR estimation.

$$\text{FDR} = \frac{2 \times \text{DecoyPSMs}}{\text{Target} + \text{DecoyPSMs}}$$

Integrated MS Data Analysis Platforms

Given the complexity of proteomic data analysis and the requirement for many steps to get from raw data to quantified proteins, there are some integrated software environments that easily allow users to complete everything in one place.

Peptide Shaker. Since each search engine may give slightly different results, and peptides identified by multiple search engines may be more confident hits, integration platforms such as PeptideShaker have been developed to combine search results.^{660,661} PeptideShaker gives an interactive overview of all the protein, peptide, and PSMs in a dataset. It also has many other features, such as PTM summaries, 3D structure mapping of detected peptides, QC, validation, and GO enrichment (see [Biological Interpretation section](#) for more details).

Trans-Proteomic Pipeline (TPP). The Trans-Proteomic Pipeline (TPP) is a free and open-source mass spectrometry data analysis suite for end-to-end analysis that remains in continual development to provide ever-expansive data analysis capabilities since its inception over twenty years ago.^{75,662–671} The current release provides tools for mass spectrometry spectral processing, spectrum searching, search validation, abundance computation, protein inference, and statistical evaluation of the data to ensure controlled false-discovery rates. Many of the tools include machine-learning modeling to extract the most information from datasets and build robust statistical models to compute probabilities that derived information is correct.

One of the major advantages of TPP is its ability to be deployed in a wide variety of environments, from personal Windows laptops to extensive large Linux clusters for automated use within cloud computing environments. While the command-line interfaces are appreciated by many power users, others prefer a graphical user interface (GUI), which is provided by the TPP GUI called Petunia, allowing users to use the TPP from any web browser on any platform. Petunia has the advantages that the same exact GUI is available on a modest Windows laptop, a powerful expandable Linux server shared by a research group, or a remote cloud computing instance running on Amazon Web Services (AWS).⁶⁶⁶

The TPP includes many statistical validation tools such as PeptideProphet,⁶⁷² ProteinProphet,⁶⁷³ iProphet,⁶⁶⁵ and PTMProphet,⁶⁷⁰ where Bayesian machine learning techniques are applied to the various search engine scores to model the correct and incorrect assignment distributions and then use these models to assign a probability of being correct based on these learned models. With these tools it is possible to validate search engine results on large-scale datasets and in short order, enabling users to select probability thresholds based on a selected tolerable false discovery rate (FDR). The TPP is made fully interoperable via the open XML-based formats pepXML and protXML for different aspects of processing data-dependent acquisition (DDA), and Data-independent acquisition (DIA) proteomics data, resulting in a complete suite of tools for processing the increasingly larger datasets from start to finish.

DIA workflows are supported via the DISCO tool which reads mzML files containing the instrument-produced spectra and uses signal processing approaches to isolate the fragment ions in the multiplexed MS2 spectra that correlate with precursors in the MS1 and writes the results to new mzML files that may then be searched with standard DDA search engines and downstream tools, including target-decoy analysis. This provides a comprehensive analysis of DIA data without the need for building a spectral library first.

From its inception, the TPP has been and will always be free and open-source software, allowing anyone to use it without cost and to inspect its source code, alter the source code for their own needs, or even incorporate parts of it into their own

products. Others have performed these tasks and include various analysis routines as addons such as TAILS N-terminomics analysis,⁶⁷⁴ quantitation analysis with PyQuant,⁶⁷⁵ SimPhospho,⁶⁷⁶ WinProphet,⁶⁷⁷ ProtyQuant,⁶⁷⁸ and inclusion of R-tools for metaproteomic analysis.⁶⁷⁹ As a collection of individual tools, they are easily amenable to pipelining in a very flexible manner to support a huge variety of combinations and workflows, and a custom program may easily be inserted into the pipeline to support technology development.

Search Engines Supported by TPP. The heart of MS proteomics DDA data continues to be the “search engine” that interprets collections of mass spectra to determine the peptide or peptides that yielded them. Spectral library search engines and de novo search engines, which are less common, are also available and are included in software suites such as the Trans Proteomic Pipeline. A sequence search engine most commonly used is the open-source version of SEQUEST called Comet, which is actively maintained and updated with new functionality as needs arise. For spectral library searching, SpectraST uses an approach where new spectra are matched against a library of previously identified spectra in the form of a spectral library.⁶⁸⁰ This approach is much faster, more sensitive, and more specific than sequence database searching, although is only as good as the reference spectral library provided. There is renewed interest in spectral libraries because of data-independent acquisition (DIA) approaches being increasingly deployed and therefore the quality and coverage of libraries is paramount and likely to improve in the coming years, aided by the standard spectral library format being developed by the PSI.⁶⁸¹ For de novo sequence analysis, Novor⁶⁸² and Casanovo⁶⁸³ are very fast and capable de novo sequence search engines that are available.

For chemical crosslinking proteomics analysis, open-source programs such as Kojak^{74,75} are available for standard or cleavable MS2-based crosslinking techniques. Crosslinking-based MS analyses are employed to elucidate protein-protein interactions and facilitate protein structure and topology predictions. Kojak is designed to identify two independent peptides covalently bonded with a crosslinker and fragmented in a single MS2 scan event using a database search approach. Kojak algorithm also includes support for cleavable cross-linkers, and identification of cross-links between 15N-labeled homomultimers and is integrated into the Trans-Proteomic Pipeline, enabling access to dozens of additional tools, in particular, the PeptideProphet and iProphet tools for validation of cross-links improve the sensitivity and accuracy of correct cross-link identifications at user-defined thresholds. Development of Kojak has continued over the last ten years culminating in many improvements and new features. These improvements include support for additional open formats and standards, further refinement to the search algorithm for efficiency, E-values to normalize the scores of the results, support for cleavable cross-linkers, and methods to identify cross-links between homomultimer subunits.

For open modification database searching, programs such as Magnum⁶⁸⁴ are also now available which is specialized in identification of non-peptide masses that are bound to peptides. The tool is capable of identifying xenobiotic mass adducts, in addition to PTMs that were uncharacterized in the search parameters.

Table 9. Quality Control Software for Raw File and Real-Time Analysis

Name	Supported instrument vendors	Website/Download	Publication	Note
QuiC	Thermo Scientific, AB SCIEX, Agilent, Bruker, Waters	QuiC		Requires Biognosys Irt Peptides
AlphaPept	Thermo Scientific, Bruker	AlphaPept	688	
RawMeat 2.1	Thermo Scientific	RawMeat		
rawDiag	Thermo Scientific	rawDiag	689	
rawrr	Thermo Scientific	rawrr	690	
rawBeans	Thermo or mzML	rawBeans	691	
SIMPATIQCO	Thermo Scientific	SIMPATIQCO	692	
QC-ART		QC-ART	693	
SprayQc	Thermo Scientific, AB SCIEX, extensible to other instrumentation	SprayQc	694	
Metriculator		Metriculator	695	
MassQC		MassQC		
OpenMS		OpenMS	696	

Quality Control

Quality control should be a central aspect of any mass spectrometry-based study to ensure reproducibility of generated results. There are two types of quality controls that can be conducted for any kind of mass spectrometry experiment. First, QC approaches should monitor instruments themselves (e.g. HPLC and mass spectrometer), and second, QC approaches should assess the quality of data from unknowns or samples. For further reading, an entire issue was published on quality control in the journal *Proteomics* in 2011,⁶⁸⁵ especially the review by Köcher et al.,⁶⁸⁶ as well as the review published by Bittremieux et al. in 2017.⁶⁸⁷

QC: Instrument Performance. It is generally advisable to monitor instrument performance regularly. Instrument calibrations in regular intervals are essential to ensure that performance is maintained. Each instrument has a specific calibration method that is required. During the calibration you can check injection times (for ion trap instruments) and intensity of the ions in the calibration mix.

After ensuring good calibration and signal with the simple calibration mixture, it is advisable to analyze complex samples, such as commercial simple peptide mixtures or even whole tryptic digests of cell lysates (e.g., K562 standard from Promega). The additional check with a complex sample ensures all aspects of the system are working together correctly, especially the liquid chromatography and emitter. These digests should be analyzed after every instrument calibration and periodically between samples when acquiring more extensive batches. Data measured from tryptic digests should be analyzed by the software of your choice and the numbers of identified peptide precursors and proteins can be compared with previous controls for consistency.

Another strategy is to analyze digested purified proteins, which easily enable discovery of retention time shifts and mass accuracy problems. It is advised that new practitioners perform this manually at first to understand their data; this can be done manually by looking at m/z values of your standard peptides across runs in Skyline or the vendor-specific software. Looking at the intensity of the extracted peaks will help identify sensitivity fluctuations.

Carry-over between different measurements can be identified from blank measurements which are subsequently analyzed with your search software of choice. Blank measurements can be injections of different buffers, water or the starting conditions of your liquid chromatography. In case of increased detection of carry-over, injections with trifluoroethanol can be performed.

Another factor to take into consideration is the stability of your electrospray. Electrospray stability tends to worsen over time as columns wear and accumulate contaminants, such as salts or detergents, which can influence the quality of the emitter or column tip. You will notice spray instabilities either in the total ion chromatogram (TIC) as thin spikes with short periods of no measured signal or if you install cameras at your ESI source. Suboptimal spray conditions will usually result in droplets forming on the emitter, being released into the mass spectrometer (also referred to as “spitting”). Real-time quality control software (listed in Tables 9 and 10 below) can help you identify instrument issues right away.

QC: Samples. Apart from instrument performance, data analysis should start with QC to identify problematic measurements and to exclude them if necessary. It is recommended to develop a standardized system for data QC early on and to keep this consistent over time. Adding indexed retention time (iRT) peptides to samples can help identify and correct gradient and retention time inconsistencies between samples at the data analysis stage. Including common contaminants in FASTA files, such as keratins, is not considered optional because these contaminants are always present, and monitoring them can help identify sample preparation issues that may be non-uniform across samples. Other parameters to check in your analysis are the consistency of the number of peptide-spectrum matches, identified peptides and proteins over all samples, as well as coefficients of variation between your replicates. Before and after data normalization (if normalization is performed) it is good to compare the median intensities of all measurements to identify potential measurement or normalization issues. Precursor charge distributions, missed cleavage numbers, peak width, as well as the number of points per peak are additional parameters that can be checked. In case you are analyzing different conditions, perform hierarchical clustering or a principal component analysis to check if your samples cluster as expected.

Quantitative Proteomic Data Analysis Overview

This section aims to provide an overview of the best practices when conducting large scale proteomics quantitative data analysis. A universal workflow for proteomic data analysis does not currently exist because the processing depends on specific attributes of each individual dataset.⁷⁰¹ Analyzing proteomic data requires knowledge of a multitude of pre-processing techniques where order matters, and it can be challenging knowing where to start. This review will cover tools to reduce bias due to nonbiological variability, statistical methods to

Table 10. Search Result QC.

Name	Website/ Download/ Publication	Publication	Note
MSSStats	MSSStats	697	can use output from MaxQuant, Proteome Discoverer, Skyline, Progenesis, Spectronaut
MSSStatsQC	MSSStatsQC	698	
PTXQC	PTXQC	699	requires MaxQuant search engine output
protti	protti	700	

identify differential expression and machine learning (ML) methods for supervised or unsupervised interpretation of proteomic data. For a no-code option of all processing methods described in this section, we recommend Perseus.⁷⁰²

Data Transformation. Peptide or protein quantities are generally assumed to be logarithm (log) transformed before any subsequent processing.^{703–706} Log transformation allows data to more closely conform to a normal distribution and reduces the effect of highly abundant proteins.⁷⁰⁴ Many normalization techniques also assume data to be symmetric, so log transformation should precede any downstream analysis in these cases.⁷⁰⁴ If there are missing values present, a simple approach would be to use $\log(1+x)$ to avoid taking the log of zero. After the transformation, zero quantities will remain as zero and the other quantities should be large enough that adding one will have a minor effect.

Data Normalization. Data normalization, the process for adjusting data to have comparable distributions between samples, should almost always be performed prior to batch correction and any subsequent data analysis.^{704,707} This is required when the assumption is that most proteins are not changed between conditions, which is not always true. For example, in some studies of post translational modifications where kinases are inhibited, there may be true shifts in total signal between conditions, and normalization would mask those differences. This is also true in co-IP experiments where one condition may truly have many fewer binding partners. Normalization removes systematic bias in peptide/protein abundances that could mask true biological discoveries or give rise to false conclusions.⁷⁰⁸ Bias may be due to factors such as measurement errors and protein degradation,⁷⁰⁴ although the causes for these variations are often unknown.⁷⁰⁶ As data scaling methods should be kept at a minimum,⁷⁰⁹ a normalization technique well suited to address the nuances specific to one's data should be selected. The assumptions for a given normalization technique should not be violated, otherwise choosing the wrong technique can lead to misleading conclusions.⁷¹⁰ There are a multitude of data normalization techniques and knowing the most suitable one for a dataset can be challenging.

Visualization of peptide or protein intensity distributions among samples is an important step prior to selecting a normalization technique. Normalization is suggested to be done on the peptide level.⁷⁰⁹ If the technical variability causes the peptide/protein abundances from each sample to be different by a constant factor, and thus intensities are graphically similar across samples, then a central tendency normalization method such as mean, median or quantile normalization may be sufficient.^{704,709} However, if there is a linear relationship between bias and the peptide/protein abundances, a different method may be more appropriate.

To visualize linear and nonlinear trends due to bias, we can plot the data in a ratio versus intensity, or a M (minus) versus A (average), plot.^{704,711} Linear regression normalization is an available technique if bias is linearly dependent on peptide/protein abundance magnitudes.^{704,705} Alternatively, local regression (LOESS) normalization assumes nonlinearity between protein intensity and bias.⁷⁰⁵ Another method, removal of unwanted variation (RUV), uses information from negative controls and a linear mixed effect model to estimate unwanted noise, which is then removed from the data.⁷¹²

If sample distributions are drastically different, for example due to different treatments or samples are obtained from various tissues, one must use a method that preserves the heterogeneity in the data, including information present in outliers, meanwhile reducing systematic bias.⁷⁰⁹ For example, Hidden Markov Model (HMM)-assisted normalization,⁷⁰⁹ RobNorm,⁷¹³ or EigenMS⁷¹⁴ may be suitable for this type of data. These techniques assume error is only due to the batch and order of processing. The first method that addresses correlation of errors between compounds by using the information from the variation of one variable to predict another is systematic error removal using random forest (SERRF).⁷¹⁵ SERRF, among 14 normalization methods, was the most effective in significantly reducing systematic error.⁷¹⁵

Studies aiming to compare these methods for omics data normalization have come to different conclusions. Ranking of different normalization methods can be done by assessing the percent decrease in median \log_2 (standard deviation) and \log_2 pooled estimate of variance (PEV) in comparison to the raw data.⁷¹⁶ One study found linear regression ranked the highest compared to central tendency, LOESS and quantile normalization for peptide abundance normalization for replicate samples with and without biological differences.⁷⁰⁴ A paper comparing multiple normalization methods using a large proteomic dataset found that mean/median centering, quantile normalization and RUV had the highest known associations between proteins and clinical variables.⁷⁰³ Rather than individually implementing normalization techniques, which can be challenging for non-domain experts, there are several R and Python packages that automate mass spectrometry data analysis and visualization. These tools assist with making an appropriate selection of a normalization technique. For example, NormalyzerDE, an R package, includes several popular methods for normalization and differential expression analysis of LC-MS data.⁷¹⁷ AlphaPeptStats,⁷¹⁸ a Python package, allows for comprehensive mass spectrometry data analysis, including normalization, imputation, batch correction, visualization, statistical analysis and graphical representations including heatmaps, volcano plots, and scatter plots. AlphaPeptStats allows for analysis of label-free proteomics data from several platforms (MaxQuant, AlphaPept, DIA-NN, Spectronaut, FragPipe) in Python but also has web version that does not require installation. For both transformation and normalization, we recommend using the options in scikit-learn in python.

Data Imputation. Missing peptide intensities, which are common in proteomic data, may need to be addressed, although this is a controversial topic in the field. Normalization should be performed before imputation since bias may not be removed to detect group differences if imputation occurs prior to normalization.⁷⁰⁶ Reasons for missing data include the peptide not being biologically present, being present but at too

low of a quantity to be detected, or present at quantifiable abundance but misidentified or incorrectly undetected.⁷⁰⁶ If the quantity is not at the detectable limit, the quantity is called censored and these values are missing not at random.⁷⁰⁶ Imputing these censored values will lead to bias as the imputed values will be overestimated.⁷⁰⁶ However, if the quantity is present at detectable limits but was missed due to a problem with the instrument, this peptide is missing completely at random (MCAR).⁷⁰⁶ While imputation of values that are MCAR using observed values would be a reasonable approach, censored peptides should not be imputed because their missingness is informative.⁷⁰⁶ Peptides MCAR are a less frequent problem compared to censored peptides.⁷⁰⁶ Understanding why the peptide is missing can be challenging,⁷⁰⁶ however there are techniques such as maximum likelihood model⁷¹⁹ or logistic regression⁷²⁰ that may distinguish censored versus MCAR values.

Commonly used imputation methods for omics data are random forest (RF) imputation,⁷²¹ k-nearest neighbors (kNN) imputation,⁷²² and single value decomposition (SVD).⁷²³ Using the mean or median of the non-missing values for a variable is an easy approach to imputation but may lead to underestimating the true biological differences.⁷⁰⁶ Choice of the appropriate imputation method is critical as how these missing values are filled in has a substantial impact on downstream analysis and conclusions.⁷²⁴ In one study, RF imputation was the most accurate among nine imputation methods across several combinations of types and rates of missingness and does not require preprocessing (e.g., does not require normal distribution) for metabolomics data.⁷²⁵ Another study found RF, among eight imputation methods, had the lowest normalized root mean squared error (NRMSE) between imputed values and the actual values when MCAR values were randomly replaced with missing values, followed by SVD and KNN using metabolomics data.⁷²⁴ Lastly, a study found RF also had the lowest NRMSE when comparing seven imputation methods using a large-scale label-free proteomics dataset.⁷²⁶ In general for imputation, we recommend using missforest, which is available as both a R and python package.

Batch Correction

Normalization is assumed to occur prior to batch effect correction.⁷⁰⁹ Batch effect correction is still a critical step after normalization as proteins may still be affected by batch effects and diagnosing a batch effect may be easier once data is normalized.⁷⁰⁹ Prior to performing any statistical analysis of data, we must start with distinguishing signals in the data due to biological versus batch effects. A batch effect occurs when differences in preparation of samples and how data was acquired between batches results in altered quantities of peptides (or genes or metabolites) which results in reduced statistical power in detecting true differences.⁷⁰⁹ This non-biological variability originates from the time of sample collection to peptide/protein quantification⁷⁰³ and is often a problem when working with large numbers of samples, involving multiple plates run by different technicians, on different instruments and/or using different reagent batches.⁷²⁷ Results from these different batches ultimately need to be aggregated and data analysis to be performed on the whole dataset, so it may be difficult to measure and then control for exact changes due to non-biological variability once the data has been aggregated.⁷⁰³ Batch correction methods remove technical variability, however they should not remove any true

biological effect.^{703,727} Although it is agreed upon that these biases should be accounted for to prevent misleading conclusions, there is no one gold standard batch correction method.

Batch effects can manifest as continuous, such as from MS signal drift, or as discrete, such as a shift that affects the entire batch.⁷⁰⁹ To visualize batch effects, one can plot the average intensity per sample in the order each was measured by the MS to see if intensities are shifted in a certain batch.⁷⁰⁹ Measuring protein-protein correlations is another method to check for batch effects; if proteins within a batch are more correlated compared to those from other batches, there are likely batch effects.⁷⁰⁹ Prior to batch correction, one should ensure the experimental design is not inherently flawed due to batch effects and whether a change in design should be implemented. Studies spanning multiple days and experiments involving samples from different centers are vulnerable to batch effects.⁷²⁸ One example of technical variability that may irreversibly flaw an experiment would be running samples at varying time points, or “as they came in”.⁷²⁹ This problem can be circumvented by balancing biological groups in each batch.⁷²⁹ Additionally, collection of samples at different institutions introduces non-biological variability due to differences in a multitude of conditions such as collection protocols, storage, and transportation.⁷⁰³ A solution to this problem would be to evenly distribute samples between centers or batches.⁷⁰³

There are several batch correction methods, the most popular method being Combating Batch Effects When Combining Batches of Gene Expression Microarray Data (ComBat), originally designed for genomics data.^{727,730} ComBat uses Bayesian inference to estimate batch effects across features in a batch and applies a modified mean shift, but requires peptides to be present in all batches which can lead to loss of a large number of peptides.⁷²⁷ ComBat is available as a python package called pycombat. Out of six batch correction methods using microarray data, ComBat was the best in reducing batch effects across several performance metrics and was effective using high dimensional data with small sample sizes.⁷³¹ ComBat may be more suitable for small datasets when the source of batch effects are known.⁷²⁷ However if potential batch variables are not known or processing time or group does not adequately control for batch effects, surrogate variable analysis (SVA) may be used where the source of batch effect is estimated from the data.^{727,728} A third option for batch effect correction uses negative control proteins to estimate unwanted variation, called “Remove Unwanted Variation, 2-step” (RUV-2).⁷³² There are many additional batch effect correction methods for single cell data, such as mutual nearest neighbors,⁷³³ or Scanorama, which generalizes mutual nearest neighbors matching.⁷³⁴

Assessment of Transformed Data

Prior to conducting any statistical analysis, the raw data matrix should be compared to the data after the above-described preprocessing steps have been performed to ensure bias is removed. We can compare data using boxplots of peptide intensities from the raw data matrix versus corrected data in sample running order to look at batch associated patterns; after correction, we should see uniform intensities across batches.⁷⁰⁹ We can also use clustering methods such as Principal Component analysis (PCA), Uniform Manifold Approxima-

tion and Projection (UMAP), or t-SNE (t-Distribute Stochastic Neighbor Embedding) and plot protein quantities colored by batches or technical versus biological samples to see how proteins cluster in space based on similarity. We can measure the variability each PC contributes; we want to see similar variability among all PCs, however if we see one PC contributing to overall variability highly then means variables are dependent.⁷³⁵ tSNE and UMAP allow for non-linear transformations and allow for clusters to be more visually distinct.⁷³⁵ Grouping of similar samples by batch or other non-biological factors, such as time or plate, indicates bias.⁷⁰⁹ Quantitative measures of whether batch effects have been removed are principal variance components analysis (PVCA), which provides information on factors driving variance between biological and technical differences, and checking correlation of samples between different batches, within the same batch and between replicates. When batch effects are present, samples in the same batch will have higher correlation than samples from different batches and between replicates.⁷⁰⁹ Once batch effects are removed, proteins in the same batch should be correlated at the same level with proteins from other batches.⁷⁰⁹ Similarity between technical replicates can be measured using pooled median absolute deviation (PMAD), pooled coefficient of variation (PCV) and pooled estimate of variance (PEV); high similarity would mean batch effects are removed and there is low non-biological effects.⁷⁰⁵

Lastly, it is also important to show that batch correction leads to improvement in finding true biological differences between samples. We can show the positive effect that batch correction has on the data by demonstrating reproducibility after batch correction. One way to provide evidence for reproducibility is to show that prior to batch correction, there was no overlap between differentially expressed proteins between groups in one batch with those found between the same groups in another batch and, after batch correction, the differentially expressed proteins between the groups become the same between batches.⁷⁰⁹ This applies generally datasets with large numbers (e.g., hundreds) of samples to allow for meaningful statistical comparisons.⁷⁰⁹

Statistical Analysis

Once the above pre-processing steps have been applied to the dataset, we can investigate if any protein quantities differ between groups. There is an urgent need for biomarkers for disease prediction and there is large potential for protein based biomarker candidates.⁷³⁶ However, omics datasets are often limited due to having many more features than number of samples, which is termed the “curse of dimensionality”.⁷⁰¹ Attributes that are redundant or not informative can reduce the accuracy of a model.⁷³⁷

Univariate statistical tests including t-tests and analysis of variance (ANOVA) provide p-values to allow ranking the importance of variables.⁷⁰¹ T-tests are used in pairwise comparisons, and ANOVA is used when there are multiple groups to ask whether any group is different from the rest. After ANOVA, the Tukey's posthoc test can reveal which pairwise differences are present among the multiple groups that were compared. There are many posthoc tests that can be used, and guidance from an expert statistician is suggested. Wilcoxon rank-sum tests can be used if the data are still not normal after the above approaches and therefore violates the assumptions required for a t-test. Kruskal-Wallis test is the

non-parametric version of ANOVA useful for three or more groups when assumptions of ANOVA are violated.

Data can be reduced using a feature selection method, which includes either feature subset selection, where irrelevant features are removed, or feature extraction, where there is a transformation that generates new, aggregated variables and do not lead to loss of information.⁷⁰¹ An example of a commonly used multivariate feature extraction method using proteomic data is principal component analysis (PCA).⁷⁰¹

Multiple hypothesis testing happens in proteomics when we make many statistical tests to check for differences of many measured proteins between conditions. For example, if we measure 1000 proteins between a pair of conditions, we may perform 1000 *t*-tests. By random chance, even in the absence of true protein quantity changes, 10 of these tests will produce a *p*-value less than 0.01. These would be false positives (i.e., a *p*-value will appear significant by chance),⁷³⁵ and multiple testing correction should be applied to manage the overall false positive rate at less than a specified cutoff.⁷⁰¹ There are many methods for multiple testing correction. Benjamini-Hochberg correction is less stringent than the Bonferroni correction, which leads to too many false negatives, and thus is a more commonly used multiple testing correction method.⁷⁰¹

Volcano plots allow visualization of differentially abundant proteins by displaying the negative log of the adjusted *p*-value as a function of the log fold change, a measure of effect size, for each protein. Points with larger *y* axis values are more statistically significant and those further away from zero on the *x* axis have a larger fold change. There are two methods for identifying differentially expressed proteins. The first method involves a combined adjusted *p*-value cutoff (*y* axis) and fold change cutoff (*x* axis) to create a “square cutoff”.⁷³⁵ The second involves a non-linear cutoff, where a systematic error is added to all the standard deviations used in the *t*-tests.⁷³⁵

There are other statistical tests to consider for quantitative proteomics data. Another popular statistical method in proteomics when dealing with high dimensional data is lasso linear regression, which removes regression coefficients from the model by applying a penalty parameter.⁷³⁸ Bayesian models are an emerging technique for protein based biomarker discovery that are more powerful than standard *t*-tests⁷³⁸ and have outperformed linear models.^{738,739} Bayesian models incorporate external information into the prior distribution; for example, knowledge of peptides that usually have more technical variability are assigned a less informative prior.⁷³⁸ Prior to implementing machine learning (ML), one can start with the simpler models, such as linear regression or naïve Bayes.⁷³⁶

In general for statistical tests, one can suggest using the standard test available in base R or in python packages *scipy* or *statsmodels*.

Machine Learning (ML)

Despite the efficacy of ML for finding signals in a high dimensional feature space to distinguish between classes,⁷³⁶ the application of ML to proteomic data analysis is still in its early stages⁷³⁶ as only 2% of proteomics studies involve ML.⁷⁴⁰ The reason for such sparse usage of ML in proteomics data is the need for very large datasets comprising 100s of samples, which is still rare in proteomics.

Supervised classification is the most common type of ML used for proteomic biomarker discovery, where an algorithm has been trained on variables to predict the class labels of

unseen test data.⁷³⁷ Supervised means the class labels, such as disease versus controls, are known.⁷⁴⁰ Decision trees are common model choice due to their many advantages: variables are not assumed to be linearly related, models are able to rank more important variables on their own, and interactions between variables do not need to be pre-specified by the user.⁷³⁸ There are three phases of model development and evaluation.⁷⁴¹ In the first step, the dataset is split into training and testing splits, commonly 70% training and 30% testing. Second, the model is constructed using only the training data, which is further subdivided into training and test sets. During this process, an internal validation strategy, or cross-validation (CV), is employed.⁷⁰¹ Commonly used CV methods in proteomics are k-fold and leave-one-out cross-validation.⁷⁰¹ The final step is to evaluate the model on the testing set that was held-out in step one. There should not be overlap between the training and testing data, and the testing data should only be evaluated once after all training has been completed. The dataset used for training and testing should be representative of the population that is to be eventually tested. If under-represented groups are lacking from models during training, these models will not generalize to these populations.⁷⁴² Proteomic data and patient specific factors derived from the electronic health record (EHR) like age, race, and smoking status can be employed as inputs to a model.⁷³⁶ However, addition of EHR data may not be informative in some instances; in studying Alzheimer's Disease, adding these patient specific variables were informative for non-Hispanic white participants, but not for African Americans.⁷⁴²

A common mistake in proteomics ML studies is allowing the test data to leak into the feature selection step.^{740,743} It has been reported that 80% of ML studies on the gut microbiome performed feature selection using all the data, including test data.⁷⁴³ Including the testing data in the feature selection step leads to development of an artificially inflated model⁷⁴³ that is overfit on the training data and performs poorly on new data.⁷³⁶ Feature selection should occur only on the training set and final model performance should be reported using the unseen testing set. The number of samples should be ten times the number of features to make statistically valid comparisons, however this may not be possible in many cases.⁷⁴⁴ If a study is limited by its number of samples, one can perform classification without feature selection.⁷⁴³

Pitfalls also arise when a ML classifier is trained using an imbalanced dataset.⁷⁴¹ Proteomics biomarker studies commonly have imbalanced groups, where the number of samples in one group is drastically different from another group. Most ML algorithms assume balanced number of samples per class and not accounting for these differences can lead to reduced performance and construction of a biased classifier.⁷⁴⁵

Care should be practiced when choosing an appropriate metric when dealing with imbalanced data. A high accuracy may be meaningless in the case of imbalanced classification; the number of correct predictions will be high even with a blind guess for the majority class.⁷⁴⁶ F1 score, Matthews correlation coefficient (MCC), and area under the precision recall curve (AUPR) are preferred metrics for imbalanced data classification.^{747,748} MCC, for example, is preferred since it is only high if the model predicts correctly on both the positive and negative classes.⁷⁴¹ Over- and under-sampling to equalize the number of samples in classes are potential methods to address class imbalance, but can be ineffective or even detrimental to the performance of the model.⁷⁴⁵ These

sampling methods may lead to a poorly calibrated model that overestimates the probability of the minority class samples and reduces the model's applicability to clinical practice.⁷⁴⁶

■ PROTEIN SEQUENCE DATABASES

Where Do You Get them?

For those looking for guidance on where to obtain a database for their organism of interest quickly, we recommend going to uniprot.org and using their "proteomes": https://www.uniprot.org/proteomes?query=*. After selecting the proteome of interest, for most applications, we recommend clicking the "download one sequence per gene (FASTA)", for example, on the left of this page for *E. coli*: <https://www.uniprot.org/proteomes/UP000000625>.

Protein Database Sources and Types

Many mass spectrometry-based proteomic techniques use search algorithms that require a defined theoretical search space to identify peptide sequences based on precursor mass and peptide fragmentation patterns, which are then used to infer the presence and abundance of a protein. Traditionally these databases are used with DDA database search algorithms, but they can also be used with new MS/MS spectra prediction algorithms to predict spectral libraries for DIA data analysis. The search space is calculated from the potential proteins in a sample, which includes the proteome (often a single species) and expected contaminants. This is called database searching and the flat file of protein sequences in FASTA format acts as a protein database. In this section, we will describe major resources for proteome FASTA files (protein sequence collections), how to retrieve them, and suggested best practices for preserving FASTA file provenance to improve reproducibility.

In general, FASTA sequence collections can be retrieved from three central clearing houses: UniProt, RefSeq, and Ensembl. These will be discussed separately below as they each have specific design goals, data products, and unique characteristics. It is important to learn the following three points for each resource: the source of the underlying data, canonical versus non-canonical sequences, and how versioning works. These points, along with general best practices, such as using a taxonomic identifier, are essential to understand and communicate search settings used in analyses of proteomic datasets. Finally, it is critical to understand that sequence collections from these three resources are not the same, nor do they offer the same sets of species.

Key terminology may vary between resources, so these terms are defined here. The term "taxon identifier" is used across resources and is based on the NCBI taxonomy database. Every taxonomic node has a number, e.g., *Homo sapiens* (genus species) is 9606 and Mammalia (class) is 40674. This can be useful when retrieving and describing protein sequence collections. Another term used is "annotation", which has different meanings in different contexts. Broadly, a "genome annotation" is the result of an annotation pipeline to predict coding sequences, and often a gene name/symbol if possible. Two examples are MAKER⁷⁴⁹ and the RefSeq annotation pipeline.⁷⁵⁰ Alternatively, "protein annotation" (or gene annotation) often refers to the annotation of proteins (gene products) using names and ontology (i.e., protein names, gene names/symbols, functional domains, gene ontology, keywords, etc.). Protein annotation is termed "biocuration" and described in detail by UniProt.⁷⁵¹ Lastly, there are established

minimum reporting guidelines for referring to FASTA files established in MIAPE: Mass Spectrometry Informatics that are taxon identifier and number of sequences.^{752,753} The FASTA file naming suggestions below are not official but are suggested as a best practice.

UniProt. The Universal Protein Resource (UniProt),^{754–757} has three different products: UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc). The numerous resources and capabilities associated with the UniProt are not explored in this section, but these are well described on UniProt's website. UniProtKB is the source of proteomes across the Tree of Life and is the resource we will be describing herein. There are broadly two types of proteome sequence collections: Swiss-Prot/TrEMBL and designated proteomes. The Swiss-Prot/TrEMBL type can be understood by discussing how data is integrated into UniProt. Most protein sequences in UniProt are derived from coding sequences submitted to EMBL-Bank, GenBank and DDBJ. These translated sequences are initially imported into TrEMBL database, which is why TrEMBL is also termed “unreviewed”. There are other sources of protein sequences, as described by UniProt.⁷⁵⁸ These include the Protein Data Bank (PDB), direct protein sequencing, sequences derived from the literature, gene prediction (from sources such as Ensembl) or in-house prediction by UniProt itself. Protein sequences can then be manually curated into the Swiss-Prot database using multiple outlined steps (described in detail by UniProt here: ref 759) and is why Swiss-Prot is also termed “reviewed”. Note that more than one TrEMBL entry may be removed and replaced by a single Swiss-Prot entry during curation. A search of “taxonomy_id:9606” at UniProtKB will retrieve both the Swiss-Prot/reviewed and TrEMBL/unreviewed sequences for *Homo sapiens*. The entries do not overlap, so users often either use just Swiss-Prot or Swiss-Prot combined with TrEMBL, the latter being the most exhaustive option. With ever-increasing numbers of high-quality genome assemblies processed with robust automated annotation pipelines, TrEMBL entries will contain higher quality protein sequences than in the past. In other words, if a mammal species has 20 000 to 40 000 entries in UniProtKB and many of these are TrEMBL, users should be comfortable using all the protein entries to define their search space (more on this later when discussing proteomes at UniProtKB). Determining the expected size of a well-annotated proteome requires additional knowledge, but tools to answer these questions continue to improve. As more and more genome annotations are generated, the backlog of manual curation continues to increase. However, automated genome annotations are also rapidly improving, blurring the line between Swiss-Prot and TrEMBL utility.

The second type of protein sequence collections available at UniProtKB are designated proteomes, with subclasses of “proteome”, “reference proteome” or “pan-proteome”. As defined by UniProt, a proteome is the set of proteins derived from the annotation of a completely sequenced genome assembly (one proteome per genome assembly). This means that a proteome will include both Swiss-Prot and TrEMBL entries present in a single genome annotation, and that all entries in the proteome can be traced to a single complete genome assembly. This aids in tracking provenance as assemblies change, and metrics of these assemblies are available. These metrics include Benchmarking Universal Single-Copy Ortholog (BUSCO) score, and “Completeness”

as Standard, Close Standard or Outlier based on the Complete Proteome Detector (CPD). Given the quality of genome annotation pipelines, using a proteome as a FASTA file for a species is the preferred method of defining search spaces now. Outside of humans, no higher eukaryotic Swiss-Prot sequence collections are complete enough for use in proteomics analyses, but this does not mean that the available Swiss-Prot plus TrEMBL protein sequence collection precludes accurate proteomic data analysis. Lastly, the difference between reference proteome and proteome is used to highlight model organisms or organisms of interest, but not to imply improved quality. UniProt also has support for the concept of “pan proteomes” (consensus proteomes for a closely related set of organisms) but this is mostly used for bacteria (e.g., strains of a given species will share a pan proteome).

When retrieving protein sequence collections as Swiss-Prot/TrEMBL or designated proteomes, there is an option of downloading “FASTA (canonical)” or “FASTA (canonical & isoform)”. The latter includes additional manually annotated isoforms for Swiss-Prot sequences. Each Swiss-Prot entry has one canonical sequence chosen by the manual curator. Any additional sequence variants (mostly from alternative slicing) are annotated as differences with respect to the canonical sequence. Specifying “canonical” will select only one protein sequence per Swiss-Prot entry while specifying “canonical & isoforms” will download additional protein sequences by including isoforms for Swiss-Prot entries. Recently, an option to “download one protein sequence per gene (FASTA)” has been added. These FASTA files include Swiss-Prot and TrEMBL sequences to number about 20 000 protein sequences for a wide range of higher eukaryotic organisms.

The number of additional isoforms in a proteome varies considerably by species. In the human, mouse, and rat proteomes of the total number of entries, 25 %, 40 % and 48 % are canonical, respectively. The choice of including isoforms is related to the search algorithm and experimental goals. For instance, if differentiating isoforms is relevant, they should be included otherwise they will not be detected. In cases where isoforms are present in the FASTA (evident by shared protein names) but these cannot be removed prior to downloading (e.g., California sea lion, *Zalophus californianus*, proteome UP000515165, release 2023_04 has no options for downloading one protein sequence per gene), non-redundant FASTA files can be manually generated (i.e., “remove_duplicates.py” via ref 760). If possible, retrieving canonical protein sequences via proteomes is the most straight forward approach and in general appropriate for most search algorithms, versus the method of searching and downloading Swiss-Prot and/or TrEMBL entries.

Though FASTA files are the typical input of many search algorithms, UniProt also offers an XML and GFF format download. In contrast to the flat FASTA file format, the XML format includes sequence information as well as associated information like PTMs, which is used in some search algorithms like MetaMorpheus.⁷⁶¹

Once a protein sequence collection has been selected and retrieved, how can the file be named and report this to others in a way that allows them to reproduce the retrieval? The minimum reporting information is the taxon identified and number of sequences used.^{752,753} The following naming format (and those below) augments this and is suggested for UniProtKB FASTA files (the use of underscores or hyphens is not critical):

[common or scientific name]-[taxon id]-uniprot-[swiss-prot/trembl/proteome]-[UP# if used]-[canonical/canonical plus isoform]-[release]

example of a *Homo sapiens* (human) protein fasta from UniProtKB:

Human-9606-uniprot-proteome-UP000005640-canonical-2023_04.fasta

The importance of the taxon identifier has already been described above and is a consistent identifier across time and shared across resources. The choices of Swiss-Prot and TrEMBL in some combination was discussed above, and Proteome can be “proteome”, “reference proteome” or “pan-proteome”. The proteome identifier (“UP” followed by 9 digits) is conserved across releases, and release information should also be included. A confusing issue to newcomers is what the term “release” means. This is a year_month format (e.g., 2023_04), but it is not the date a FASTA file was downloaded or created, nor does it imply there are monthly updates. This release “date” is a traceable release identifier that is listed on UniProt’s website. Including all this information ensures that the exact provenance of a FASTA file is known and allows the FASTA file to be regenerated.

RefSeq. NCBI is a clearing house of numerous types of data and databases. Specific to protein sequence collections, NCBI Reference Sequence Database (RefSeq) provides annotated genomes across the Tree of Life. The newly developed NCBI Datasets portal⁷⁶² is the preferred method for accessing the myriad of NCBI data products, though protein sequence collections can also be retrieved from RefSeq directly.^{763,764} Like UniProt described above, most of the additional functionality and information available through NCBI Datasets and RefSeq will not be described here, although the Eukaryotic RefSeq annotation dashboard⁷⁶⁵ is a noteworthy resource to monitor the progress of new or re-annotations. We recommend exploring the resources available from NCBI,⁷⁶⁶ utilizing their tutorials and help requests.

RefSeq is akin to the “proteome” sequence collection from UniProtKB, where a release is based on a single genome assembly. If a more complete genome assembly is deposited or additional secondary evidence (e.g., RNA sequencing) is deposited, RefSeq can update the annotation with a new annotation release. Every annotation release will have an annotation report that contains information on the underlying genome assembly, the new genome annotation, secondary evidence used, and various statistics about what was updated. The current annotation release is referred to as the “reference annotation”, but each annotation is numbered sequentially starting at 100 (the first release), though a recent naming change has abandoned the sequential release numbering and instead is the RefSeq assembly “-RS” and then the year month when it was annotated (e.g., the current human reference annotation is GCF_000001405.40-RS_2023_10). Certain species are on scheduled re-annotation, like human and mouse, while other species are updated as needed based on new data and community feedback (ex. release 100 of taxon 9704 was in 2018, but a more contiguous genome assembly resulted in re-annotation to release 101 in 2020). This general process for new and existing species is described by Heck and Neely.⁷⁶⁷

Since RefSeq is genome assembly-centric, its protein sequence collections are retrieved for each species. This contrasts with being able to use a higher-level taxon identifier like 40674 (Mammalia) in UniProt to retrieve a single FASTA.

To accomplish this same search in NCBI Datasets requires a Mammalia search, followed by browsing all 2847 genomes and then filtering the results to reference genomes with RefSeq annotations, and those resulting 223 could be bulk downloaded, though this will still be 223 individual FASTA files. It is possible to download a single FASTA from an upper-level taxon identifier using the NCBI Taxonomy Browser, though this service may be redundant with the new NCBI Datasets portal. Given the constant development of NCBI Datasets, these functionalities may change, but the general RefSeq philosophy of single species FASTA should be kept in mind. Likewise, when retrieving genome annotations there is no ability to specify canonical entries only, but it is possible to use computational tools to remove redundant entries (“remove_duplicates.py” from ref 760).

Similar to the UniProtKB FASTA file naming suggestion, the following naming format is suggested for RefSeq protein sequence collection FASTA (the use of underscores or hyphens is not critical):

[common or scientific name]-[taxon id]-refseq-[release number]

Example of a *Equus caballus* (horse) protein FASTA from RefSeq:

Equus_caballus-9796-refseq-103.fasta

The release number starts at 100 and is consecutively numbered. Note, the human releases previously had a much longer number to be included (e.g., NCBI Release 109.20211119), then began following a consecutive numbering for Release 110, but have now switched to the new format related to assembly and annotation date. Also, in a few species (Human, Chinese hamster, and Dog, currently), there is a reference and an alternate assembly, both with an available annotation. In these cases, including the underlying assembly identifier would be needed. Note that when you retrieve the protein FASTA from NCBI it will include two more identifiers that aren’t required in the file name since it can be determined from the taxon identifier and release number. These are the genome assembly used (this is generated by the depositor and follows no naming scheme) and the RefSeq identifier (GCF followed by a number string). These aren’t essential for FASTA naming, but are for comparing between UniProt, RefSeq and Ensembl when the same underlying assembly is used (or not, indicating how up to date one is versus the other).

Ensembl. There are two main web portals for Ensembl sequence collections: the Ensembl genome browser⁷⁶⁸ has vertebrate organisms, and the Ensemble Genome project⁷⁶⁹ has specific web portals for different non-vertebrate branches of the Tree of Life. This contrasts with NCBI and UniProt where all branches are centrally available. Recently, Ensembl has created a new portal “Rapid Release” focusing on quickly making annotations available (replacing the “Pre-Ensemble” portal), albeit without the full functionality of the primary Ensembl resources. Overall, Ensembl provides diverse comparative and genomic tools that should be explored, but, specific to this discussion, they provide species-specific genome annotation products similar to RefSeq.

To retrieve a protein sequence collection from Ensembl at any of the portals, a species can be searched using a name, which will then have taxon identifier displayed (but searching by identifier is not readily apparent). From the results you can select your species and follow links for genome annotation. Caution should be used when browsing the annotation

products since the protein coding sequence (abbreviated “cds”) annotations are nucleic acid sequences (a useable via 3-frame translation if using certain software), while actual translated peptide sequences are in the “pep” folders. The pep folders contain file names with “ab initio” and “all” in the FASTA file names (file extensions are “fa” for FASTA and “gz” indicating gzip compression algorithm), while there may only be one pep product for certain species in the “Rapid Release” portal. The “ab initio” FASTA files contain mostly predicted gene products. The “all” FASTA files are the usable protein sequence collections. Ensembl FASTA files usually have some protein sequence redundancy.

Ensembl provides a release number for all the databases within each portal. Similar to the UniProt file naming suggestion, the following naming format is suggested for Ensembl protein sequence collection FASTA (the use of underscores or hyphens is not critical):

[common or scientific name]-[taxon id]-ensembl-[abinitio/all]-[rapid]-[release number]

Example of a *Sus scrofa* (pig) protein FASTA from Ensembl: Pig-9823-ensembl-all-106.fasta

Similar to the FASTA download from RefSeq, the downloaded file name can include additional identifying information related to the underlying genome assembly. Again, this is not required for labeling, but is useful to easily compare assembly versions.

Since much of the data from Ensembl is also regularly processed into UniProt, using UniProt sequence collections instead may be preferred. That said, they are not on the same release schedule nor will the FASTA files contain the same proteins. Ensembl sequences still must go through the established protein sequence pipeline at UniProt to remove redundancy and conform to UniProt accession and FASTA header formats. Moreover, the gene-centric and comparative tools built into Ensembl may be more experimentally appropriate and using an Ensembl protein sequence collection can better leverage those tools.

Other Resources. There are other locations of protein sequence collections, and these will likewise have different FASTA file formatting; sequences may have unusual characters, and formats of accessions and FASTA header lines may need to be reformatted to be compatible with search software. These alternatives include institutes like the Joint Genome Institute’s microbial genome clearing house, species-specific community resource (e.g., PomBase, FlyBase, WormBase, TrypDB, etc.), and one-off websites tenuously hosting in-house annotations. It is preferred to use protein sequence collection from the main three sources described here, since provenance can be tracked, and versions maintained. It is beyond the scope of this discussion to address other genome annotation resources, how they are versioned, or the best way to describe FASTA files retrieved from those sources. In these cases, defaulting to the minimum requirements of listing number of entries and supplying the FASTA along with data are necessary.

Contaminants. Samples are rarely comprised of only proteins from the species of interest. There can be protein contamination during sample collection or processing. This may include proteins from human skin, wool from clothing, particles from latex, and even trypsin itself, all of which contain proteins that can be digested along with the intended sample and analyzed in the mass spectrometer. Avoiding unwanted matching of mass spectra originating from contaminant

proteins to the cellular proteins due to sequence similarities is important to the identification and quantitation of as many cellular proteins as possible. To avoid these spectra matching to the wrong peptides, repositories of supplementary sequences for contaminant proteins have been added to a reference database for MS data searches. Appending a contaminants database to the reference database allows the identification of peptides that are not exclusive to one species.

As early as 2004, The Global Proteome Machine was providing a protein sequence collection of these common Repository of Adventitious Proteins (cRAP), while another contaminant list was published in 2008.⁷⁷⁰ The current cRAP version (v1.0) was described in 2012⁷⁷¹ and is still widely in use today. cRAP is the contaminant protein list used in nearly all modern database searching software, though the documentation, versioning or updating of many of these “built-in” contaminant sequence collections is difficult to follow. There is also another contaminant sequence collection distributed with MaxQuant. Together, the cRAP and MaxQuant contaminant protein sequence collections are found in some form across most software, including MetaMorpheus and Philosopher (available in FragPipe).⁷⁷² This list of known frequently contaminating proteins can either be automatically included by the software or can be retrieved as a FASTA to be used along with the primary search FASTA(s). Recently the Hao Lab has revisited these common contaminant sequences in an effort to update the protein sequences (ProtContLib), test their utility on experimental data, and add or remove entries.⁷⁷³

In addition to these environmentally unintended contaminants, there are known contaminants that also have available protein sequence collections (or can be generated using the steps above) and should be included in the search space. These can include the media cells were grown in (e.g., fetal bovine serum,^{774,775} food fed to cells/animals (e.g., *Caenorhabditis elegans* grown on *Escherichia coli*) or known non-specific binders in affinity purification (i.e., CRAPome⁷⁷⁶). The common Repository of Fetal Bovine Serum Proteins (cRFP)⁷⁷⁷ are protein lists of common protein contaminants and fetal serum bovine sequences used to reduced the number of falsely identified proteins in cell culture experiments. Cells washed or cultured in contaminant free media before harvest or the collection of secreted proteins depletes most high abundance contaminant proteins but the sequence similarity between contaminant and secreted proteins can cause false identifications and overestimation of the true protein abundance leading to wasted resources and time on validating false leads. As emphasized throughout this section, accurately defining the search space is essential for accurate results and, especially in the case of contaminants, requires knowledge of the experiment and sample processing to adequately define possible background proteins.

Choosing the Right Database

Proteomics data analysis requires carefully matching the search space (defined by the database choice) with the expected proteins. A properly chosen database will minimize false positives and false negatives. Choosing a database that is too large will increase the number of false positives, or decoy hits, which in turn will reduce the total number of identifiable proteins. For this reason it is ill advised to search against all possible protein sequences ever predicted from any genomic sequence. On the other hand, choosing a database that is too small may increase false negatives, or missed protein

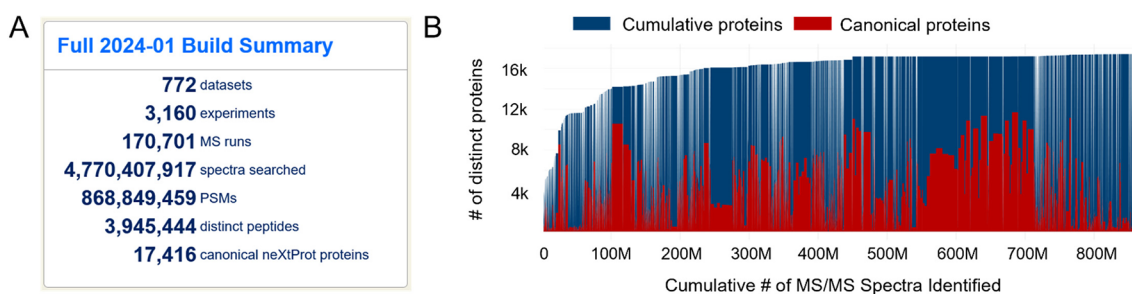


Figure 17. Human Peptide Atlas as of 2024. (A) Current total search space and identified elements of the 2024 human PeptideAtlas. (B) Historical cumulative plot of the identified total proteins (blue vertical bars) and the unique proteins identified per dataset (red vertical bars) over the period of 2005–2024

identifications, because in order for a protein to be identified it must be present in the database. Some search algorithms can self-correct when a database is overly large such that higher identity thresholds are required for identification to minimize false positives (e.g., Mascot), while smaller experiment-specific search spaces (also referred to as “subsets”) can have unintended effects on false positives if not managed appropriately^{778–780} or may even improve protein identifications.⁷⁸¹ Whether to employ a search space that is sample-specific (i.e., subset), species-specific (with only canonical proteins, described below), exhaustive species-specific (including all isoforms), or even larger clade-level protein sequence set (e.g., the over 14 million protein sequences associated with Fungi, taxon identifier 4751) is a complex issue that is experiment and software dependent. Moreover, in cases where no species-specific protein sequence collection exists, homology-based searching can be used (as described in ref 767). In each of these cases, proteomics practitioners must understand their specific experimental sample and search algorithm in order to know how to best define the search space, which is essential to yielding accurate results.

DATA REPOSITORIES AND KNOWLEDGE BASES

Proteomics Raw Data Repositories

An essential part of the proteomics publication cycle is raw data sharing. This is important so that others can reproduce results and utilize data for new investigations. Computational researchers may use published data to develop new algorithms or combine multiple datasets into a meta study. There are many websites that serve as data repositories for publication. These include: PRIDE,^{782,783} Massive,⁶⁴⁸ and Chorus.⁷⁸⁴

PeptideAtlas and SRMatlas

It would be beneficial to analyze all the data *in toto* to derive a knowledge base of all detectable proteins in an organism. A challenge in attempting this is that, given the vast array of software for MS data analysis, the results are not directly comparable nor combinable given the problem of false discovery rates (FDR) that must be added up when dataset results are combined. For example, if we combine 3 datasets that were each filtered to 1% FDR, the maximum FDR of the combined dataset is now 3% because it is unlikely that the random decoy hits are shared across each dataset. To address this, in 2005 the PeptideAtlas concept was started to ingest as many publically available datasets as possible per organism, search the data through a single pipeline together and arrive at a total controlled 1% protein level FDR.^{785,786} The PeptideAtlas website (www.peptideatlas.org) is a multi-

organism, publicly accessible compendium of peptides identified in large sets of tandem mass spectrometry proteomics experiments. Mass spectrometer output files are collected for human, mouse, yeast, and many other organisms of research interest, and searched using the latest search engines and genome derived protein sequences. All results of sequence and spectral library searching on PeptideAtlas are processed through the Trans Proteomic Pipeline to derive a probability of correct identification for all results in a uniform manner to insure a high quality database, along with false discovery rates at the whole atlas level.

The most recognizable MS data compendium is the Human PeptideAtlas which is produced yearly since 2005 to derive all the peptide sequence knowledge of the current human proteome (Figure 17A). As of 2024, the Human PeptideAtlas contains the knowledge of over 93% of the human proteome, with over 170K MS runs and 4.8B spectra searched resulting in 3.9m peptides identified and 17,416 canonical proteins identified from the 19,600 total proteins possible. The number of proteins has been incrementally increases year over year as new public data becomes available (Figure 17B).

For the presentation of selected reaction monitoring (SRM) targeted peptide assays, there are two components of the PeptideAtlas ecosystem where the PeptideAtlas SRM Experiment Library (PASSEL) is presented to enable submission, dissemination, and reuse of SRM experimental results from analysis of biological samples.^{787,788} The PASSEL system acts as a data repository by allowing researchers with SRM data to deposit their data in parallel with journal publication, and other users can search existing data to obtain the parameters for replication in their own laboratory. Another unique component for SRM data repositories is the SRMatlas website, which provides definitive coordinates for all possible proteins within an organism to conduct targeted SRM assays that conclusively identify the respective peptide in biological samples. As an example, the Human SRMatlas provides data on 166,174 synthetic proteotypic human peptides, providing multiple, independent assays to quantify any human protein and numerous spliced variants, non-synonymous mutations, and post-translational modifications.⁷⁸⁹ The data are freely accessible as a resource at <http://www.srmatlas.org/>.

Other Knowledge Bases

There are many other knowledgebases that will be useful to proteomics researchers. These include the proteomics standards initiative (PSI),⁷⁹⁰ Proteometools,⁷⁹¹ and iProX.⁷⁹² Panorama⁷⁹³ is a resource for sharing processed proteomics data including the extracted ion chromatograms, which can

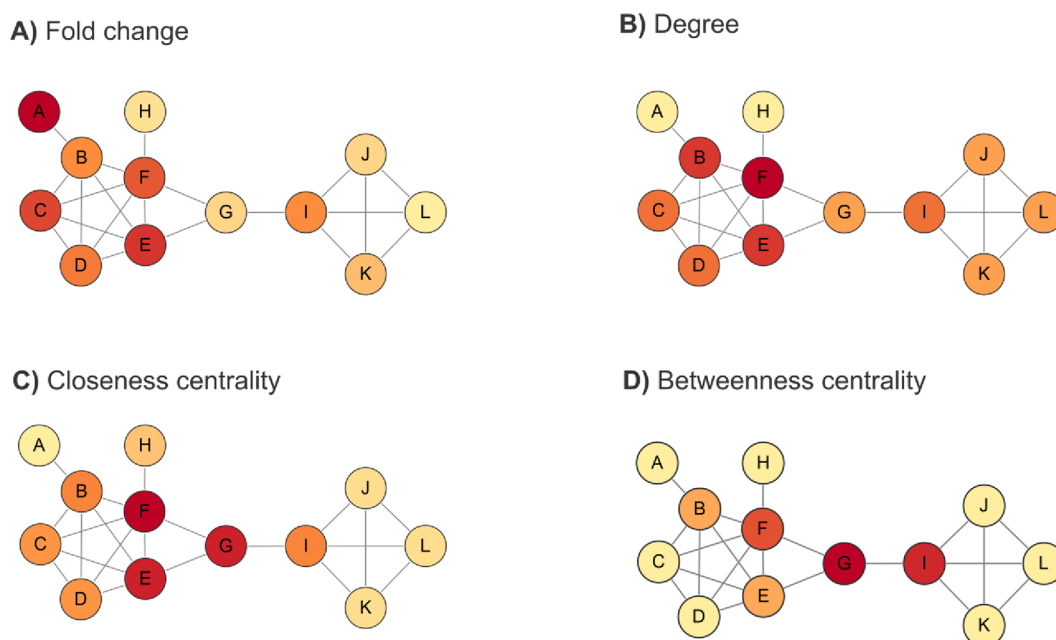


Figure 18. Analysis of a simple network using different centrality measurements. Nodes are colored according to each metric using a yellow-to-red gradient (yellow: lowest value, red: highest value). Network visualization and analysis were performed in Cytoscape.

improve transparency by enabling easy data inspection on the web.

BIOLOGICAL INTERPRETATION

The most common untargeted proteomics experiment will produce a list of proteins or peptides of interest which require further validation and biological interpretation. This list usually results from statistical data analysis; the typical output of differentially expressed proteins usually contains hundreds of hits. In this section, we aim to present a concise overview of how proteomic data can be effectively contextualized and used to generate new hypotheses.

The simplest approach is to start manual lookup of every protein in the list to uncover groups that function together. Starting with a list of hundreds of protein changes, a smaller list can be prioritized by considering the level of significance and effect size. For example, proteins with the smallest p-values (significance) and largest abundance fold-changes (effect size). It is tempting to focus on proteins with the most extreme fold changes. In this case, the assumption is that the more significant the fold change (in either direction, up- or down-regulation), the higher the impact of those proteins on cellular behavior. This assumption is not always valid because protein signal in MS depends on abundance. The manual data interpretation approach is typically infeasible due to the number of proteins that would need to be individually looked up one-by-one.

A better strategy is to use computational methods. These methods may consider the whole list of proteins including some ranking by significance or fold change. One common interpretation method is to construct a protein network, which then lends itself to network analyses. Another method is to consider functional enrichment through annotation databases. These databases offer insights by examining the enrichment of certain functional annotations amongst the interesting proteins. Secondly, one could consider other evolutionary, structurally or regulatory based methods to enable interpreta-

tion of the data. To fully interpret analysis, it may be required to perform or examine other data such as data from biophysical, biochemical and alternative proteomic approaches. Finally, the data can further be interpreted using multi-omic, native or clinical approaches. Below we summarize these approaches and point out potential pitfalls with these methods.

Constructing a Protein Network

A network is a representation of the relations between objects. Nodes are the entities of the network (e.g., users of a social platform, train stations, proteins), while edges are the connections between them (e.g., friendship, routes, and protein interactions, respectively). In the case of protein-protein interactions, evidence for the functional associations between proteins can be obtained experimentally. For example, co-immunoprecipitation, crosslinking, and proximity labeling can be used to reveal physical interactions.⁷⁹⁴ The data is presented in a table with nodes and edges (e.g., “protein A interacts with protein B”) from which the network can be constructed. A considerable wealth of protein-protein association data is stored in free databases like IntAct, which contain interactions derived from literature curation or direct user submissions.⁷⁹⁵ Protein interactions can also be predicted by classifiers that consider many features, like orthology and co-localization, to produce a posterior odds ratio of interaction.^{796,797}

Large repositories like STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) collect and integrate protein-protein interaction data from several databases.⁷⁹⁷ STRING also provides a web-based interface to survey the data, and users only have to feed a search box with the identifiers of the protein(s) of interest. STRING will retrieve the network and show the evidence supporting each interaction. Importantly, these databases do not indicate the direction of the interaction, so they produce undirected networks.

There are many other options for generating and working with networks. For example, geneMANIA can generate a

network from data.⁷⁹⁸ Cytoscape is a free tool useful for generating and interacting with networks.⁷⁹⁹

Network Analysis

Network analysis is a group of techniques that explore and investigate the network, yielding valuable knowledge about its structure and unveiling key players regulating the flow of information. One of the first steps in network analysis relates to centrality measurements. Centralities are indicators of the relative importance of a node corresponding to its position in the network, and each centrality measure provides new insights to interpret the data in new ways.^{800,801}

Degree Centrality. The degree of a node measures the number of edges incident to that node. Nodes with a high degree interact with many other nodes, called first neighbors. In particular, the node degree distribution in protein networks is highly skewed, with most nodes having a low degree and a few having high degrees, known as hubs. Hubs are usually regulatory proteins, being notable examples oncogenes and transcription factors. Moreover, hubs are attractive targets for directed interventions, as their alteration has a profound effect on the stability of the network.⁸⁰²

Closeness Centrality. The route from one node to another is a path, and the shortest path is the one connecting them in the least amount of steps. Closeness centrality is the inverse of the average length of a node's shortest paths to all other nodes in the network. Nodes with a high closeness score have the shortest distances to all the others, so closeness centrality calculations detect nodes that can spread information very efficiently, as they are in a better position in the network for this task.^{803,804}

Betweenness Centrality. This centrality index is related to the amount of shortest paths transversing a node. Nodes with a high betweenness centrality usually bridge different parts of the network and strongly influence the flow of information, as they lie in communication paths. These connector hubs (or bottlenecks) are also interesting for follow-up experiments because their removal can disconnect different regions of the network.⁸⁰⁵

Centrality measurements add new layers of information and allow for ranking differentially expressed proteins apart from their fold-change in abundance. Figure 18 depicts a simple network consisting of proteins A to L, with A having the highest fold change (10) and L the lowest (2). In Panel A, the fill color for the nodes indicates this metric, where it can be easily seen that A stands out. However, protein A is a peripheral protein, only interacting with B. In Panel B, nodes are colored according to node degree. Clearly, protein F has the highest number of interactions and is also the closest to all other nodes, which can be appreciated when nodes are colored according to closeness centrality (Panel C). On the other hand, protein G acts as a bridge between two regions of the network and thus, has the highest betweenness centrality (Panel D). Except for fold change, node A has the lowest indices, and it will be up to the researcher to decide whether this protein warrants further examination.

Network Clustering

In the small network presented in Figure 18, two groups of densely connected nodes exist. This topology suggests that these communities (or "clusters") work together or participate in a protein complex. Dividing a network into clusters helps identify underlying relationships among nodes, which is especially useful in large networks. In a broad sense, network

clustering groups nodes according to a topological property, generally interconnectedness. There are many network clustering algorithms, each with its own merits and approaches.^{806,807} The MCL (Markov CLustering) algorithm is suitable for protein networks in most situations. On the other hand, the Molecular COMplex DETection (MCODE) algorithm helps detect very densely connected nodes, thus unveiling protein complexes.⁸⁰⁸ In this regard, network clustering is useful for tentatively assigning the function of an uncharacterized protein. If the protein appears in a cluster, its function should be closely related to the cluster members, a principle known as "guilty by association."⁸⁰⁹

Network Visualization

A critical step in network analysis is to display the data in a structured and uncluttered graph. Networks can rapidly become a hairball unamenable to interpretation. Software platforms like Cytoscape can be used to visualize networks orderly by applying layout algorithms and format styles.⁸¹⁰ Since many of these platforms are open source, community-designed plugins enhance their capabilities. In Cytoscape, the stringApp adds a search bar to query the STRING database with accession numbers or protein names.⁸¹¹ The network is directly retrieved into Cytoscape, where its built-in network analyzer can be used to calculate centralities. Moreover, user-defined information, like fold-change values, can be integrated and mapped into the network.

Functional Term Enrichment Analysis: KEGG, String, GO, GSEA, ORA, Reactome, and Others

Term enrichment analysis is performed to assess whether particular 'functional terms' are over-represented in a list of proteins (e.g. from a proteomics experiment).^{812–814} For example, after a differential abundance analysis, we may wish to examine whether there is any shared function amongst the proteins which were determined to have significant changes. The simplest analysis to test whether this subset contains more of any particular functional terms than we would expect given the background of proteins. For example, the Gene Ontology is split into the classes: Cellular Component, Molecular Function and Biological Function and we might be interested as to whether our proteins may be more likely to localize to a particular subcellular niche.⁸¹⁵ The Cellular Component terms could give us a starting point if this might be the case, by examining if Cellular Component annotations are enriched.

There are a number of databases and tools to perform such analysis, which can even be extended to examine whole pathways, networks, post-translational modification and literature representation. For example, databases such as KEGG,⁸¹⁶ String,⁷⁹⁷ Reactome,⁸¹⁷ and PhosphoSitePlus⁸¹⁸ can be used to test or annotate a list of proteins. For example, proteomics analysis of human cardiac 3D microtissue exposed to anthracyclines (drugs used in cancer chemotherapy) unearthed several proteins with altered levels.⁸¹⁹ Many of these were specifically grouped under GO terms related to mitochondrial dysfunction, indicating the detrimental effects of these drugs on the organelle. GO terms⁸¹⁵ or descriptors from other annotation libraries (like KEGG⁸¹⁶ or REACTOME⁸²⁰) can be retrieved from STRING when constructing a network or from other freely available compendiums. We refer to a number of articles on the topics, including tools, reviews and best-practice.^{821–823} The main points from such analysis is that we can obtain an insight about protein function by looking at whether our list of proteins have similar or the same

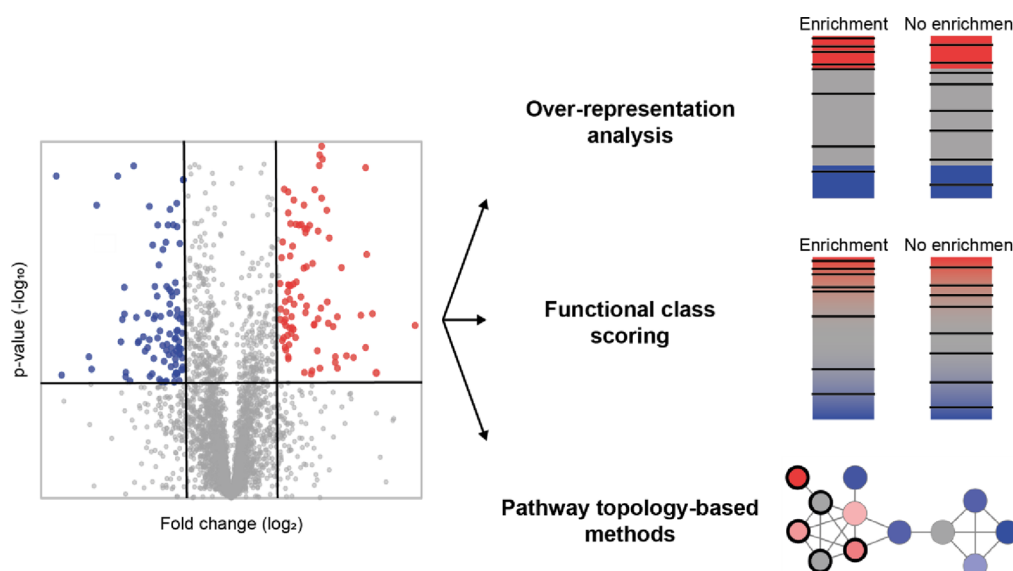


Figure 19. Types of functional enrichment methods. In the volcano plot (left), proteins with altered values are colored blue or red according to arbitrarily chosen cut-off values for significance and fold change. Black bars or thick-bordered nodes indicate members of a GO category.

annotations. A number of limitations should be taken into account for interpretation. The first is that proteins that are more abundant are more likely to be studied, measured and examined in the literature. Hence, abundant proteins will have more annotations than less abundant ones. One key part of the analysis is also to correctly select the background set; that is, the universe of protein which our list is being compared against. By including contaminants or proteins that are not expressed in our system within the list, the results may be unfaithful.

We may also have access to our own curated set of annotations derived either computationally or experimentally. One may be interested in seeing whether we have enrichment of these annotations amongst the differentially abundant proteins. Our list of proteins could be divided into two groups: differentially abundant or not. These groups could be divided into whether they have a particular annotation: yes or no. This information can be summarized in a two-by-two table, to which we can apply a statistical test to examine whether that annotation is enriched within our differentially abundant proteins. One test that could be used is the hypergeometric test, and another would be a Fisher Exact test.

There are many methods for performing functional enrichment analysis on the data, but they can mainly be classified into three categories (Figure 19), as follows.

Over-representation Analysis. In modern proteomics analysis, usually thousands of proteins are identified and quantified. Fold-change and significance thresholds are chosen (e.g., fold-change ≥ 2 and $p \leq 0.05$) to obtain a list of proteins with altered levels among the tested conditions. In over-representation methods, a contingency table is created for every protein set to establish whether proteins with altered abundance show an enrichment or a depletion of the ontology term compared to the background observed proteome.⁸²⁴ For example, suppose that 2000 proteins were quantified in a proteomics analysis, being 40 of these members of the set “tricarboxylic acid cycle (TCA).” Also, let us assume that 200 proteins showed altered abundance, with 15 belonging to the TCA set. Then, the contingency table can be constructed as follows (Table 11):

Table 11. Example Term Enrichment Analysis

	Proteins with altered abundance	Proteins with unaltered abundance	Total
Proteins in TCA set	15	25	40
Proteins not in TCA set	185	1775	1960
Total	200	1800	2000

Then, a suitable statistical test is conducted to ascertain if proteins with altered levels are enriched in members of the TCA cycle (in this case, they are; $p < 0.00001$). This is commonly achieved using Fisher’s exact test.⁸²⁵ The process is then repeated for every set as desired. Since multiple comparisons are tested, p -values must be adjusted by a false discovery rate.⁸²⁶ There are also several free tools for term enrichment analysis, including Enrichr,⁸²⁷ GSEA,⁸²⁸ and DAVID.⁸²⁹

Functional Class Scoring. The caveat of over-representation methods is that they rely on a list of differentially expressed genes or proteins with altered abundance, selected due to arbitrarily chosen cut-off values. For example, if we set a fold change cutoff of 2, a protein with a fold-change of 1.99 would not be included in the analysis. Moreover, several proteins belonging to the same set may have altered levels but are below the fold change threshold. However, moderate alterations of their abundance as a group could drive the observed phenotype, even more so than a single protein over the cutoff. Functional class scoring strategies aim at countering these limitations by disregarding thresholds altogether. GSEA (Gene Set Enrichment Analysis) is a widely used functional class scoring method in which all detected entities are first ranked according to a quantitative measurement (fold change, p -value, or their combination).⁸³⁰ Then, the distribution of members of a set is obtained. A scoring scheme based on the Kolmogorov–Smirnov test is used to assess whether there is an enrichment of the category towards the top or bottom of the ranked list.

Pathway Topology-Based Methods. Both methods mentioned above do not consider the functional relationships

among proteins put forth by network analysis; i.e., they assume functional independence. Topology-based enrichment methods incorporate this information by, for example, assigning an importance value to a set when its members also participate in a pathway or cluster together in a network.⁸³¹ Figure 19 shows how topology-based methods consider non-significant hits (grey nodes) that other strategies may not pick up, due to their position in a network.

Other Computational Approaches: Network Analysis, Isoform Correlation Analysis, AlphaFold, BLAST, Protein Language Models

Additional computational analysis of a list of interesting proteins may uncover additional substructure, correlation or biologically useful hypothesis. Building a network between the proteins based on the experiments performed might be a useful approach to identify additional structure. For example, co-expression network analysis can be used to build a network from these proteins.⁸³² In these networks, proteins are nodes and edges describe relationships between those proteins. Network-specific methods can then be applied, such as community detection algorithms which could uncover clusters of proteins with shared functions.^{833,834}

One way the proteome generates complexity is through alternative-splicing, which results in protein isoforms.⁸³⁵

Recently, a number of tools have been proposed to identify peptide isoforms that are quantitatively different across conditions by using a principle called peptide correlation analysis.^{836,837} The idea is that the quantitative behavior of peptides should match each other. If there are subgroups that behave coherently within the group but not across groups suggest that peptide may have come from a different proteoform. These approaches can be used to identify specific proteoforms that are functional across different conditions.

For many, a protein's structure reveals important functional details.⁸³⁸ There are a plethora of approaches to predict a protein's structure.^{839–841} Recently, AlphaFold and RoseTTA-Fold have become dominant methods for predicting protein structures with high resolution.^{840,841} If intrinsically disordered domains are of particular interest, methods explicitly designed for this task are recommended.⁸⁴² Once a structure is obtained more elaborate computational methods might be useful such as docking or molecular dynamics.^{843,844} These approaches can give insight into how protein or molecules fit together and the dynamics of a protein's structure (conformational heterogeneity). A complete discussion of these topics is beyond the scope of this section.

Another way to obtain insights into a protein function is to look for protein with similar sequences or motifs. Using BLAST, a sequence alignment tool, one can align two or more protein sequences and determine their level of similarity.⁸⁴⁵ For example, if a human protein of unknown function has a similar sequence to a yeast protein with known function this may be a starting place for the putative function of that protein.

Novel approaches to representing the similarity of proteins have proved successful at predicting the functional properties of proteins. Protein language models seek to learn "representation" of proteins, these are usually numerical vectors that represent a protein sequence.^{846,847} Abstractly, these vectors preserve protein similarity or a notion of "proteinness". This usually means that two proteins that have a close vector may share similarities in protein function. These representations are also advantageous because they can easily

become the inputs for machine learning algorithms to predict valuable protein properties; for example, thermal stability values,⁸⁴⁸ protein–protein binding affinities,⁸⁴⁹ secondary protein structure, and more.

■ ORTHOGONAL VALIDATION

The Importance of Orthogonal Experimental Validation

The computational workflows to interpret mass spectrometry data are sophisticated, powerful tools, but also show important limitations and caveats due to their dependence on limited prior knowledge, specific experimental parameters or data quality restraints (see section "Raw Data Analysis"). These inherent biases can give rise to ambiguous or spurious interpretation of the data even when these workflows are applied correctly and to the best of the experimenter's knowledge. Therefore, researchers will oftentimes be asked by scientific journals to provide independent orthogonal validation of their proteomics data and not performing such can be a major roadblock in the publication process.

The aim of validating data obtained by proteomics approaches should always be two-fold by demonstrating that the conclusions arrived at by proteomics data acquisition and analysis are, firstly, valid and, secondly, relevant. Depending on the question at hand, researchers can draw on an overabundance of techniques to validate MS-derived hypotheses in appropriate cellular, organismal or *in vitro* models. In the following paragraphs we aim to present only a high-level, stringent, non-exhaustive selection of orthogonal validation approaches and emphasize the importance of implementing assays that challenge assumptions gained from proteomics data analysis pipelines.

Before embarking on orthogonal validation of any hit, the success of the experiment should be established by assessing (internal) positive controls. Internal positive controls can be proteins whose behavior under the experimental conditions applied can be deduced from prior knowledge (i.e. the scientific literature or public databases). Once the expected changes in internal controls have been confirmed by computational analysis (see the above section), the orthogonal experimental validation of novel, perhaps unexpected findings can begin.

Orthogonal validation of new insights obtained from quantitative proteomics experiments can be a very time-consuming process and often requires familiarity with techniques not directly related to proteomics workflows. Given these challenges, the method(s) of choice warrant(s) careful consideration and is highly context-dependent. Importantly, proteomics experiments in one way or another generally yield comprehensive lists of potentially interesting candidate proteins or pathways, the researcher will have to shortlist candidates to be taken forward to the validation stage of the project. Which candidates should you validate by an orthogonal approach and which ones might not require further validation?

In general, candidates representing abundant proteins that show high sequence coverage and are detected with high confidence might not necessarily need extensive orthogonal validation when compared with proteins of intermediate to low abundance that might be more challenging to faithfully quantify by proteomics alone, i.e., many membrane proteins or transcription factors. Similarly, since the proteome is rarely comprehensively quantified in any single proteomics experi-

ment, proteins of interest (POIs) that are critical for an observed biological change might not be part of the dataset. In these cases, additional, targeted analyses might help to support or discredit proteomics-based hypotheses.

Validation techniques are as manifold as biological questions and discussions thereof may easily fill multiple textbooks. The following sections are therefore merely meant to paint with a broad brush stroke a picture of useful methodologies with which to validate and follow up MS-data derived observations. As this is meant to orient the reader, wherever possible, we will explicitly point out useful literature reviews for a deeper dive into each of these techniques.

General Considerations

Once POIs have been selected based on prior agreed-upon selection criteria (i.e. (adjusted) p value and/or fold change thresholds), orthogonal validation experiments should ideally be conducted under physiologically relevant conditions to mitigate artificial and misleading outcomes. Therefore, *in vitro* experiments, while useful to isolate and dissect particular aspects of a biological system, can give highly artificial results as conditions are far removed from the POI's native environment. To investigate the biological function of a protein or pathway, direct genetic manipulation of the biological system at hand (e.g., modulating the expression of a POI by overexpression or knockout-/down experiments) can be minimally invasive when performed correctly. Should the POI be encoded by an essential gene, by definition, a complete and stable knockout might not be advisable.^{850,851} In these extreme cases, attenuated expression (i.e., using RNA interference (RNAi) or controlled degradation, see below) rather than complete repression of a gene can be used to probe for protein function. Epitope tagging and/or exogenous expression of a gene of interest can be a powerful approach in assessing PPIs and investigating proteins of low abundance. However, overexpression artifacts are common.⁸⁵²

It is not always possible to fully avoid the pleiotropic effects of protein (over-)expression or depletion, but a number of mitigation strategies (i.e., inducible expression, the use of multiple independent RNAi strategies) will be discussed below.

Extensive biochemical characterization of any overexpressed gene is critical to ensure it closely reflects the functions of its endogenous counterpart. These assays might involve assessing protein localization (i.e., by imaging techniques such as microscopy and flow cytometry), protein abundance (i.e., by mass spectrometry or immunoblot analysis) and phenotypic assays where applicable and practical.

Functional Genomics Techniques in the Validation of MS Hits

Typical follow-up experiments to validate mass-spectrometry derived insights often involve the acute depletion or induction of a POI and assessing the impact on specific cellular phenotypes. Here we present a selection of methodologies to effectively modulate gene expression and discuss important considerations when planning functional genomics experiments for target validation.

Gene deletion or knockdown to prevent production of a functional protein is a powerful means to interrogate the role of one or more proteins in the phenotype(s) under investigation. To this end, well-established technologies deserving mention at this point are RNA interference (RNAi) in the form of siRNA/shRNA- or miRNA-mediated gene knockdown by CRISPR/Cas9- or TALEN-mediated gene

knockout.⁸⁵³ Since each one of these technologies comes with its own unique advantages and caveats, the approach taken depends on the biological question at hand.

Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas-based gene deletion technologies allow for the targeting of individual genes with relative ease, high efficiency and specificity.⁸⁵⁴ When expressed in mammalian cells, the bacterially-derived Cas9 endonuclease can be guided with the help of a short guide RNA (gRNA) to a genomic location of interest, where it creates a DNA double strand break in a highly controlled manner (for a detailed discussion see ref 855). The cell's DNA double-strand break repair machinery then introduces base pair insertions or deletions (indels) via non-homologous-end-joining (NHEJ), thus causing missense, and frameshift mutations (i.e. resulting in premature stop codons), leading to premature termination of gene expression or non-functional, aberrant gene products. Similarly, the concomitant provision of a complementary DNA donor template encoding a desired gene modification (i.e. insertion of a stretch of DNA or base pair modification) will trigger homology-directed repair (HDR), resulting in gene knock in or base editing.⁸⁵⁵ Practical considerations of CRISPR/Cas9-mediated gene knock-in and base editing will not be addressed in detail but are expertly discussed in refs 856–859.

The relative ease-of-use and high efficiency of the CRISPR/Cas9 gene editing technology has rendered it the method of choice for gene manipulation in many fields of cell biology. However, it should be noted that CRISPR/Cas9-mediated gene deletion is not free from off-target effects (ref 860 for advice on how to minimize these off-target effects). Moreover, long-term depletion (or upregulation) of a POI itself can in some cases have dramatic systemic consequences and constitute an acute selection pressure leading to compensatory stress-induced adaptation that might obfuscate primary loss-of-function phenotypes and pose a substantial hurdle to the interpretability of biological data. As these compensatory mechanisms often manifest with time, controlled, transient genetic manipulation (gene depletion or transgene expression) is advised. Small interfering RNA (siRNA)-mediated knock-down by transient transfection is typically achieved at shorter time frames (24–96 h), depending on the turnover of the POI. On an even shorter timescale, targeted, degron-based degradation systems enable depletion of a POI within minutes and further reduce off-target effects, but require the exogenous expression of a transgene and therefore some genetic manipulation. A more comprehensive discussion of a selection of these systems (anchor-away, deGradFP, auxin-inducible degron (AID), dTAG technologies) and their advantages and potential pitfalls is presented in ref 861.

Multiple eukaryotic and prokaryotic transcription-based systems have been developed that allow for the controlled biosynthesis or depletion of one or more POIs. Amongst these, a popular and dependable choice for mammalian cells are tetracycline-controlled operon systems, which allow up- or downregulation of a POI in the presence of the antibiotic tetracycline or its derivative doxycycline. These systems rely on the insertion of a bacteria-derived Tet operon (TetO) between the promoter and coding sequence of the gene of interest. In this configuration, the TetO binds a co-expressed Tet-repressor protein blocking transcription of the gene of interest. When tetracycline is added to the cells, the repressor then dissociates from the operon, thus de-repressing the gene of interest. Different variations of this potent system exist,

allowing for more flexibility in experimental design. For instance, in the Tet-OFF system, the Tet repressor is fused to a eukaryotic transactivator (the chimeric fusion construct is termed tTA) and addition of tetracycline, or the related doxycycline, abolishes TetO binding and thus suppresses transcriptional activation.⁸⁶² Alternatively, a mutant form of tTA (rtTA) binds the TetO only in the presence of tetracycline, allowing for tetracycline-induced gene expression. For a detailed discussion of these systems, we refer the reader to an excellent review.⁸⁶³

When generating stable expression cell lines, being able to precisely control the genomic integration site of the transgene reduces overall genetic heterogeneity in a cell population and thereby reduces potential off-target or pleiotropic effects. This ability is realised in the FlpIn-T-REx technology which harnesses Flp-recombinase mediated DNA recombination at a strictly defined genomic locus (the FRT site).⁸⁶⁴ Site-directed isogenic integration of any gene of interest at the FRT site, which is under a tetracycline-inducible promoter and a hygromycin resistance gene, allows for facile generation of tetracycline/doxycycline-inducible isogenic expression cell lines with minimal leaky expression (for an example, see ref 865).

Validation and Interpretation of Protein Abundance Changes

To validate protein abundance changes observed by quantitative bottom-up proteomics or simply assess the success of targeted genetic manipulation as part of an orthogonal follow-up experiment (see above), the experimenter typically resorts to antibody-based techniques such as immunoblotting analysis or immunofluorescence and immunohistological imaging of POIs. The latter also allows for validation of protein expression and localization in intact tissue or cells. However, these semi-quantitative methods are strongly influenced by the quality of the antibodies used and might not be sensitive enough to detect small changes in protein levels. In this case, more accurate orthogonal quantitation of proteins might be achieved by stable isotope labelling (SILAC/TMT/iTRAQ) and/or SRM/PRM (see section “Types of Experiments”). SDS-PAGE and immunoblot analysis are powerful and facile low-throughput tools to quickly validate protein abundance changes. However, short of introducing epitope tags to the endogenous POI, the success of immunoblotting is contingent on the availability of specific antibodies, which can present a formidable problem when investigating poorly characterized proteins or working with model organisms for which the commercial availability of specific antibodies is limited (this is particularly problematic for ‘unconventional’ or even well-established model organisms such as yeast). A detailed discussion of the strengths and pitfalls of immunoblotting for validation of semi-quantitative proteomics data can be found in an excellent review by Handler et al.⁸⁶⁶

Protein abundance changes detected in a proteomics experiment can be the result of a range of different cellular processes. The abundance of a protein in a complex sample (e.g., cell lysate or biological fluid) directly reflects a combination of the protein’s intrinsic stability and the translational rate under the conditions of interest.

Both protein stability as well as gene expression activity can be quantified independently. Altered protein stability might be a direct consequence of specific or global changes in protein

turnover. Radioisotope labelling is a well-established, accurate way to monitor protein synthesis, maturation and turnover.^{867,868} This “pulse-chase” methodology relies on the incorporation (“pulsing”) of radioisotopes (typically ³⁵S-labelled cysteine and methionine) into de-novo synthesized proteins. Upon withdrawal of the labeled amino acids from the culture medium, the decay of signal is monitored over time (‘the chase’) by SDS-PAGE and phosphoimaging, resulting in a temporal readout of protein abundances. The advantage of this technology is that a subpopulation (newly synthesized proteins) can be monitored directly, giving an accurate assessment of protein stability. Once a change in protein stability has been validated, the underlying mechanisms can be addressed by inhibiting protein degradation pathways; prominently proteasome-mediated degradation (using specific proteasome inhibitors such as bortezomib/velcade or MG132), autophagy (pharmacologically inhibiting autophagic flux) or degradation by proteases (using protease inhibitors). The type of radiolabeling described above is relatively labor-intensive, of low-throughput and has the obvious disadvantage of requiring radioactive material, which needs to be handled under strict safety precautions. Moreover, it critically depends on the presence of one or more methionines and/or cysteines in the POIs.

It is also possible to measure protein stability within complex protein mixtures (i.e. cell lysates or biological fluids) using an array of specialized mass spectrometry techniques as discussed in refs 869 and 116.

For purified proteins, well-established *in vitro* spectrometric and calorimetric methods such as circular dichroism, differential scanning calorimetry or differential scanning fluorimetry can be used, but the relatively high sample amounts might be restrictive.

Finally, gene expression changes can also be determined with high fidelity using quantitative real-time PCR (qRT-PCR) or RNA-Seq can measure changes in gene transcription or mRNA turnover (for an extensive discussion of both technologies, please see refs 870 and 871, respectively).

Validation of Protein–Protein Interactions

The interaction of a protein with other proteins determines its function. Protein-protein interactions (PPIs) can be either mostly static (i.e. core subunits of a protein complex) or dynamic, varying with cellular state (i.e. cell cycle phase or cellular stress responses, posttranslational modifications) or environmental factors (i.e., availability of nutrients, presence of extracellular ligands of cell-surface receptors). Therefore, any given protein can typically bind a range of interaction partners in a spatially and temporally restricted manner, thus forming complex PPI networks (the interactome of a protein). The method of choice to experimentally examine altered PPI states depends on the model system and biological question (i.e., purified proteins vs complex protein mixtures, monitoring of PPIs in live cells or cell lysate etc). Popular methods for the validation of PPIs *in vivo* include protein fragment complementation (split protein systems), 2-hybrid assays (mammalian, yeast and bacterial), proximity ligation, proximity labelling and FRET / BRET.

Protein fragment complementation assays rely on the principle that the two self-associating halves of reporter proteins can be expressed in an inactive form but when in spatial proximity bind one another to complement the functional, active reporter. When these split reporters are

fused to two interacting proteins (so-called bait and prey proteins), the binding of bait to prey induces the spatial restriction needed to fully complement the reporter. Commonly used reporter complementation systems are split fluorescent proteins (i.e., GFP, YFP),⁸⁷² ubiquitin,⁸⁷³ luciferase,⁸⁷⁴ TEV protease,⁸⁷⁵ beta-lactamase,⁸⁷⁶ beta-galactosidase, Gal4, or DHFR.⁸⁷⁷ The resulting functional readout of these complementation system depends on which split reporter is used. In general, the split luciferase system shows enhanced sensitivity over fluorescence-based systems as background luminescence is low.

Two-hybrid assays are based on a similar functional complementation strategy as fragment complementation systems. Conventionally, two self-complementing transcription factor fragments are fused to bait and prey proteins, respectively, leading to the restoration of a functional transcription factor only upon prey-bait interaction. The complemented transcription factor then induces the expression of a reporter gene that can be measured. Multiple variations of this system abound for different model organisms, but they almost always involve transcriptional activation or repression of a reporter gene (ref 878 for a detailed discussion).

The yeast-2-hybrid system (Y2H) is deserving of mention here as it had been the very first 2-hybrid system established⁸⁷⁹ and has ever since proven to be extremely versatile (multiple auxotrophic reporters and markers of phenotypic sensitivity available), cheap, lends itself to functional high-throughput screening and variants have been developed that allow for the investigation of membrane-protein interactions (i.e. membrane Y2H).^{878,880}

Despite the many advantages the Y2H offers, critical drawbacks include the potential of misfolding of bait and prey proteins when fused to a complementation reporter, expression at non-physiological levels, the lack of control over posttranslational modifications that might be important for the PPI under investigation, and the potential requirement of kingdom- or species-specific folding factors for the bait/prey under investigation (i.e. when probing PPI of mammalian proteins in Y2H). Principles of the Y2H technology have also been adapted to mammalian systems, which circumvent some of the aforementioned drawbacks of Y2H.⁸⁸¹

Perhaps the most commonly applied method of detecting and validating PPIs *in vitro* is affinity purification (AP, also known as affinity chromatography) of co-immunoprecipitation (Co-IP) either coupled with SDS-PAGE/immunoblotting or mass spectrometry to determine the identity of interacting proteins. AP typically relies on the isolation of a transgenic POI by an epitope tag (using epitope-specific matrix-conjugated proteins (antibodies or epitope-binding proteins)), while Co-IP harnesses specific antibodies directly targeting the POI. Specific interactors are expected to be enriched compared to the negative control (i.e. an isotype control antibody, a knockout cell line or empty matrix). AP is not solely restricted to detecting PPIs, but can also be adapted to protein interactions with other biomolecules such as RNA.⁸⁸² It should be noted that AP and Co-IP can return multiple potential binding partners, many of which might be artefactual due to loss of cellular compartmentalization during sample preparation.

To reduce the probability of such artefacts and increase the confidence of a specific interaction, reciprocal affinity purification (by pulldown of each interaction partner) or *in situ* imaging might be performed (i.e. using fluorescence

resonance energy transfer (FRET),⁸⁸³ split-protein systems,⁸⁸⁴ proximity ligation assay,⁸⁸⁵ and immunofluorescence microscopy).

Förster and bioluminescence resonance energy transfer (FRET/BRET) can be used for *in situ* visualization of protein proximities and therefore PPIs. In FRET, non-radiative energy transfer between donor and receptor chromophores (each fused to prey and bait proteins, respectively), results in the emission of a characteristic fluorescence signal only when both prey and bait are in very close proximity (1–10 nm distance) and a suitable light source for donor excitation is provided.⁸⁸⁶

The underlying principle of BRET is similar to that of FRET but with the exception of using a chemical substrate which activates bioluminescent donor, such as luciferase, resulting in energy transfer to a fluorescent acceptor molecule.^{887,888} The main advantages of BRET over FRET are independence from an external light source (which can result in photobleaching), but requires at least one of the POIs to be fused to the donor (while in FRET, donor and acceptor can be chemically conjugated to POI-specific antibodies).⁸⁸⁷ FRET can be particularly useful in investigating cell surface protein interactions when using specific antibodies conjugated to donor and acceptor probes as antibodies are not cell-permeable and therefore restricted to targets presented on the cell surface in the absence of membrane permeabilization agents. Other fluorescence-based PPI assays encompass Fluorescence correlation spectroscopy (FCS) and fluorescence cross-correlation spectroscopy (FCCS). These methods use small volumes of fluorescently labelled proteins and can determine their diffusion coefficients, which change in when proteins form a complex.⁸⁸⁹

Proximity labelling methods (Proximity ligation and enzymatic proximity labelling (BirA, APEX2, HRP) can surveil labile or transient interaction in live cells in a high-throughput format when coupled with target identification by MS.^{890,891} These approaches harness a biotin ligase (i.e. BirA, BioID2, AirID, BASU, APEX2, HRP) fused to a POI whose interactome is to be determined. In the presence of biotin (for BirA, BioID2, AirID, BASU, APEX2 and HRP) or a biotin-phenol derivative (for APEX2), the biotin ligase will activate the biotin(-phenol) which then covalently biotinylates any protein in close proximity. The activated biotin has a short half-life, ensuring that the effective labelling radius is typically restricted to approximately 10 nm. Biotinylated proteins are isolated by affinity purification with streptavidin-conjugated beads and identified by mass spectrometry or SDS-PAGE/immunoblotting. TurboID, miniTurboID and ultraID, promiscuous biotin ligases faster than BirA, have been developed allowing for shorter treatment times and decreased background signal. The choice of a biotin ligase variant depends on the POI and experimental setup, but in general HRP does not work in cytoplasmic environments where conditions are chemically reducing, but is suitable for labelling proteins extracellular face of the plasma membrane or in the endoplasmic reticulum and golgi apparatus. While TurboID and similar variants have fast kinetics, they can cause depletion of endogenous biotin and therefore cytotoxicity.

A major drawback shared by all variants described above is that they necessitate fusion to the POI, which might alter its physiological behavior and give rise to false positives or false negatives. Moreover, detecting a biotin-labelled protein does not unequivocally designate it as an interaction partner as spatial proximity to the POI-biotin ligase fusion protein

without direct binding can result in biotinylation. The inclusion of controls, such as expression of the biotinylating enzyme alone in the cellular compartment of interest, is therefore particularly important for enzymatic proximity labelling methods.

The *in situ* proximity ligation assay (PLA) combines the specificity of antibodies with the signal amplification capacity of a DNA polymerase reaction. Here, two antibodies, each conjugated to a short single-strand DNA (ssDNA) tag and each specific to one of the two proteins whose interaction is under investigation, are added to fixed cells or tissue. Once bound to their respective targets and only when in direct proximity, the addition of two connector oligonucleotides complementary to each tag ssDNA tag and phi29 DNA polymerase, triggers isothermal rolling circle amplification, eventually resulting in the generation of continuous stretches of repetitive DNA. These DNA products can then be visualized by *in situ* hybridization with fluorescently labelled oligonucleotides (see ref 892 for a detailed discussion). PLA has the advantage of visualizing the two interacting proteins in their native environment when high-resolution microscopy is used as a readout.

Chemical cross-linking (XL) of proteins can determine PPIs with amino-acid level resolution, and can thereby give valuable insights into the orientation of two or more proteins relative to one another.⁸⁹³ Recent technical advances also enabled the visualization of protein-RNA interaction.⁸⁹⁴ Various XL chemistries are available (amine-reactive, sulfhydryl and photoreactive crosslinkers; reversible vs irreversible) and cross-linked proteins detected by mass spectrometry.⁸⁹⁵ In general, applying XL-MS to a mixture of interacting, purified proteins is preferable to *in situ* XL of complex protein mixtures (i.e., cell lysate) as detection and deconvolution of XL peptides is technically and computationally challenging.

Surface plasmon resonance can accurately measure several key kinetics of PPIs with high accuracy (e.g. association and dissociation kinetics, stoichiometry, affinity).⁸⁹⁶ It relies on the quantification of refractive index changes of polarized light shone onto a sensor chip containing a prey protein immobilized on a metal surface (typically gold). When prey and bait proteins interact, the mass concentration at the metal interface changes, altering the refractive index and SPR angle (intensity of the refracted light).

AUTHOR INFORMATION

Corresponding Author

Jesse G. Meyer – Department of Computational Biomedicine, Cedars Sinai Medical Center, Los Angeles, California 90048, United States; Smidt Heart Institute and Advanced Clinical Biosystems Research Institute, Cedars Sinai Medical Center, Los Angeles, California 90048, United States; orcid.org/0000-0003-2753-3926; Email: jesse.meyer@cshs.org

Authors

Yuming Jiang – Department of Computational Biomedicine, Cedars Sinai Medical Center, Los Angeles, California 90048, United States; Smidt Heart Institute and Advanced Clinical Biosystems Research Institute, Cedars Sinai Medical Center, Los Angeles, California 90048, United States

†Devasahayam Arokia Balaya Rex – Center for Systems Biology and Molecular Medicine, Yenepoya Research Centre, Yenepoya (Deemed to be University), Mangalore 575018, India

†Dina Schuster – Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich 8093, Switzerland; Department of Biology, Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich 8093, Switzerland; Laboratory of Biomolecular Research, Division of Biology and Chemistry, Paul Scherrer Institute, Villigen 5232, Switzerland; orcid.org/0000-0001-6611-8237

***Benjamin A. Neely** – Chemical Sciences Division, National Institute of Standards and Technology, NIST, Charleston, South Carolina 29412, United States; orcid.org/0000-0001-6120-7695

***Germán L. Rosano** – Mass Spectrometry Unit, Institute of Molecular and Cellular Biology of Rosario, Rosario 2000, Argentina

***Norbert Volkmar** – Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich 8093, Switzerland

Amanda Momenzadeh – Department of Computational Biomedicine, Cedars Sinai Medical Center, Los Angeles, California 90048, United States; Smidt Heart Institute and Advanced Clinical Biosystems Research Institute, Cedars Sinai Medical Center, Los Angeles, California 90048, United States; orcid.org/0000-0002-8614-0690

Trenton M. Peters-Clarke – Department of Pharmaceutical Chemistry, University of California—San Francisco, San Francisco, California 94158, United States; orcid.org/0000-0002-9153-2525

Susan B. Egbert – Department of Chemistry, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

Simion Kreimer – Smidt Heart Institute and Advanced Clinical Biosystems Research Institute, Cedars Sinai Medical Center, Los Angeles, California 90048, United States; orcid.org/0000-0001-6627-3771

Emma H. Doud – Center for Proteome Analysis, Indiana University School of Medicine, Indianapolis, Indiana 46202-3082, United States

Oliver M. Crook – Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford OX1 3LB, United Kingdom; orcid.org/0000-0001-5669-8506

Amit Kumar Yadav – Translational Health Science and Technology Institute, Faridabad, Haryana 121001, India; orcid.org/0000-0002-9445-8156

Muralidharan Vanuopadath – School of Biotechnology, Amrita Vishwa Vidyapeetham, Kollam 690 525 Kerala, India

Adrian D. Hegeman – Departments of Horticultural Science and Plant and Microbial Biology, University of Minnesota, Twin Cities, Minnesota 55108, United States; orcid.org/0000-0003-1008-6066

Martín L. Mayta – School of Medicine and Health Sciences, Center for Health Sciences Research, Universidad Adventista del Plata, Libertador San Martín 3103, Argentina; Molecular Biology Department, School of Pharmacy and Biochemistry, Universidad Nacional de Rosario, Rosario 2000, Argentina

Anna G. Duboff – Department of Chemistry, University of Washington, Seattle, Washington 98195, United States; orcid.org/0009-0002-7316-3831

Nicholas M. Riley – Department of Chemistry, University of Washington, Seattle, Washington 98195, United States; orcid.org/0000-0002-1536-2966

Robert L. Moritz – Institute for Systems biology, Seattle, Washington 98109, United States; orcid.org/0000-0002-3216-9447

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acsmeasuresci.3c00068>

Author Contributions

J.G.M. was assigned last author as the project initiator and leader. R.L.M. was assigned second to last author based on their leadership role in curating all sections. All other authors were ordered by estimating their contributions using a quantitative score. The score for each author was a sum of the number of sentences added plus 33 lines for each figure added. Scores were adjusted for confounding factors, such as split contributions between multiple contributors. Authors with similar scores were assigned equal contributions. CRediT: **Yuming Jiang** visualization, writing-original draft; **Devasahayam Arokia Balaya Rex** visualization, writing-original draft; **Dina Schuster** visualization, writing-original draft, writing-review & editing; **Benjamin A. Neely** writing-original draft, writing-review & editing; **German L Rosano** writing-original draft, writing-review & editing; **Norbert Volkmar** writing-original draft, writing-review & editing; **Amanda Momenzadeh** writing-original draft, writing-review & editing; **Trenton M. Peters-Clarke** visualization, writing-original draft, writing-review & editing; **Susan B Egbert** writing-original draft; **Simion Kreimer** writing-original draft; **Emma H Doud** writing-original draft; **Oliver M. Crook** writing-original draft; **Amit Kumar Yadav** writing-original draft, writing-review & editing; **Muralidharan Vanuopadath** visualization, writing-original draft; **Adrian D. Hegeman** writing-review & editing; **Martín L Mayta** writing-original draft, writing-review & editing; **Anna G Duboff** visualization, writing-original draft, writing-review & editing; **Nicholas M Riley** visualization, writing-original draft, writing-review & editing; **Robert L. Moritz** visualization, writing-original draft, writing-review & editing; **Jesse G. Meyer** conceptualization, project administration, supervision, visualization, writing-original draft, writing-review & editing.

Funding

United States National Institute of Standards of Technology; Argentina Agencia I+D+i (grant PICT 2019-02971); Department of Biotechnology, India (BT/PR16456/BID/7/624/2016, Translational Research Program (TRP) at THSTI); Department of Health Research, Indian Council of Medical Research, Government of India (File No.R.12014/31/2022-HR); Nation Science Foundation (grants MCB-2225057, IOS-2025297, IIBR-1920268); National Institutes of Health (grants R00GM147304, R01GM087221, R24GM127667, U19AG023122, S10OD026936, R21AG074234, R35GM142502).

Notes

The authors declare no competing financial interest.

[†]D.A.B.R and D.S. contributed equally.

[‡]B.A.N., G.L.R., and N.V. contributed equally.

ACKNOWLEDGMENTS

The authors thank Phil Wilmarth for helpful input. Identification of certain commercial equipment, instruments, software, or materials does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose. The authors thank Dasom Hwang for help with graphic design. The authors

thank Anthony Gitter and Daniem Himmelstein for assistance using manubot. The authors thank Jordan Burton and Pierre-Alexander Mücke for minor edits to the text. This manuscript was written collaboratively using manubot.⁸⁹⁷ The live and evolving version where anyone can contribute can be found here <https://github.com/jessegmeyerlab/proteomics-tutorial/tree/main>.

REFERENCES

- (1) Martin-Banlindres, P.; Lan, W.-H.; Board, S.; Romero-Ruiz, M.; Garcia-Manyes, S.; Qing, Y.; Bayley, H. Enzyme-Less Nanopore Detection of Post-Translational Modifications Within Long Polypeptides. *Nat. Nanotechnol.* **2023**, *18* (11), 1335–1340.
- (2) Wang, X.; Thomas, T.-M.; Ren, R.; Zhou, Y.; Zhang, P.; Li, J.; Cai, S.; Liu, K.; Ivanov, A. P.; Herrmann, A.; Edel, J. B. Nanopore Detection Using Supercharged Polypeptide Molecular Carriers. *J. Am. Chem. Soc.* **2023**, *145* (11), 6371–6382.
- (3) Yusko, E. C.; Bruhn, B. R.; Eggenberger, O. M.; Houghtaling, J.; Rollings, R. C.; Walsh, N. C.; Nandivada, S.; Pindrus, M.; Hall, A. R.; Sept, D.; Li, J.; Kalonia, D. S.; Mayer, M. Real-Time Shape Approximation and Fingerprinting of Single Proteins Using a Nanopore. *Nature Nanotech* **2017**, *12* (4), 360–367.
- (4) Afshar Bakshloo, M.; Kasianowicz, J. J.; Pastoriza-Gallego, M.; Mathé, J.; Daniel, R.; Piguet, F.; Oukhaled, A. Nanopore-Based Protein Identification. *J. Am. Chem. Soc.* **2022**, *144* (6), 2716–2725.
- (5) Swaminathan, J.; Boulgakov, A. A.; Hernandez, E. T.; Bardo, A. M.; Bachman, J. L.; Marotta, J.; Johnson, A. M.; Anslyn, E. V.; Marcotte, E. M. Highly Parallel Single-Molecule Identification of Proteins in Zeptomole-Scale Mixtures. *Nat. Biotechnol.* **2018**, *36* (11), 1076–1082.
- (6) Hunt, D. F.; Buko, A. M.; Ballard, J. M.; Shabanowitz, J.; Giordani, A. B. Sequence Analysis of Polypeptides by Collision Activated Dissociation on a Triple Quadrupole Mass Spectrometer. *Biol. Mass Spectrom.* **1981**, *8* (9), 397–408.
- (7) Gibson, B. W.; Biemann, K. Strategy for the Mass Spectrometric Verification and Correction of the Primary Structures of Proteins Deduced from Their DNA Sequences. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81* (7), 1956–1960.
- (8) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64–71.
- (9) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-flight Mass Spectrometry. *Rapid Comm Mass Spectrometry* **1988**, *2* (8), 151–153.
- (10) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
- (11) Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001**, *73* (23), 5683–5690.
- (12) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Anal. Chem.* **2003**, *75* (17), 4646–4658.
- (13) Elias, J. E.; Gygi, S. P. Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry. *Nat. Methods* **2007**, *4* (3), 207–214.
- (14) Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, *537* (7620), 347–355.
- (15) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top Down Versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **1999**, *121* (4), 806–812.
- (16) Smith, L. M.; Kelleher, N. L. Proteoform: A Single Term Describing Protein Complexity. *Nat. Methods* **2013**, *10* (3), 186–187.

- (17) Kelleher, N. L. Peer Reviewed: Top-Down Proteomics. *Anal. Chem.* **2004**, *76* (11), 196 A–203 A.
- (18) Cupp-Sutton, K. A.; Wu, S. High-Throughput Quantitative Top-down Proteomics. *Mol. Omics* **2020**, *16* (2), 91–99.
- (19) Smith, L. M.; Kelleher, N. L. Proteoforms as the Next Proteomics Currency. *Science* **2018**, *359* (6380), 1106–1107.
- (20) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schlüter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlén, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschläger, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B. How Many Human Proteoforms Are There? *Nat Chem Biol* **2018**, *14* (3), 206–214.
- (21) Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y. Top-Down Proteomics: Ready for Prime Time? *Anal. Chem.* **2018**, *90* (1), 110–127.
- (22) Meyer, J. G.; Niemi, N. M.; Pagliarini, D. J.; Coon, J. J. Quantitative Shotgun Proteome Analysis by Direct Infusion. *Nat Methods* **2020**, *17* (12), 1222–1228.
- (23) Jiang, Y.; Hutton, A.; Cranney, C. W.; Meyer, J. G. Label-Free Quantification from Direct Infusion Shotgun Proteome Analysis (DISPA-LFQ) with CsoDIAq Software. *Anal. Chem.* **2022**, DOI: 10.1021/acs.analchem.2c02249.
- (24) Jiang, Y.; Salladay-Perez, I.; Momenzadeh, A.; Covarrubias, A. J.; Meyer, J. G. Simultaneous Multi-Omics Analysis by Direct Infusion Mass Spectrometry (SMAD-MS). *bioRxiv*, 2023. DOI: 10.1101/2023.06.26.546628.
- (25) Omenn, G. S.; Lane, L.; Overall, C. M.; Pineau, C.; Packer, N. H.; Cristea, I. M.; Lindskog, C.; Weintraub, S. T.; Orchard, S.; Roehrl, M. H. A.; Nice, E.; Liu, S.; Bandeira, N.; Chen, Y.-J.; Guo, T.; Aebersold, R.; Moritz, R. L.; Deutsch, E. W. The 2022 Report on the Human Proteome from the HUPO Human Proteome Project. *J Proteome Res* **2023**, *22* (4), 1024–1042.
- (26) Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R. The Quantitative Proteome of a Human Cell Line. *Molecular Systems Biology* **2011**, *7* (1), 549.
- (27) Jiang, L.; Wang, M.; Lin, S.; Jian, R.; Li, X.; Chan, J.; Dong, G.; Fang, H.; Robinson, A. E.; Snyder, M. P.; Aguet, F.; Anand, S.; Ardlie, K. G.; Gabriel, S.; Getz, G.; Graubert, A.; Hadley, K.; Handsaker, R. E.; Huang, K. H.; Kashin, S.; MacArthur, D. G.; Meier, S. R.; Nedzel, J. L.; Nguyen, D. Y.; Segrè, A. V.; Todres, E.; Balliu, B.; Barbeira, A. N.; Battle, A.; Bonazzola, R.; Brown, A.; Brown, C. D.; Castel, S. E.; Conrad, D.; Cotter, D. J.; Cox, N.; Das, S.; de Goede, O. M.; Dermitzakis, E. T.; Engelhardt, B. E.; Eskin, E.; Eulalio, T. Y.; Ferraro, N. M.; Flynn, E.; Fresard, L.; Gamazon, E. R.; Garrido-Martín, D.; Gay, N. R.; Guigó, R.; Hamel, A. R.; He, Y.; Hoffman, P. J.; Hormozdiari, F.; Hou, L.; Im, H. K.; Jo, B.; Kasela, S.; Kellis, M.; Kim-Hellmuth, S.; Kwong, A.; Lappalainen, T.; Li, X.; Liang, Y.; Mangul, S.; Mohammadi, P.; Montgomery, S. B.; Muñoz-Aguirre, M.; Nachun, D. C.; Nobel, A. B.; Oliva, M.; Park, Y.; Park, Y.; Parsana, P.; Reverter, F.; Rouhana, J. M.; Sabatti, C.; Saha, A.; Skol, A. D.; Stephens, M.; Stranger, B. E.; Strober, B. J.; Teran, N. A.; Viñuela, A.; Wang, G.; Wen, X.; Wright, F.; Wucher, V.; Zou, Y.; Ferreira, P. G.; Li, G.; Melé, M.; Yeger-Lotem, E.; Barcus, M. E.; Bradbury, D.; Krubit, T.; McLean, J. A.; Qi, L.; Robinson, K.; Roche, N. V.; Smith, A. M.; Sobin, L.; Tabor, D. E.; Undale, A.; Bridge, J.; Brigham, L. E.; Foster, B. A.; Gillard, B. M.; Hasz, R.; Hunter, M.; Johns, C.; Johnson, M.; Karasik, E.; Kopen, G.; Leinweber, W. F.; McDonald, A.; Moser, M. T.; Myer, K.; Ramsey, K. D.; Roe, B.; Shad, S.; Thomas, J. A.; Walters, G.; Washington, M.; Wheeler, J.; Jewell, S. D.; Rohrer, D. C.; Valley, D. R.; Davis, D. A.; Mash, D. C.; Branton, P. A.; Barker, L. K.; Gardiner, H. M.; Mosavel, M.; Siminoff, L. A.; Flicek, P.; Haeussler, M.; Juettemann, T.; Kent, W. J.; Lee, C. M.; Powell, C. C.; Rosenbloom, K. R.; Ruffier, M.; Sheppard, D.; Taylor, K.; Trevanion, S. J.; Zerbino, D. R.; Abell, N. S.; Akey, J.; Chen, L.; Demanelis, K.; Doherty, J. A.; Feinberg, A. P.; Hansen, K. D.; Hickey, P. F.; Jasmine, F.; Kaul, R.; Kibria, M. G.; Li, J. B.; Li, Q.; Linder, S. E.; Pierce, B. L.; Rizzardi, L. F.; Smith, K. S.; Stamatoyannopoulos, J.; Tang, H.; Carithers, L. J.; Guan, P.; Koester, S. E.; Little, A. R.; Moore, H. M.; Nierras, C. R.; Rao, A. K.; Vaught, J. B.; Volpi, S. A Quantitative Proteome Map of the Human Body. *Cell* **2020**, *183* (1), 269–283.e19.
- (28) Sinitcyn, P.; Richards, A. L.; Weatheritt, R. J.; Brademan, D. R.; Marx, H.; Shishkova, E.; Meyer, J. G.; Hebert, A. S.; Westphall, M. S.; Blencowe, B. J.; Cox, J.; Coon, J. J. Global Detection of Human Variants and Isoforms by Deep Proteome Sequencing. *Nat Biotechnol* **2023**, *41* (12), 1776–1786.
- (29) Ramazi, S.; Zahir, J. Post-Translational Modifications in Proteins: Resources, Tools and Prediction Methods. *Database* **2021**, 2021, baab012.
- (30) Xu, H.; Wang, Y.; Lin, S.; Deng, W.; Peng, D.; Cui, Q.; Xue, Y. PTMD: A Database of Human Disease-Associated Post-Translational Modifications. *Genomics, Proteomics & Bioinformatics* **2018**, *16* (4), 244–251.
- (31) Wang, Y.-C.; Peterson, S. E.; Loring, J. F. Protein Post-Translational Modifications and Regulation of Pluripotency in Human Stem Cells. *Cell Res* **2014**, *24* (2), 143–160.
- (32) Rogers, L. D.; Overall, C. M. Proteolytic Post-Translational Modification of Proteins: Proteomic Tools and Methodology. *Mol Cell Proteomics* **2013**, *12* (12), 3532–3542.
- (33) Kitamura, N.; Galligan, J. J. A Global View of the Human Post-Translational Modification Landscape. *Biochemical Journal* **2023**, *480* (16), 1241–1265.
- (34) Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Lo Muzio, L. The Crucial Role of Protein Phosphorylation in Cell Signaling and Its Use as Targeted Therapy (Review). *Int J Mol Med* **2017**, *40* (2), 271–280.
- (35) Niemi, N. M.; MacKeigan, J. P. Mitochondrial Phosphorylation in Apoptosis: Flipping the Death Switch. *Antioxid Redox Signal* **2013**, *19* (6), 572–582.
- (36) Ruvolo, P. P.; Deng, X.; May, W. S. Phosphorylation of Bcl2 and Regulation of Apoptosis. *Leukemia* **2001**, *15* (4), 515–522.
- (37) Lecker, S. H.; Goldberg, A. L.; Mitch, W. E. Protein Degradation by the Ubiquitin-Proteasome Pathway in Normal and Disease States. *J Am Soc Nephrol* **2006**, *17* (7), 1807–1819.
- (38) Eifler, K.; Vertegaal, A. C. O. SUMOylation-Mediated Regulation of Cell Cycle Progression and Cancer. *Trends Biochem Sci* **2015**, *40* (12), 779–793.
- (39) Lumpkin, R. J.; Gu, H.; Zhu, Y.; Leonard, M.; Ahmad, A. S.; Clauser, K. R.; Meyer, J. G.; Bennett, E. J.; Komives, E. A. Site-Specific Identification and Quantitation of Endogenous SUMO Modifications Under Native Conditions. *Nat Commun* **2017**, *8* (1), 1171.
- (40) Romancino, D. P.; Buffà, V.; Caruso, S.; Ferrara, I.; Raccosta, S.; Notaro, A.; Campos, Y.; Noto, R.; Martorana, V.; Cupane, A.; Giallongo, A.; d’Azzo, A.; Manno, M.; Bongiovanni, A. Palmitoylation Is a Post-Translational Modification of Alix Regulating the Membrane Organization of Exosome-Like Small Extracellular Vesicles. *Biochim Biophys Acta Gen Subj* **2018**, *1862* (12), 2879–2887.
- (41) Clark, M. C.; Baum, L. G. T Cells Modulate Glycans on CD43 and CD45 During Development and Activation, Signal Regulation, and Survival. *Ann N Y Acad Sci* **2012**, *1253*, 58–67.
- (42) Marth, J. D.; Grewal, P. K. Mammalian Glycosylation in Immunity. *Nat Rev Immunol* **2008**, *8* (11), 874–887.
- (43) Conibear, A. C. Deciphering Protein Post-Translational Modifications Using Chemical Biology Tools. *Nat Rev Chem* **2020**, *4* (12), 674–695.
- (44) Zhu, H.; Klemic, J. F.; Chang, S.; Bertone, P.; Casamayor, A.; Klemic, K. G.; Smith, D.; Gerstein, M.; Reed, M. A.; Snyder, M. Analysis of Yeast Protein Kinases Using Protein Chips. *Nat Genet* **2000**, *26* (3), 283–289.

- (45) Reimer, U.; Reineke, U.; Schneider-Mergener, J. Peptide Arrays: From Macro to Micro. *Curr Opin Biotechnol* **2002**, *13* (4), 315–320.
- (46) Martyn, G. D.; Veggiani, G.; Kusebauch, U.; Morrone, S. R.; Yates, B. P.; Singer, A. U.; Tong, J.; Manczyk, N.; Gish, G.; Sun, Z.; Kurinov, I.; Sicheri, F.; Moran, M. F.; Moritz, R. L.; Sidhu, S. S. Engineered SH2 Domains for Targeted Phosphoproteomics. *ACS Chem Biol* **2022**, *17* (6), 1472–1484.
- (47) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **1995**, *57* (1), 289–300.
- (48) Agard, N. J.; Wells, J. A. Methods for the Proteomic Identification of Protease Substrates. *Current Opinion in Chemical Biology* **2009**, *13* (5–6), 503–509.
- (49) Villén, J.; Beausoleil, S. A.; Gerber, S. A.; Gygi, S. P. Large-Scale Phosphorylation Analysis of Mouse Liver. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (5), 1488–1493.
- (50) Bodenmiller, B.; Malmstrom, J.; Gerrits, B.; Campbell, D.; Lam, H.; Schmidt, A.; Rinner, O.; Mueller, L. N.; Shannon, P. T.; Pedrioli, P. G.; Panse, C.; Lee, H.; Schlapbach, R.; Aebersold, R. PhosphoPep—a Phosphoproteomic Resource for Systems Biology Research in *Drosophila* Kc167 Cells. *Molecular Systems Biology* **2007**, *3* (1), 139.
- (51) Kim, J.-E.; White, F. M. Quantitative Analysis of Phosphotyrosine Signaling Networks Triggered by CD3 and CD28 Costimulation in Jurkat Cells. *The Journal of Immunology* **2006**, *176* (5), 2833–2843.
- (52) Mayya, V.; Lundgren, D. H.; Hwang, S.-I.; Rezaul, K.; Wu, L.; Eng, J. K.; Rodionov, V.; Han, D. K. Quantitative Phosphoproteomic Analysis of T Cell Receptor Signaling Reveals System-Wide Modulation of Protein-Protein Interactions. *Sci. Signal.* **2009**, *2* (84), ra46.
- (53) White, F. M. Quantitative Phosphoproteomic Analysis of Signaling Network Dynamics. *Current Opinion in Biotechnology* **2008**, *19* (4), 404–409.
- (54) Bagdonaite, I.; Malaker, S. A.; Polasky, D. A.; Riley, N. M.; Schjoldager, K.; Vakhrushev, S. Y.; Halim, A.; Aoki-Kinoshita, K. F.; Nesvizhskii, A. I.; Bertozzi, C. R.; Wandall, H. H.; Parker, B. L.; Thaysen-Andersen, M.; Scott, N. E. Glycoproteomics. *Nat Rev Methods Primers* **2022**, *2* (1), 48.
- (55) Varki, A. Biological Roles of Glycans. *Glycobiology* **2017**, *27* (1), 3–49.
- (56) Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **2006**, *126* (5), 855–867.
- (57) Freeze, H. H.; Eklund, E. A.; Ng, B. G.; Patterson, M. C. Neurological Aspects of Human Glycosylation Disorders. *Annu. Rev. Neurosci.* **2015**, *38* (1), 105–125.
- (58) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat Methods* **2009**, *6* (5), 359–362.
- (59) Zougman, A.; Selby, P. J.; Banks, R. E. Suspension Trapping (STrap) Sample Preparation Method for Bottom-up Proteomics Analysis. *Proteomics* **2014**, *14* (9), 1006–1000.
- (60) Batth, T. S.; Tollenaere, M. X.; Rüther, P.; Gonzalez-Franquesa, A.; Prabhakar, B. S.; Bekker-Jensen, S.; Deshmukh, A. S.; Olsen, J. V. Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation*. *Molecular & Cellular Proteomics* **2019**, *18* (5), 1027a–1035.
- (61) HaileMariam, M.; Egue, R. V.; Singh, H.; Bekele, S.; Ameni, G.; Pieper, R.; Yu, Y. S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *J. Proteome Res.* **2018**, *17* (9), 2917–2924.
- (62) Cao, J.; Boatner, L. M.; Desai, H. S.; Burton, N. R.; Armenta, E.; Chan, N. J.; Castellón, J. O.; Backus, K. M. Multiplexed CuAAC Suzuki–Miyaura Labeling for Tandem Activity-Based Chemoproteomic Profiling. *Anal. Chem.* **2021**, *93* (4), 2610–2618.
- (63) Yan, T.; Desai, H.; Boatner, L.; Yen, S.; Cao, J.; Palafox, M.; Jami-Alahmadi, Y.; Backus, K. SP3-FAIMS Chemoproteomics for High Coverage Profiling of the Human Cysteineome. *ChemRxiv*, 2020. DOI: 10.26434/chemrxiv.13487364.v1.
- (64) Meyer, J. G.; Kim, S.; Maltby, D. A.; Ghassemlian, M.; Bandeira, N.; Komives, E. A. Expanding Proteome Coverage with Orthogonal-Specificity α -Lytic Proteases. *Molecular & Cellular Proteomics* **2014**, *13* (3), 823–835.
- (65) Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **2014**, *83* (1), 553–584.
- (66) Fatima, K.; Naqvi, F.; Younas, H. A Review: Molecular Chaperone-Mediated Folding, Unfolding and Disaggregation of Expressed Recombinant Proteins. *Cell Biochem Biophys* **2021**, *79* (2), 153–174.
- (67) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, 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.
- (68) Jumper, J.; Hassabis, D. The Protein Structure Prediction Revolution and Its Implications for Medicine: 2023 Albert Lasker Basic Medical Research Award. *JAMA* **2023**, *330* (15), 1425–1426.
- (69) Holding, A. N. XL-MS: Protein Cross-Linking Coupled with Mass Spectrometry. *Methods* **2015**, *89*, 54–63.
- (70) Steigenberger, B.; Albanese, P.; Heck, A. J. R.; Scheltema, R. A. To Cleave or Not To Cleave in XL-MS? *J. Am. Soc. Mass Spectrom.* **2020**, *31* (2), 196–206.
- (71) Yu, C.; Huang, L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* **2018**, *90* (1), 144–165.
- (72) Kaake, R. M.; Wang, X.; Burke, A.; Yu, C.; Kandur, W.; Yang, Y.; Novitsky, E. J.; Second, T.; Duan, J.; Kao, A.; Guan, S.; Vellucci, D.; Rychnovsky, S. D.; Huang, L. A New in Vivo Cross-Linking Mass Spectrometry Platform to Define Protein–Protein Interactions in Living Cells. *Molecular & Cellular Proteomics* **2014**, *13* (12), 3533–3543.
- (73) Yang, B.; Wu, Y.-J.; Zhu, M.; Fan, S.-B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y.-X.; Chen, H.-F.; Luo, S.-K.; Ding, Y.-H.; Wang, L.-H.; Hao, Z.; Xiu, L.-Y.; Chen, S.; Ye, K.; He, S.-M.; Dong, M.-Q. Identification of Cross-Linked Peptides from Complex Samples. *Nat Methods* **2012**, *9* (9), 904–906.
- (74) Hoopmann, M. R.; Zelter, A.; Johnson, R. S.; Riffle, M.; MacCoss, M. J.; Davis, T. N.; Moritz, R. L. Kojak: Efficient Analysis of Chemically Cross-Linked Protein Complexes. *J. Proteome Res.* **2015**, *14* (5), 2190–2198.
- (75) Hoopmann, M. R.; Shteynberg, D. D.; Zelter, A.; Riffle, M.; Lyon, A. S.; Agard, D. A.; Luan, Q.; Nolen, B. J.; MacCoss, M. J.; Davis, T. N.; Moritz, R. L. Improved Analysis of Cross-Linking Mass Spectrometry Data with Kojak 2.0, Advanced by Integration into the Trans-Proteomic Pipeline. *J. Proteome Res.* **2023**, *22* (2), 647–655.
- (76) Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L. N.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. Identification of Cross-Linked Peptides from Large Sequence Databases. *Nat Methods* **2008**, *5* (4), 315–318.
- (77) Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. Proteome-Wide Profiling of Protein Assemblies by Cross-Linking Mass Spectrometry. *Nature Methods* **2015**, *12* (12), 1179–1184.
- (78) Riffle, M.; Jaschob, D.; Zelter, A.; Davis, T. N. ProXL (Protein Cross-Linking Database): A Platform for Analysis, Visualization, and Sharing of Protein Cross-Linking Mass Spectrometry Data. *J. Proteome Res.* **2016**, *15* (8), 2863–2870.
- (79) Riffle, M.; Jaschob, D.; Zelter, A.; Davis, T. N. ProXL (Protein Cross-Linking Database): A Public Server, QC Tools, and Other Major Updates. *J. Proteome Res.* **2019**, *18* (2), 759–764.
- (80) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or

Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125* (7), 1731–1737.

(81) Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; Garlish, R.; Griffin, P. R.; Gross, M. L.; Guttman, M.; Hamuro, Y.; Heck, A. J. R.; Houde, D.; Jacob, R. E.; Jørgensen, T. J. D.; Kaltashov, I. A.; Klinman, J. P.; Konermann, L.; Man, P.; Mayne, L.; Pascal, B. D.; Reichmann, D.; Skehel, M.; Snijder, J.; Strutzenberg, T. S.; Underbakke, E. S.; Wagner, C.; Wales, T. E.; Walters, B. T.; Weis, D. D.; Wilson, D. J.; Wintrode, P. L.; Zhang, Z.; Zheng, J.; Schriemer, D. C.; Rand, K. D. Recommendations for Performing, Interpreting and Reporting Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Experiments. *Nat Methods* **2019**, *16* (7), 595–602.

(82) Konermann, L.; Pan, J.; Liu, Y.-H. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40* (3), 1224–1234.

(83) Hvidt, A.; Nielsen, S. O. Hydrogen Exchange in Proteins. *Adv Protein Chem* **1966**, *21*, 287–386.

(84) Englander, S. W.; Kallenbach, N. R. Hydrogen Exchange and Structural Dynamics of Proteins and Nucleic Acids. *Quart. Rev. Biophys.* **1983**, *16* (4), 521–655.

(85) Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R. Differential Hydrogen/Deuterium Exchange Mass Spectrometry Analysis of Protein–Ligand Interactions. *Expert Review of Proteomics* **2011**, *8* (1), 43–59.

(86) Jacob, R. E.; Engen, J. R. Hydrogen Exchange Mass Spectrometry: Are We Out of the Quicksand? *J. Am. Soc. Mass Spectrom.* **2012**, *23* (6), 1003–1010.

(87) Moroco, J. A.; Engen, J. R. Replication in Bioanalytical Studies with HDX MS: Aim as High as Possible. *Bioanalysis* **2015**, *7* (9), 1065–1067.

(88) Rand, K. D.; Jørgensen, T. J. D. Development of a Peptide Probe for the Occurrence of Hydrogen (H^1/H^2) Scrambling Upon Gas-Phase Fragmentation. *Anal. Chem.* **2007**, *79* (22), 8686–8693.

(89) Xu, G.; Chance, M. R. Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* **2007**, *107* (8), 3514–3543.

(90) Wang, L.; Chance, M. R. Structural Mass Spectrometry of Proteins Using Hydroxyl Radical Based Protein Footprinting. *Anal. Chem.* **2011**, *83* (19), 7234–7241.

(91) Tullius, T. D.; Greenbaum, J. A. Mapping Nucleic Acid Structure by Hydroxyl Radical Cleavage. *Current Opinion in Chemical Biology* **2005**, *9* (2), 127–134.

(92) Tullius, T. D. Probing DNA Structure with Hydroxyl Radicals. *CP Nucleic Acid Chemistry* **2001**, *7* (1), 6.7.1.

(93) Biswas, P. K.; Chakraborty, S. Targeted DNA Oxidation and Trajectory of Radical DNA Using DFT Based QM/MM Dynamics. *Nucleic Acids Research* **2019**, *47* (6), 2757–2765.

(94) Niu, B.; Gross, M. L. MS-Based Hydroxyl Radical Footprinting: Methodology and Application of Fast Photochemical Oxidation of Proteins (FPOP). *Mass Spectrometry-Based Chemical Proteomics* **2019**, 363–416.

(95) Feng, J.; Chu, C.; Ma, Z. Fenton and Fenton-Like Catalysts for Electrochemical Immunoassay: A Mini Review. *Electrochemistry Communications* **2021**, *125*, No. 106970.

(96) Nieto-Juarez, J. I.; Pierzchla, K.; Sienkiewicz, A.; Kohn, T. Inactivation of MS2 Coliphage in Fenton and Fenton-Like Systems: Role of Transition Metals, Hydrogen Peroxide and Sunlight. *Environ. Sci. Technol.* **2010**, *44* (9), 3351–3356.

(97) Houée-Levin, C.; Bobrowski, K. The Use of the Methods of Radiolysis to Explore the Mechanisms of Free Radical Modifications in Proteins. *Journal of Proteomics* **2013**, *92*, 51–62.

(98) Hernández-Corroto, E.; Boussetta, N.; Marina, M. L.; García, M. C.; Vorobiev, E. High Voltage Electrical Discharges Followed by Deep Eutectic Solvents Extraction for the Valorization of Pomegranate Seeds (*Punica Granatum* L.). *Innovative Food Science & Emerging Technologies* **2022**, *79*, 103055.

(99) Jolivet, P.; Aymé, L.; Giuliani, A.; Wien, F.; Chardot, T.; Gohon, Y. Structural Proteomics: Topology and Relative Accessibility of Plant Lipid Droplet Associated Proteins. *Journal of Proteomics* **2017**, *169*, 87–98.

(100) Stulić, V.; Vukušić, T.; Butorac, A.; Popović, D.; Herceg, Z. Proteomic Analysis of *Saccharomyces Cerevisiae* Response to Plasma Treatment. *International Journal of Food Microbiology* **2019**, *292*, 171–183.

(101) Hambly, D. M.; Gross, M. L. Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (12), 2057–2063.

(102) Johnson, D. T.; Di Stefano, L. H.; Jones, L. M. Fast Photochemical Oxidation of Proteins (FPOP): A Powerful Mass Spectrometry–Based Structural Proteomics Tool. *Journal of Biological Chemistry* **2019**, *294* (32), 11969–11979.

(103) Hambly, D.; Gross, M. Laser Flash Photochemical Oxidation to Locate Heme Binding and Conformational Changes in Myoglobin. *International Journal of Mass Spectrometry* **2007**, *259* (1–3), 124–129.

(104) Chea, E. E.; Jones, L. M. Modifications Generated by Fast Photochemical Oxidation of Proteins Reflect the Native Conformations of Proteins. *Protein Science* **2018**, *27* (6), 1047–1056.

(105) Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L. Fast Photochemical Oxidation of Protein Footprints Faster Than Protein Unfolding. *Anal. Chem.* **2009**, *81* (16), 6563–6571.

(106) Luchini, A.; Espina, V.; Liotta, L. A. Protein Painting Reveals Solvent-Excluded Drug Targets Hidden Within Native Protein–Protein Interfaces. *Nat Commun* **2014**, *5* (1), 4413.

(107) Hammond, A.; Dey, D.; Carter, R.; Dailing, A.; Nara, V.; Nara, P.; Venkatayogi, S.; Paige, M.; Liotta, L.; Luchini, A. Protein Painting, an Optimized MS-Based Technique, Reveals Functionally Relevant Interfaces of the PD-1/PD-L1 Complex and the YAP2/ZO-1 Complex. *Journal of Biological Chemistry* **2019**, *294* (29), 11180–11198.

(108) Pepelnjak, M.; de Souza, N.; Picotti, P. Detecting Protein–Small Molecule Interactions Using Limited Proteolysis–Mass Spectrometry (LiP-MS). *Trends in Biochemical Sciences* **2020**, *45* (10), 919–920.

(109) Schopper, S.; Kahraman, A.; Leuenberger, P.; Feng, Y.; Piazza, I.; Müller, O.; Boersema, P. J.; Picotti, P. Measuring Protein Structural Changes on a Proteome-Wide Scale Using Limited Proteolysis–Coupled Mass Spectrometry. *Nat Protoc* **2017**, *12* (11), 2391–2410.

(110) Malinowska, L.; Cappelletti, V.; Kohler, D.; Piazza, I.; Tsai, T.-H.; Pepelnjak, M.; Stalder, P.; Dörig, C.; Sesterhenn, F.; Elsässer, F.; Kralickova, L.; Beaton, N.; Reiter, L.; de Souza, N.; Vitek, O.; Picotti, P. Proteome-Wide Structural Changes Measured with Limited Proteolysis–Mass Spectrometry: An Advanced Protocol for High-Throughput Applications. *Nat Protoc* **2023**, *18* (3), 659–682.

(111) Barret, D. C. A.; Schuster, D.; Rodrigues, M. J.; Leitner, A.; Picotti, P.; Schertler, G. F. X.; Kaupp, U. B.; Korkhov, V. M.; Marino, J. Structural Basis of Calmodulin Modulation of the Rod Cyclic Nucleotide-Gated Channel. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120* (15), e2300309120.

(112) <https://www.eubopen.org/sites/www.eubopen.org/files/attachments/2022/LiP-MS%20protocol.pdf>.

(113) Molina, D. M.; 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.

(114) Prabhu, N.; Dai, L.; Nordlund, P. CETSA in Integrated Proteomics Studies of Cellular Processes. *Current Opinion in Chemical Biology* **2020**, *54*, 54–62.

(115) Mateus, A.; Kurzawa, N.; Becher, I.; Sridharan, S.; Helm, D.; Stein, F.; Typas, A.; Savitski, M. M. Thermal Proteome Profiling for Interrogating Protein Interactions. *Molecular Systems Biology* **2020**, *16* (3), e9232.

(116) Le Sueur, C.; Hammarén, H. M.; Sridharan, S.; Savitski, M. M. Thermal Proteome Profiling: Insights into Protein Modifications,

Associations, and Functions. *Current Opinion in Chemical Biology* **2022**, *71*, No. 102225.

(117) King, D. T.; Serrano-Negrón, J. E.; Zhu, Y.; Moore, C. L.; Shoulders, M. D.; Foster, L. J.; Vocadlo, D. J. Thermal Proteome Profiling Reveals the O-GlcNAc-Dependent Meltome. *J. Am. Chem. Soc.* **2022**, *144* (9), 3833–3842.

(118) Sridharan, S.; Günthner, I.; Becher, I.; Savitski, M.; Bantscheff, M. Target Discovery Using Thermal Proteome Profiling. *Mass Spectrometry-Based Chemical Proteomics* **2019**, 267–291.

(119) Marcilla, M.; Albar, J. P. Quantitative Proteomics: A Strategic Ally to Map Protein Interaction Networks. *IUBMB Life* **2013**, *65* (1), 9–16.

(120) Altelaar, A. F. M.; Munoz, J.; Heck, A. J. R. Next-Generation Proteomics: Towards an Integrative View of Proteome Dynamics. *Nat Rev Genet* **2013**, *14* (1), 35–48.

(121) Varjosalo, M.; Sacco, R.; Stukalov, A.; van Drogen, A.; Planyavsky, M.; Hauri, S.; Aebersold, R.; Bennett, K. L.; Colinge, J.; Gstaiger, M.; Superti-Furga, G. Interlaboratory Reproducibility of Large-Scale Human Protein-Complex Analysis by Standardized AP-MS. *Nat Methods* **2013**, *10* (4), 307–314.

(122) 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.; Vaite, 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.

(123) Huttlin, E. L.; Bruckner, R. J.; Paulo, J. A.; Cannon, J. R.; Ting, L.; Baltier, K.; Colby, G.; Gebreab, F.; Gygi, M. P.; Parzen, H.; Szpyt, J.; Tam, S.; Zarraga, G.; Pontano-Vaite, L.; Swarup, S.; White, A. E.; Schweppe, D. K.; Rad, R.; Erickson, B. K.; Obar, R. A.; Gururharsha, K. G.; Li, K.; Artavanis-Tsakonas, S.; Gygi, S. P.; Harper, J. W. Architecture of the Human Interactome Defines Protein Communities and Disease Networks. *Nature* **2017**, *545* (7655), 505–509.

(124) 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.; Vaite, 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.e28.

(125) Martell, J. D.; Deerinck, T. J.; Sancak, Y.; Poulos, T. L.; Mootha, V. K.; Sosinsky, G. E.; Ellisman, M. H.; Ting, A. Y. Engineered Ascorbate Peroxidase as a Genetically Encoded Reporter for Electron Microscopy. *Nat Biotechnol* **2012**, *30* (11), 1143–1148.

(126) Rhee, H.-W.; Zou, P.; Udeshi, N. D.; Martell, J. D.; Mootha, V. K.; Carr, S. A.; Ting, A. Y. Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science* **2013**, *339* (6125), 1328–1331.

(127) Hung, V.; Zou, P.; Rhee, H.-W.; Udeshi, N. D.; Cracan, V.; Svinkina, T.; Carr, S. A.; Mootha, V. K.; Ting, A. Y. Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Molecular Cell* **2014**, *55* (2), 332–341.

(128) Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. A Promiscuous Biotin Ligase Fusion Protein Identifies Proximal and Interacting Proteins in Mammalian Cells. *Journal of Cell Biology* **2012**, *196* (6), 801–810.

(129) Samavarchi-Tehrani, P.; Samson, R.; Gingras, A.-C. Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches. *Molecular & Cellular Proteomics* **2020**, *19* (5), 757–773.

(130) Varnaité, R.; MacNeill, S. A. Meet the Neighbors: Mapping Local Protein Interactomes by Proximity-dependent Labeling with BioID. *Proteomics* **2016**, *16* (19), 2503–2518.

(131) Kim, D. I.; Jensen, S. C.; Noble, K. A.; KC, B.; Roux, K. H.; Motamedchaboki, K.; Roux, K. J. An Improved Smaller Biotin Ligase for BioID Proximity Labeling. *MBoC* **2016**, *27* (8), 1188–1196.

(132) Cho, K. F.; Branon, T. C.; Udeshi, N. D.; Myers, S. A.; Carr, S. A.; Ting, A. Y. Proximity Labeling in Mammalian Cells with TurboID and Split-TurboID. *Nat Protoc* **2020**, *15* (12), 3971–3999.

(133) Li, P.; Meng, Y.; Wang, L.; Di, L. BioID: A Proximity-Dependent Labeling Approach in Proteomics Study. In *Functional Proteomics*; Springer New York, 2018; pp 143–151. DOI: 10.1007/978-1-4939-8814-3_10.

(134) Burgess, R. R.; Deutscher, M. P. *Guide to protein purification*, 2nd ed.; Methods in enzymology; Elsevier/Academic Press: Amsterdam Boston, 2009.

(135) *The protein protocols handbook*, 3. ed.; Walker, J. M., Ed.; Springer protocols handbooks; Humana Press: New York, NY, 2009.

(136) Linn, S. Strategies and Considerations for Protein Purifications. *Methods Enzymol* **2009**, *463*, 9–19.

(137) *Proteins and Proteomics: A Laboratory Manual*; Simpson, R. J., Ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2003.

(138) Feist, P.; Hummon, A. B. Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples. *Int J Mol Sci* **2015**, *16* (2), 3537–3563.

(139) Salvi, G.; De Los Rios, P.; Vendruscolo, M. Effective Interactions Between Chaotropic Agents and Proteins. *Proteins* **2005**, *61* (3), 492–499.

(140) Proc, J. L.; Kuzyk, M. A.; Hardie, D. B.; Yang, J.; Smith, D. S.; Jackson, A. M.; Parker, C. E.; Borchers, C. H. A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. *J Proteome Res* **2010**, *9* (10), 5422–5437.

(141) Sun, S.; Zhou, J.-Y.; Yang, W.; Zhang, H. Inhibition of Protein Carbamylation in Urea Solution Using Ammonium-Containing Buffers. *Anal Biochem* **2014**, *446*, 76–81.

(142) Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whiteledge, J. P.; Paša-Tolić, L.; Chamot-Rooke, J.; Danis, P. O.; Smith, L. M.; Tsybin, Y. O.; Loo, J. A.; Ge, Y.; Kelleher, N. L.; Agar, J. N. Best Practices and Benchmarks for Intact Protein Analysis for Top-down Mass Spectrometry. *Nat Methods* **2019**, *16* (7), 587–594.

(143) Zhang, X. Less Is More: Membrane Protein Digestion Beyond Urea–Trypsin Solution for Next-Level Proteomics. *Molecular & Cellular Proteomics* **2015**, *14* (9), 2441–2453.

(144) Jiang, L.; He, L.; Fountoulakis, M. Comparison of Protein Precipitation Methods for Sample Preparation Prior to Proteomic Analysis. *J Chromatogr A* **2004**, *1023* (2), 317–320.

(145) Burgess, R. R. Protein Precipitation Techniques. *Methods Enzymol* **2009**, *463*, 331–342.

(146) Wojtkiewicz, M.; Berg Luecke, L.; Kelly, M. I.; Gundry, R. L. Facile Preparation of Peptides for Mass Spectrometry Analysis in Bottom-Up Proteomics Workflows. *Curr Protoc* **2021**, *1* (3), No. e85.

(147) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat Methods* **2009**, *6* (5), 359–362.

(148) Wiśniewski, J. R. Filter-Aided Sample Preparation for Proteome Analysis. *Methods Mol Biol* **2018**, *1841*, 3–10.

(149) Kodama, H.; Tokuman, N.; Yasui, I.; Gyobu, Y.; Kashiwagi, Y. Group and Type Distribution of Hemolytic Streptococci Isolated from Clinical Specimens—Prevalence of Group A Type 3 Isolates in 1985 in Toyama Prefecture. *Kansenshogaku Zasshi* **1987**, *61* (4), 482–488.

(150) Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J Proteome Res* **2017**, *16* (11), 4060–4072.

(151) Batth, T. S.; Tollenaere, M. X.; Rüther, P.; Gonzalez-Franquesa, A.; Prabhakar, B. S.; Bekker-Jensen, S.; Deshmukh, A. S.;

- Olsen, J. V. Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol Cell Proteomics* **2019**, *18* (5), 1027a–1035.
- (152) Ludwig, K. R.; Schroll, M. M.; Hummon, A. B. Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies. *J Proteome Res* **2018**, *17* (7), 2480–2490.
- (153) Masuda, T.; Tomita, M.; Ishihama, Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. *J Proteome Res* **2008**, *7* (2), 731–740.
- (154) Yeung, Y.-G.; Nieves, E.; Angeletti, R. H.; Stanley, E. R. Removal of Detergents from Protein Digests for Mass Spectrometry Analysis. *Analytical Biochemistry* **2008**, *382* (2), 135–137.
- (155) Sun, Z.; Asmann, Y. W.; Kalari, K. R.; Bot, B.; Eckel-Passow, J. E.; Baker, T. R.; Carr, J. M.; Khrebtkova, I.; Luo, S.; Zhang, L.; Schroth, G. P.; Perez, E. A.; Thompson, E. A. Integrated Analysis of Gene Expression, CpG Island Methylation, and Gene Copy Number in Breast Cancer Cells by Deep Sequencing. *PLoS One* **2011**, *6* (2), e17490.
- (156) Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Séraphin, B. A Generic Protein Purification Method for Protein Complex Characterization and Proteome Exploration. *Nat Biotechnol* **1999**, *17* (10), 1030–1032.
- (157) Chang, Y.-H.; Gregorich, Z. R.; Chen, A. J.; Hwang, L.; Guner, H.; Yu, D.; Zhang, J.; Ge, Y. New Mass-Spectrometry-Compatible Degradable Surfactant for Tissue Proteomics. *J Proteome Res* **2015**, *14* (3), 1587–1599.
- (158) Knott, S. J.; Brown, K. A.; Josyer, H.; Carr, A.; Inman, D.; Jin, S.; Friedl, A.; Ponik, S. M.; Ge, Y. Photocleavable Surfactant-Enabled Extracellular Matrix Proteomics. *Anal Chem* **2020**, *92* (24), 15693–15698.
- (159) Mbeunkui, F.; Goshe, M. B. Investigation of Solubilization and Digestion Methods for Microsomal Membrane Proteome Analysis Using Data-Independent LC-MSE. *Proteomics* **2011**, *11* (5), 898–911.
- (160) Lin, Y.; Huo, L.; Liu, Z.; Li, J.; Liu, Y.; He, Q.; Wang, X.; Liang, S. Sodium Laurate, a Novel Protease- and Mass Spectrometry-Compatible Detergent for Mass Spectrometry-Based Membrane Proteomics. *PLoS One* **2013**, *8* (3), e59779.
- (161) Zhou, J.; Zhou, T.; Cao, R.; Liu, Z.; Shen, J.; Chen, P.; Wang, X.; Liang, S. Evaluation of the Application of Sodium Deoxycholate to Proteomic Analysis of Rat Hippocampal Plasma Membrane. *J Proteome Res* **2006**, *5* (10), 2547–2553.
- (162) Ahmadi, S.; Winter, D. Identification of Poly(ethylene Glycol) and Poly(ethylene Glycol)-Based Detergents Using Peptide Search Engines. *Anal Chem* **2018**, *90* (11), 6594–6600.
- (163) León, I. R.; Schwämmle, V.; Jensen, O. N.; Sprenger, R. R. Quantitative Assessment of in-Solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis. *Mol Cell Proteomics* **2013**, *12* (10), 2992–3005.
- (164) Tobin, A. K.; Bowshe, C. G. Subcellular Fractionation of Plant Tissues: Isolation of Plastids and Mitochondria. In *Protein Purification Protocols*; Humana Press; pp 53–64. DOI: 10.1385/1-59259-655-x:53.
- (165) Lande, N. V.; Barua, P.; Gayen, D.; Kumar, S.; Chakraborty, S.; Chakraborty, N. Proteomic Dissection of the Chloroplast: Moving Beyond Photosynthesis. *Journal of Proteomics* **2020**, *212*, No. 103542.
- (166) Millar, A. H.; Liddell, A.; Leaver, C. J. Isolation and Subfractionation of Mitochondria from Plants. *Methods Cell Biol* **2007**, *80*, 65–90.
- (167) Wilken, L. R.; Nikolov, Z. L. Recovery and Purification of Plant-Made Recombinant Proteins. *Biotechnology Advances* **2012**, *30* (2), 419–433.
- (168) Plaxton, W. C. Avoiding Proteolysis During the Extraction and Purification of Active Plant Enzymes. *Plant and Cell Physiology* **2019**, *60* (4), 715–724.
- (169) Pierpoint, W. S. The Extraction of Enzymes From Plant Tissues Rich in Phenolic Compounds. In *Protein Purification Protocols*; Humana Press; pp 65–74. DOI: 10.1385/1-59259-655-x:65.
- (170) Isaacson, T.; Damasceno, C. M. B.; Saravanan, R. S.; He, Y.; Catalá, C.; Saladié, M.; Rose, J. K. C. Sample Extraction Techniques for Enhanced Proteomic Analysis of Plant Tissues. *Nat Protoc* **2006**, *1* (2), 769–774.
- (171) Wu, X.; Gong, F.; Wang, W. Protein Extraction from Plant Tissues for 2DE and Its Application in Proteomic Analysis. *Proteomics* **2014**, *14* (6), 645–658.
- (172) Niu, L.; Yuan, H.; Gong, F.; Wu, X.; Wang, W. Protein Extraction Methods Shape Much of the Extracted Proteomes. *Front. Plant Sci.* **2018**, *9*, 802.
- (173) Mergner, J.; Kuster, B. Plant Proteome Dynamics. *Annu. Rev. Plant Biol.* **2022**, *73* (1), 67–92.
- (174) Song, G.; Hsu, P. Y.; Walley, J. W. Assessment and Refinement of Sample Preparation Methods for Deep and Quantitative Plant Proteome Profiling. *Proteomics* **2018**, *18* (17), 1800220.
- (175) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-Pot, Solid-Phase-Enhanced Sample Preparation for Proteomics Experiments. *Nat Protoc* **2019**, *14* (1), 68–85.
- (176) Mikulášek, K.; Konečná, H.; Potěšil, D.; Holánková, R.; Havliš, J.; Zdráhal, Z. SP3 Protocol for Proteomic Plant Sample Preparation Prior LC-MS/MS. *Front. Plant Sci.* **2021**, *12*, 635550.
- (177) Gupta, R.; Kim, S. T. Depletion of RuBisCO Protein Using the Protamine Sulfate Precipitation Method. In *Methods in Molecular Biology*; Springer New York, 2015; pp 225–233. DOI: 10.1007/978-1-4939-2550-6_17.
- (178) Smolikova, G.; Gorbach, D.; Lukasheva, E.; Mavropoulos-Stolyarenko, G.; Bilova, T.; Soboleva, A.; Tsarev, A.; Romanovskaya, E.; Podolskaya, E.; Zhukov, V.; Tikhonovich, I.; Medvedev, S.; Hoehenwarter, W.; Frolov, A. Bringing New Methods to the Seed Proteomics Platform: Challenges and Perspectives. *IJMS* **2020**, *21* (23), 9162.
- (179) Gomes, T. A.; Zanette, C. M.; Spier, M. R. An Overview of Cell Disruption Methods for Intracellular Biomolecules Recovery. *Preparative Biochemistry & Biotechnology* **2020**, *50* (7), 635–654.
- (180) Salomon, I.; Janssen, H.; Neeffes, J. Mechanical Forces Used for Cell Fractionation Can Create Hybrid Membrane Vesicles. *Int J Biol Sci* **2010**, *6* (7), 649–654.
- (181) Molnar, A.; Lakat, T.; Hosszu, A.; Szebeni, B.; Balogh, A.; Orfi, L.; Szabo, A. J.; Fekete, A.; Hodrea, J. Lyophilization and Homogenization of Biological Samples Improves Reproducibility and Reduces Standard Deviation in Molecular Biology Techniques. *Amino Acids* **2021**, *53* (6), 917–928.
- (182) Cai, X.; Xue, Z.; Wu, C.; Sun, R.; Qian, L.; Yue, L.; Ge, W.; Yi, X.; Liu, W.; Chen, C.; Gao, H.; Yu, J.; Xu, L.; Zhu, Y.; Guo, T. High-Throughput Proteomic Sample Preparation Using Pressure Cycling Technology. *Nat Protoc* **2022**, *17* (10), 2307–2325.
- (183) Olson, B. J. S. C.; Markwell, J. Assays for Determination of Protein Concentration. *Curr Protoc Protein Sci* **2007**, *48*, 3.4.
- (184) Jones, L. J.; Haugland, R. P.; Singer, V. L. Development and Characterization of the NanoOrange Protein Quantitation Assay: A Fluorescence-Based Assay of Proteins in Solution. *Biotechniques* **2003**, *34* (4), 850–854. 856, 858 passim
- (185) Brady, P. N.; Macnaughtan, M. A. Evaluation of Colorimetric Assays for Analyzing Reductively Methylated Proteins: Biases and Mechanistic Insights. *Anal Biochem* **2015**, *491*, 43–51.
- (186) Contreras-Martos, S.; Nguyen, H. H.; Nguyen, P. N.; Hristozova, N.; Macossay-Castillo, M.; Kovacs, D.; Bekesi, A.; Oemig, J. S.; Maes, D.; Pauwels, K.; Tompa, P.; Lebrun, P. Quantification of Intrinsically Disordered Proteins: A Problem Not Fully Appreciated. *Front Mol Biosci* **2018**, *5*, 83.
- (187) Sapan, C. V.; Lundblad, R. L.; Price, N. C. Colorimetric Protein Assay Techniques. *Biotechnol Appl Biochem* **1999**, *29* (2), 99–108.
- (188) Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C. Measurement of Protein Using Bicinchoninic Acid. *Anal Biochem* **1985**, *150* (1), 76–85.

- (189) Walker, J. M. The Bicinchoninic Acid (BCA) Assay for Protein Quantitation. *Methods Mol Biol* **1994**, 32, 5–8.
- (190) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal Biochem* **1976**, 72, 248–254.
- (191) Kielkopf, C. L.; Bauer, W.; Urbatsch, I. L. Bradford Assay for Determining Protein Concentration. *Cold Spring Harb Protoc* **2020**, 2020 (4), 102269.
- (192) Jones, A.; Razniewska, T.; Lesser, B. H.; Siqueira, R.; Berk, D.; Behie, L. A.; Gaucher, G. M. An Assay for the Measurement of the Protein Content of Cells Immobilized in Carrageenan. *Can J Microbiol* **1984**, 30 (4), 475–481.
- (193) Duncombe, T. A.; Ponti, A.; Seebeck, F. P.; Dittrich, P. S. UV-Vis Spectra-Activated Droplet Sorting for Label-Free Chemical Identification and Collection of Droplets. *Anal Chem* **2021**, 93 (38), 13008–13013.
- (194) Wiśniewski, J. R.; Gaugaz, F. Z. Fast and Sensitive Total Protein and Peptide Assays for Proteomic Analysis. *Anal Chem* **2015**, 87 (8), 4110–4116.
- (195) You, W. W.; Haugland, R. P.; Ryan, D. K.; Haugland, R. P. 3-(4-Carboxybenzoyl)quinoline-2-Carboxaldehyde, a Reagent with Broad Dynamic Range for the Assay of Proteins and Lipoproteins in Solution. *Anal Biochem* **1997**, 244 (2), 277–282.
- (196) Suttapitugsakul, S.; Xiao, H.; Smeekens, J.; Wu, R. Evaluation and Optimization of Reduction and Alkylation Methods to Maximize Peptide Identification with MS-Based Proteomics. *Mol Biosyst* **2017**, 13 (12), 2574–2582.
- (197) Hale, J. E.; Butler, J. P.; Gelfanova, V.; You, J.-S.; Knierman, M. D. A Simplified Procedure for the Reduction and Alkylation of Cysteine Residues in Proteins Prior to Proteolytic Digestion and Mass Spectral Analysis. *Anal Biochem* **2004**, 333 (1), 174–181.
- (198) Müller, T.; Winter, D. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-Containing Reagents. *Mol Cell Proteomics* **2017**, 16 (7), 1173–1187.
- (199) Gao, X.-H.; Li, L.; Parisien, M.; Wu, J.; Bederman, I.; Gao, Z.; Krokowski, D.; Chirieleison, S. M.; Abbott, D.; Wang, B.; Arvan, P.; Cameron, M.; Chance, M.; Willard, B.; Hatzoglou, M. Discovery of a Redox Thiol Switch: Implications for Cellular Energy Metabolism. *Mol Cell Proteomics* **2020**, 19 (5), 852–870.
- (200) Irvine, G. W.; Stillman, M. J. Residue Modification and Mass Spectrometry for the Investigation of Structural and Metalation Properties of Metallothionein and Cysteine-Rich Proteins. *Int J Mol Sci* **2017**, 18 (5), 913.
- (201) Murray, C. I.; Van Eyk, J. E. Chasing Cysteine Oxidative Modifications: Proteomic Tools for Characterizing Cysteine Redox Status. *Circ Cardiovasc Genet* **2012**, 5 (5), 591.
- (202) Vandermarliere, E.; Mueller, M.; Martens, L. Getting Intimate with Trypsin, the Leading Protease in Proteomics. *Mass Spec Rev* **2013**, 32 (6), 453–465.
- (203) Swaney, D. L.; Wenger, C. D.; Coon, J. J. Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *J Proteome Res* **2010**, 9 (3), 1323–1329.
- (204) Tsiatsiani, L.; Heck, A. J. R. Proteomics Beyond Trypsin. *FEBS J* **2015**, 282 (14), 2612–2626.
- (205) Meyer, J. G. *ISRN Comput Biol* **2014**, 2014, 960902.
- (206) Meyer, J. G. *<I>In Silico</i> Proteome Cleavage Reveals Iterative Digestion Strategy for High Sequence Coverage. *ISRN Computational Biology* **2014**, 2014, 1–7.*
- (207) Choudhary, G.; Wu, S.-L.; Shieh, P.; Hancock, W. S. Multiple Enzymatic Digestion for Enhanced Sequence Coverage of Proteins in Complex Proteomic Mixtures Using Capillary LC with Ion Trap MS/MS. *J Proteome Res* **2003**, 2 (1), 59–67.
- (208) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. Six Alternative Proteases for Mass Spectrometry-Based Proteomics Beyond Trypsin. *Nat Protoc* **2016**, 11 (5), 993–1006.
- (209) Blank-Landeshammer, B.; Teichert, I.; Märker, R.; Nowrousian, M.; Kück, U.; Sickmann, A. Combination of Proteogenomics with Peptide. *mBio* **2019**, 10 (5), 10.1128/mbio.02367-19.
- (210) Yang, H.; Li, Y.-C.; Zhao, M.-Z.; Wu, F.-L.; Wang, X.; Xiao, W.-D.; Wang, Y.-H.; Zhang, J.-L.; Wang, F.-Q.; Xu, F.; Zeng, W.-F.; Overall, C. M.; He, S.-M.; Chi, H.; Xu, P. Precision. *Mol Cell Proteomics* **2019**, 18 (4), 773–785.
- (211) Vanuopadath, M.; Sajeev, N.; Murali, A. R.; Sudish, N.; Kangosseri, N.; Sebastian, I. R.; Jain, N. D.; Pal, A.; Raveendran, D.; Nair, B. G.; Nair, S. S. Mass Spectrometry-Assisted Venom Profiling of Hypnale Hypnale Found in the Western Ghats of India Incorporating de Novo Sequencing Approaches. *Int J Biol Macromol* **2018**, 118 (Pt B), 1736–1746.
- (212) Vanuopadath, M.; Raveendran, D.; Nair, B. G.; Nair, S. S. Venomics and Antivenomics of Indian Spectacled Cobra (*Naja Naja*) from the Western Ghats. *Acta Tropica* **2022**, 228, No. 106324.
- (213) Guthals, A.; Clauser, K. R.; Frank, A. M.; Bandeira, N. Sequencing-Grade *<i>De Novo</i>* Analysis of MS/MS Triplets (CID/HCD/ETD) From Overlapping Peptides. *J. Proteome Res.* **2013**, 12 (6), 2846–2857.
- (214) Wiśniewski, J. R.; Wegler, C.; Artursson, P. Multiple-Enzyme-Digestion Strategy Improves Accuracy and Sensitivity of Label- and Standard-Free Absolute Quantification to a Level That Is Achievable by Analysis with Stable Isotope-Labeled Standard Spiking. *J Proteome Res* **2018**, 18 (1), 217–224.
- (215) Miller, R. M.; Millikin, R. J.; Hoffmann, C. V.; Solntsev, S. K.; Sheynkman, G. M.; Shortreed, M. R.; Smith, L. M. Improved Protein Inference from Multiple Protease Bottom-Up Mass Spectrometry Data. *J. Proteome Res.* **2019**, 18 (9), 3429–3438.
- (216) Kaulich, P. T.; Cassidy, L.; Bartel, J.; Schmitz, R. A.; Tholey, A. Multi-Protease Approach for the Improved Identification and Molecular Characterization of Small Proteins and Short Open Reading Frame-Encoded Peptides. *J. Proteome Res.* **2021**, 20 (5), 2895–2903.
- (217) Vincent, D.; Ezernieks, V.; Rochfort, S.; Spangenberg, G. A. Multiple Protease Strategy to Optimize the Shotgun Proteomics of Mature Medicinal Cannabis Buds. *IJMS* **2019**, 20 (22), 5630.
- (218) Guo, X.; Trudgian, D. C.; Lemoff, A.; Yadavalli, S.; Mirzaei, H. Confetti: A Multiprotease Map of the HeLa Proteome for Comprehensive Proteomics. *Molecular & Cellular Proteomics* **2014**, 13 (6), 1573–1584.
- (219) Giansanti, P.; Aye, T. T.; van den Toorn, H.; Peng, M.; van Breukelen, B.; Heck, A. J. R. An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas. *Cell Reports* **2015**, 11 (11), 1834–1843.
- (220) Jekel, P. A.; Weijer, W. J.; Beintema, J. J. Use of Endoprotease Lys-C from *Lysobacter* Enzymogenes in Protein Sequence Analysis. *Anal Biochem* **1983**, 134 (2), 347–354.
- (221) Glatzer, T.; Ludwig, C.; Ahrné, E.; Aebersold, R.; Heck, A. J. R.; Schmidt, A. Large-Scale Quantitative Assessment of Different in-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion. *J Proteome Res* **2012**, 11 (11), 5145–5156.
- (222) Epstein, D. M.; Wensink, P. C. The Alpha-Lytic Protease Gene of *Lysobacter* Enzymogenes. The Nucleotide Sequence Predicts a Large Prepro-Peptide with Homology to Pro-Peptides of Other Chymotrypsin-Like Enzymes. *J Biol Chem* **1988**, 263 (32), 16586–16590.
- (223) Drapeau, G. R.; Boily, Y.; Houmar, J. Purification and Properties of an Extracellular Protease of *Staphylococcus Aureus*. *J Biol Chem* **1972**, 247 (20), 6720–6726.
- (224) Liu, S.; Moulton, K. R.; Auclair, J. R.; Zhou, Z. S. Mildly Acidic Conditions Eliminate Deamidation Artifact During Proteolysis: Digestion with Endoprotease Glu-C at pH 4.5. *Amino Acids* **2016**, 48 (4), 1059–1067.
- (225) Ingrosso, D.; Fowler, A. V.; Bleibaum, J.; Clarke, S. Specificity of Endoprotease Asp-N (*Pseudomonas* Fragi): Cleavage at Glutamyl Residues in Two Proteins. *Biochem Biophys Res Commun* **1989**, 162 (3), 1528–1534.

- (226) Appel, W. Chymotrypsin: Molecular and Catalytic Properties. *Clin Biochem* **1986**, *19* (6), 317–322.
- (227) Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A. M.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; Mathieson, T.; Lemeer, S.; Schnatbaum, K.; Reimer, U.; Wenschuh, H.; Mollenhauer, M.; Slotta-Huspenina, J.; Boese, J.-H.; Bantscheff, M.; Gerstmair, A.; Faerber, F.; Kuster, B. Mass-Spectrometry-Based Draft of the Human Proteome. *Nature* **2014**, *509* (7502), 582–587.
- (228) Guo, X.; Trudgian, D. C.; Lemoff, A.; Yadavalli, S.; Mirzaei, H. Confetti: A Multiprotease Map of the HeLa Proteome for Comprehensive Proteomics. *Mol Cell Proteomics* **2014**, *13* (6), 1573–1584.
- (229) Low, T. Y.; van Heesch, S.; van den Toorn, H.; Giansanti, P.; Cristobal, A.; Toonen, P.; Schafer, S.; Hübner, N.; van Breukelen, B.; Mohammed, S.; Cuppen, E.; Heck, A. J. R.; Guryev, V. Quantitative and Qualitative Proteome Characteristics Extracted from in-Depth Integrated Genomics and Proteomics Analysis. *Cell Rep* **2013**, *5* (5), 1469–1478.
- (230) Peng, M.; Taouatas, N.; Cappadona, S.; van Breukelen, B.; Mohammed, S.; Scholten, A.; Heck, A. J. R. Protease Bias in Absolute Protein Quantitation. *Nat Methods* **2012**, *9* (6), 524–525.
- (231) Porter, W. H.; Cunningham, L. W.; Mitchell, W. M. Studies on the Active Site of Clostripain. The Specific Inactivation by the Chloromethyl Ketone Derived from -N-Tosyl-L-Lysine. *J Biol Chem* **1971**, *246* (24), 7675–7682.
- (232) Huesgen, P. F.; Lange, P. F.; Rogers, L. D.; Solis, N.; Eckhard, U.; Kleifeld, O.; Goulas, T.; Gomis-Rüth, F. X.; Overall, C. M. LysargiNase Mirrors Trypsin for Protein C-Terminal and Methylation-Site Identification. *Nat Methods* **2015**, *12* (1), 55–58.
- (233) Hohmann, L.; Sherwood, C.; Eastham, A.; Peterson, A.; Eng, J. K.; Eddes, J. S.; Shteynberg, D.; Martin, D. B. Proteomic Analyses Using Grifola Frondosa Metalloendoprotease Lys-N. *J Proteome Res* **2009**, *8* (3), 1415–1422.
- (234) Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S. Straightforward Ladder Sequencing of Peptides Using a Lys-N Metalloendopeptidase. *Nat Methods* **2008**, *5* (5), 405–407.
- (235) Rajmakers, R.; Neerincx, P.; Mohammed, S.; Heck, A. J. R. Cleavage Specificities of the Brother and Sister Proteases Lys-C and Lys-N. *Chem Commun (Camb)* **2009**, *46* (46), 8827–8829.
- (236) Fruton, J. S. A History of Pepsin and Related Enzymes. *Q Rev Biol* **2002**, *77* (2), 127–147.
- (237) Northrop, J. H. Crystalline Pepsin: I. Isolation and Tests of Purity. *J Gen Physiol* **1930**, *13* (6), 739–766.
- (238) Northrop, J. H. Crystalline Pepsin: II. General Properties and Experimental Methods. *J Gen Physiol* **1930**, *13* (6), 767–780.
- (239) Northrop, J. H. Crystalline Pepsin. *Science* **1929**, *69* (1796), 580.
- (240) The Nobel Prize in Chemistry 1946. NobelPrize.org. <https://www.nobelprize.org/prizes/chemistry/1946/speedread/> (accessed 2024-05-07).
- (241) Gorman, J. J.; Wallis, T. P.; Pitt, J. J. Protein Disulfide Bond Determination by Mass Spectrometry. *Mass Spectrom Rev* **2002**, *21* (3), 183–216.
- (242) Liu, F.; van Breukelen, B.; Heck, A. J. R. Facilitating Protein Disulfide Mapping by a Combination of Pepsin Digestion, Electron Transfer Higher Energy Dissociation (ETHCD), and a Dedicated Search Algorithm SlinkS. *Mol Cell Proteomics* **2014**, *13* (10), 2776–2786.
- (243) Jones, L. M.; Zhang, H.; Vidavsky, I.; Gross, M. L. Online, High-Pressure Digestion System for Protein Characterization by Hydrogen/Deuterium Exchange and Mass Spectrometry. *Anal. Chem.* **2010**, *82* (4), 1171–1174.
- (244) Kostyukevich, Y.; Acter, T.; Zhrebker, A.; Ahmed, A.; Kim, S.; Nikolaev, E. Hydrogen/Deuterium Exchange in Mass Spectrometry. *Mass Spectrometry Reviews* **2018**, *37* (6), 811–853.
- (245) Ebeling, W.; Hennrich, N.; Klockow, M.; Metz, H.; Orth, H. D.; Lang, H. Proteinase K from Triticachium Album Limber. *Eur J Biochem* **1974**, *47* (1), 91–97.
- (246) Saenger, W. Proteinase K. In *Handbook of Proteolytic Enzymes*; Elsevier, 2013; pp 3240–3242. DOI: 10.1016/b978-0-12-382219-2.00714-6.
- (247) Schopper, S.; Kahraman, A.; Leuenberger, P.; Feng, Y.; Piazza, I.; Müller, O.; Boersema, P. J.; Picotti, P. Measuring Protein Structural Changes on a Proteome-Wide Scale Using Limited Proteolysis-Coupled Mass Spectrometry. *Nat Protoc* **2017**, *12* (11), 2391–2410.
- (248) Udenfriend, S.; Stein, S.; Böhlen, P.; Dairman, W.; Leimgruber, W.; Weigle, M. Fluorescamine: A Reagent for Assay of Amino Acids, Peptides, Proteins, and Primary Amines in the Picomole Range. *Science* **1972**, *178* (4063), 871–872.
- (249) Bantan-Polak, T.; Kassai, M.; Grant, K. B. A Comparison of Fluorescamine and Naphthalene-2,3-Dicarboxaldehyde Fluorogenic Reagents for Microplate-Based Detection of Amino Acids. *Anal Biochem* **2001**, *297* (2), 128–136.
- (250) Wiśniewski, J. R.; Gaugaz, F. Z. Fast and Sensitive Total Protein and Peptide Assays for Proteomic Analysis. *Anal. Chem.* **2015**, *87* (8), 4110–4116.
- (251) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics* **2002**, *1* (5), 376–386.
- (252) Mertins, P.; Udeshi, N. D.; Clauser, K. R.; Mani, D.; Patel, J.; Ong, S.; Jaffe, J. D.; Carr, S. A. iTRAQ Labeling Is Superior to mTRAQ for Quantitative Global Proteomics and Phosphoproteomics. *Molecular & Cellular Proteomics* **2012**, *11* (6), M111.014423.
- (253) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics. *Nat Protoc* **2009**, *4* (4), 484–494.
- (254) Zanivan, S.; Meves, A.; Behrendt, K.; Schoof, E. M.; Neilson, L. J.; Cox, J.; Tang, H. R.; Kalna, G.; van Ree, J. H.; van Deursen, J. M.; Trempus, C. S.; Machesky, L. M.; Lindner, R.; Wickstrom, S. A.; Fassler, R.; Mann, M. In Vivo SILAC-Based Proteomics Reveals Phosphoproteome Changes During Mouse Skin Carcinogenesis. *Cell Reports* **2013**, *3* (2), 552–566.
- (255) Monetti, M.; Nagaraj, N.; Sharma, K.; Mann, M. Large-Scale Phosphosite Quantification in Tissues by a Spike-in SILAC Method. *Nat Methods* **2011**, *8* (8), 655–658.
- (256) Derks, J.; Leduc, A.; Wallmann, G.; Huffman, R. G.; Willetts, M.; Khan, S.; Specht, H.; Ralser, M.; Demichev, V.; Slavov, N. Increasing the Throughput of Sensitive Proteomics by plexDIA. *Nat Biotechnol* **2023**, *41* (1), 50–59.
- (257) Thielert, M.; Itang, E. C.; Ammar, C.; Rosenberger, F. A.; Bludau, I.; Schweizer, L.; Nordmann, T. M.; Skowronek, P.; Wahle, M.; Zeng, W.; Zhou, X.; Brunner, A.; Richter, S.; Levesque, M. P.; Theis, F. J.; Steger, M.; Mann, M. Robust Dimethyl-based Multiplex-DIA Doubles Single-cell Proteome Depth via a Reference Channel. *Molecular Systems Biology* **2023**, *19* (9), e11503.
- (258) Meyer, J. G.; D'Souza, A. K.; Sorensen, D. J.; Rardin, M. J.; Wolfe, A. J.; Gibson, B. W.; Schilling, B. Quantification of Lysine Acetylation and Succinylation Stoichiometry in Proteins Using Mass Spectrometric Data-Independent Acquisitions (SWATH). *J. Am. Soc. Mass Spectrom.* **2016**, *27* (11), 1758–1771.
- (259) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75* (8), 1895–1904.
- (260) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in Saccharomyces Cerevisiae Using Amine-Reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics* **2004**, *3* (12), 1154–1169.
- (261) Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L. <I>N</i>, <i>N</i>-Dimethyl Leucines as Novel Isobaric Tandem Mass Tags for Quantitative Proteomics and Peptidomics. *Anal. Chem.* **2010**, *82* (7), 2817–2825.

- (262) Rauniyar, N.; Yates, J. R. Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *J Proteome Res* **2014**, *13* (12), 5293–5309.
- (263) Sinitcyn, P.; Rudolph, J. D.; Cox, J. Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data. *Annu. Rev. Biomed. Data Sci.* **2018**, *1* (1), 207–234.
- (264) Cheng, L.; Pisitkun, T.; Knepper, M. A.; Hoffert, J. D. Peptide Labeling Using Isobaric Tagging Reagents for Quantitative Phosphoproteomics. *Methods Mol Biol* **2016**, *1355*, 53–70.
- (265) Wu, W. W.; Wang, G.; Insel, P. A.; Hsiao, C.-T.; Zou, S.; Martin, B.; Maudsley, S.; Shen, R.-F. Discovery- and Target-Based Protein Quantification Using iTRAQ and Pulsed Q Collision Induced Dissociation (PQD). *J Proteomics* **2012**, *75* (8), 2480–2487.
- (266) Orsburn, B. C.; Yuan, Y.; Bumpus, N. N. Insights into Protein Post-Translational Modification Landscapes of Individual Human Cells by Trapped Ion Mobility Time-of-Flight Mass Spectrometry. *Nat Commun* **2022**, *13* (1), 7246.
- (267) Zecha, J.; Satpathy, S.; Kanashova, T.; Avanesian, S. C.; Kane, M. H.; Clauser, K. R.; Mertins, P.; Carr, S. A.; Kuster, B. TMT Labeling for the Masses: A Robust and Cost-Efficient, In-Solution Labeling Approach. *Mol Cell Proteomics* **2019**, *18* (7), 1468–1478.
- (268) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in *Saccharomyces Cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Mol Cell Proteomics* **2004**, *3* (12), 1154–1169.
- (269) Aggarwal, S.; Yadav, A. K. Dissecting the iTRAQ Data Analysis. In *Methods in Molecular Biology*; Springer: New York, 2016; pp 277–291. DOI: 10.1007/978-1-4939-3106-4_18.
- (270) Pichler, P.; Köcher, T.; Holzmann, J.; Mazanek, M.; Taus, T.; Ammerer, G.; Mechtler, K. Peptide Labeling with Isobaric Tags Yields Higher Identification Rates Using iTRAQ 4-Plex Compared to TMT 6-Plex and iTRAQ 8-Plex on LTQ Orbitrap. *Anal Chem* **2010**, *82* (15), 6549–6558.
- (271) Pottiez, G.; Wiederin, J.; Fox, H. S.; Ciborowski, P. Comparison of 4-Plex to 8-Plex iTRAQ Quantitative Measurements of Proteins in Human Plasma Samples. *J Proteome Res* **2012**, *11* (7), 3774–3781.
- (272) Palmese, A.; De Rosa, C.; Chiappetta, G.; Marino, G.; Amoresano, A. Novel Method to Investigate Protein Carbonylation by iTRAQ Strategy. *Anal Bioanal Chem* **2012**, *404* (6–7), 1631–1635.
- (273) Nyström, T. Role of Oxidative Carbonylation in Protein Quality Control and Senescence. *EMBO J* **2005**, *24* (7), 1311–1317.
- (274) Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; Berfelde, A.; Martinez-Pinna, R.; Farztdinov, V.; Jung, S.; Kuhn, K.; Pike, I. TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Anal. Chem.* **2019**, *91* (24), 15941–15950.
- (275) Li, J.; Cai, Z.; Bomgarden, R. D.; Pike, I.; Kuhn, K.; Rogers, J. C.; Roberts, T. M.; Gygi, S. P.; Paulo, J. A. TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing. *J. Proteome Res.* **2021**, *20* (5), 2964–2972.
- (276) He, Y.; Shishkova, E.; Peters-Clarke, T. M.; Brademan, D. R.; Westphall, M. S.; Bergen, D.; Huang, J.; Huguet, R.; Senko, M. W.; Zabrouskov, V.; McAlister, G. C.; Coon, J. J. Evaluation of the Orbitrap Ascend Tribrid Mass Spectrometer for Shotgun Proteomics. *Anal. Chem.* **2023**, *95* (28), 10655–10663.
- (277) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 Eliminates Ratio Distortion in Isobaric Multiplexed Quantitative Proteomics. *Nat Methods* **2011**, *8* (11), 937–940.
- (278) McAlister, G. C.; Nusinow, D. P.; Jedrychowski, M. P.; Wühr, M.; Huttlin, E. L.; Erickson, B. K.; Rad, R.; Haas, W.; Gygi, S. P. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression Across Cancer Cell Line Proteomes. *Anal. Chem.* **2014**, *86* (14), 7150–7158.
- (279) Lee, K. W.; Peters-Clarke, T. M.; Mertz, K. L.; McAlister, G. C.; Syka, J. E. P.; Westphall, M. S.; Coon, J. J. Infrared Photoactivation Boosts Reporter Ion Yield in Isobaric Tagging. *Anal. Chem.* **2022**, *94* (7), 3328–3334.
- (280) Schweppe, D. K.; Prasad, S.; Belford, M. W.; Navarrete-Perea, J.; Bailey, D. J.; Huguet, R.; Jedrychowski, M. P.; Rad, R.; McAlister, G.; Abbatiello, S. E.; Wouters, E. R.; Zabrouskov, V.; Dunyach, J.-J.; Paulo, J. A.; Gygi, S. P. Characterization and Optimization of Multiplexed Quantitative Analyses Using High-Field Asymmetric-Waveform Ion Mobility Mass Spectrometry. *Analytical Chemistry* **2019**, *91* (6), 4010–4016.
- (281) Pan, K.-T.; Chen, Y.-Y.; Pu, T.-H.; Chao, Y.-S.; Yang, C.-Y.; Bomgarden, R. D.; Rogers, J. C.; Meng, T.-C.; Khoo, K.-H. Mass Spectrometry-Based Quantitative Proteomics for Dissecting Multiplexed Redox Cysteine Modifications in Nitric Oxide-Protected Cardiomyocyte Under Hypoxia. *Antioxid Redox Signal* **2014**, *20* (9), 1365–1381.
- (282) Qu, Z.; Meng, F.; Bomgarden, R. D.; Viner, R. I.; Li, J.; Rogers, J. C.; Cheng, J.; Greenlief, C. M.; Cui, J.; Lubahn, D. B.; Sun, G. Y.; Gu, Z. Proteomic Quantification and Site-Mapping of S-Nitrosylated Proteins Using Isobaric iodoTMT Reagents. *J Proteome Res* **2014**, *13* (7), 3200–3211.
- (283) Hahne, H.; Neubert, P.; Kuhn, K.; Etienne, C.; Bomgarden, R.; Rogers, J. C.; Kuster, B. Carbonyl-Reactive Tandem Mass Tags for the Proteome-Wide Quantification of N-Linked Glycans. *Anal Chem* **2012**, *84* (8), 3716–3724.
- (284) Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L. N,N-Dimethyl Leucines as Novel Isobaric Tandem Mass Tags for Quantitative Proteomics and Peptidomics. *Anal Chem* **2010**, *82* (7), 2817–2825.
- (285) Chen, Z.; Wang, Q.; Lin, L.; Tang, Q.; Edwards, J. L.; Li, S.; Liu, S. Comparative Evaluation of Two Isobaric Labeling Tags, DiART and iTRAQ. *Anal Chem* **2012**, *84* (6), 2908–2915.
- (286) Zhang, J.; Wang, Y.; Li, S. Deuterium Isobaric Amine-Reactive Tags for Quantitative Proteomics. *Anal Chem* **2010**, *82* (18), 7588–7595.
- (287) Zhang, Y.; Askenazi, M.; Jiang, J.; Luckey, C. J.; Griffin, J. D.; Marto, J. A. A Robust Error Model for iTRAQ Quantification Reveals Divergent Signaling Between Oncogenic FLT3 Mutants in Acute Myeloid Leukemia. *Mol Cell Proteomics* **2010**, *9* (5), 780–790.
- (288) Ramsabramaniam, N.; Tao, F.; Li, S.; Marten, M. R. Cost-Effective Isobaric Tagging for Quantitative Phosphoproteomics Using DiART Reagents. *Mol Biosyst* **2013**, *9* (12), 2981–2987.
- (289) Dephoure, N.; Gygi, S. P. Hyperplexing: A Method for Higher-Order Multiplexed Quantitative Proteomics Provides a Map of the Dynamic Response to Rapamycin in Yeast. *Sci. Signal.* **2012**, *5* (217), rs2.
- (290) Liang, Y.; Truong, T.; Saxton, A. J.; Boekweg, H.; Payne, S. H.; Van Ry, P. M.; Kelly, R. T. HyperSCP: Combining Isotopic and Isobaric Labeling for Higher Throughput Single-Cell Proteomics. *Anal. Chem.* **2023**, *95* (20), 8020–8027.
- (291) Aggarwal, S.; Talukdar, N. C.; Yadav, A. K. Advances in Higher Order Multiplexing Techniques in Proteomics. *J Proteome Res* **2019**, *18* (6), 2360–2369.
- (292) Aggarwal, S.; Kumar, A.; Jamwal, S.; Midha, M. K.; Talukdar, N. C.; Yadav, A. K. HyperQuant-A Computational Pipeline for Higher Order Multiplexed Quantitative Proteomics. *ACS Omega* **2020**, *5* (19), 10857–10867.
- (293) Bowser, B. L.; Robinson, R. A. S. Enhanced Multiplexing Technology for Proteomics. *Annu Rev Anal Chem (Palo Alto Calif)* **2023**, *16* (1), 379–400.
- (294) Jayapal, K. P.; Sui, S.; Philp, R. J.; Kok, Y.-J.; Yap, M. G. S.; Griffin, T. J.; Hu, W.-S. Multitagging Proteomic Strategy to Estimate Protein Turnover Rates in Dynamic Systems. *J. Proteome Res.* **2010**, *9* (5), 2087–2097.
- (295) Gu, L.; Evans, A. R.; Robinson, R. A. S. Sample Multiplexing with Cysteine-Selective Approaches: cysDML and cPILOT. *J Am Soc Mass Spectrom* **2015**, *26* (4), 615–630.
- (296) Evans, A. R.; Gu, L.; Guerrero, R.; Robinson, R. A. S. Global cPILOT Analysis of the APP/PS-1 Mouse Liver Proteome. *Proteomics Clin Appl* **2015**, *9* (9–10), 872–884.

- (297) Evans, A. R.; Robinson, R. A. S. Global Combined Precursor Isotopic Labeling and Isobaric Tagging (cPILOT) Approach with Selective MS(3) Acquisition. *Proteomics* **2013**, *13* (22), 3267–3272.
- (298) Gu, L.; Robinson, R. A. S. High-Throughput Endogenous Measurement of S-Nitrosylation in Alzheimer's Disease Using Oxidized Cysteine-Selective cPILOT. *Analyst* **2016**, *141* (12), 3904–3915.
- (299) Dephoure, N.; Gygi, S. P. Hyperplexing: A Method for Higher-Order Multiplexed Quantitative Proteomics Provides a Map of the Dynamic Response to Rapamycin in Yeast. *Sci Signal* **2012**, *5* (217), rs2.
- (300) Schlage, P.; Kockmann, T.; Kizhakkedathu, J. N.; auf dem Keller, U. Monitoring Matrix Metalloproteinase Activity at the Epidermal-Dermal Interface by SILAC-iTRAQ-TAILS. *Proteomics* **2015**, *15* (14), 2491–2502.
- (301) Welle, K. A.; Zhang, T.; Hryhorenko, J. R.; Shen, S.; Qu, J.; Ghaemmaghami, S. Time-Resolved Analysis of Proteome Dynamics by Tandem Mass Tags and Stable Isotope Labeling in Cell Culture (TMT-SILAC) Hyperplexing. *Mol Cell Proteomics* **2016**, *15* (12), 3551–3563.
- (302) Kumar, A.; Jamwal, S.; Midha, M. K.; Hamza, B.; Aggarwal, S.; Yadav, A. K.; Rao, K. V. S. Dataset Generated Using Hyperplexing and Click Chemistry to Monitor Temporal Dynamics of Newly Synthesized Macrophage Secretome Post Infection by Mycobacterial Strains. *Data Brief* **2016**, *9*, 349–354.
- (303) Rothenberg, D. A.; Taliaferro, J. M.; Huber, S. M.; Begley, T. J.; Dedon, P. C.; White, F. M. A Proteomics Approach to Profiling the Temporal Translational Response to Stress and Growth. *iScience* **2018**, *9*, 367–381.
- (304) Potel, C. M.; Lin, M.-H.; Heck, A. J. R.; Lemeer, S. Widespread Bacterial Protein Histidine Phosphorylation Revealed by Mass Spectrometry-Based Proteomics. *Nat Methods* **2018**, *15* (3), 187–190.
- (305) Kleinnijenhuis, A. J.; Kjeldsen, F.; Kallipolitis, B.; Haselmann, K. F.; Jensen, O. N. Analysis of Histidine Phosphorylation Using Tandem MS and Ion–Electron Reactions. *Anal. Chem.* **2007**, *79* (19), 7450–7456.
- (306) Leijten, N. M.; Heck, A. J. R.; Lemeer, S. Histidine Phosphorylation in Human Cells; a Needle or Phantom in the Haystack? *Nat Methods* **2022**, *19* (7), 827–828.
- (307) Schmidt, A.; Ammerer, G.; Mechtler, K. Studying the Fragmentation Behavior of Peptides with Arginine Phosphorylation and Its Influence on Phospho-site Localization. *Proteomics* **2013**, *13* (6), 945–954.
- (308) Salomon, A. R.; Ficarro, S. B.; Brill, L. M.; Brinker, A.; Phung, Q. T.; Ericson, C.; Sauer, K.; Brock, A.; Horn, D. M.; Schultz, P. G.; Peters, E. C. Profiling of Tyrosine Phosphorylation Pathways in Human Cells Using Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (2), 443–448.
- (309) Stopfer, L. E.; Flower, C. T.; Gajadhar, A. S.; Patel, B.; Gallien, S.; Lopez-Ferrer, D.; White, F. M. High-Density, Targeted Monitoring of Tyrosine Phosphorylation Reveals Activated Signaling Networks in Human Tumors. *Cancer Research* **2021**, *81* (9), 2495–2509.
- (310) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **2006**, *127* (3), 635–648.
- (311) Keck, F.; Ataey, P.; Amaya, M.; Bailey, C.; Narayanan, A. Phosphorylation of Single Stranded RNA Virus Proteins and Potential for Novel Therapeutic Strategies. *Viruses* **2015**, *7* (10), 5257–5273.
- (312) Ochoa, D.; Jarnuczak, A. F.; Viéitez, C.; Gehre, M.; Soucheray, M.; Mateus, A.; Kleefeldt, A. A.; Hill, A.; Garcia-Alonso, L.; Stein, F.; Krogan, N. J.; Savitski, M. M.; Swaney, D. L.; Vizcaíno, J. A.; Noh, K.-M.; Beltrao, P. The Functional Landscape of the Human Phosphoproteome. *Nat Biotechnol* **2020**, *38* (3), 365–373.
- (313) Savage, S. R.; Zhang, B. Using Phosphoproteomics Data to Understand Cellular Signaling: A Comprehensive Guide to Bioinformatics Resources. *Clin Proteomics* **2020**, *17*, 27.
- (314) Urban, J. A Review on Recent Trends in the Phosphoproteomics Workflow. From Sample Preparation to Data Analysis. *Anal Chim Acta* **2022**, *1199*, No. 338857.
- (315) Cohen, P. The Role of Protein Phosphorylation in Human Health and Disease. The Sir Hans Krebs Medal Lecture. *Eur J Biochem* **2001**, *268* (19), 5001–5010.
- (316) Zhao, Y.; Jensen, O. N. Modification-Specific Proteomics: Strategies for Characterization of Post-Translational Modifications Using Enrichment Techniques. *Proteomics* **2009**, *9* (20), 4632–4641.
- (317) Iliuk, A. B.; Arrington, J. V.; Tao, W. A. Analytical Challenges Translating Mass Spectrometry-Based Phosphoproteomics from Discovery to Clinical Applications. *Electrophoresis* **2014**, *35* (24), 3430–3440.
- (318) Hsu, C.-C.; Xue, L.; Arrington, J. V.; Wang, P.; Paez Paez, J. S.; Zhou, Y.; Zhu, J.-K.; Tao, W. A. Estimating the Efficiency of Phosphopeptide Identification by Tandem Mass Spectrometry. *J Am Soc Mass Spectrom* **2017**, *28* (6), 1127–1135.
- (319) Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Van Eyk, J. E. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-up Proteomics Workflow. *Curr Protoc Mol Biol* **2010**, Unit10.25.
- (320) Potel, C. M.; Lemeer, S.; Heck, A. J. R. Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Anal Chem* **2019**, *91* (1), 126–141.
- (321) Fila, J.; Honys, D. Enrichment Techniques Employed in Phosphoproteomics. *Amino Acids* **2012**, *43* (3), 1025–1047.
- (322) Rex, D. A. B.; Subbannayya, Y.; Modi, P. K.; Palollathil, A.; Gopalakrishnan, L.; Bhandary, Y. P.; Prasad, T. S. K.; Pinto, S. M. Temporal Quantitative Phosphoproteomics Profiling of Interleukin-33 Signaling Network Reveals Unique Modulators of Monocyte Activation. *Cells* **2022**, *11* (1), 138.
- (323) Posewitz, M. C.; Tempst, P. Immobilized Gallium(III) Affinity Chromatography of Phosphopeptides. *Anal Chem* **1999**, *71* (14), 2883–2892.
- (324) Ojida, A.; Inoue, M.; Mito-oka, Y.; Tsutsumi, H.; Sada, K.; Hamachi, I. Effective Disruption of Phosphoprotein-Protein Surface Interaction Using Zn(II) Dicolylamine-Based Artificial Receptors via Two-Point Interaction. *J Am Chem Soc* **2006**, *128* (6), 2052–2058.
- (325) Andersson, L.; Porath, J. Isolation of Phosphoproteins by Immobilized Metal (Fe³⁺) Affinity Chromatography. *Anal Biochem* **1986**, *154* (1), 250–254.
- (326) Machida, M.; Kosako, H.; Shirakabe, K.; Kobayashi, M.; Ushiyama, M.; Inagawa, J.; Hirano, J.; Nakano, T.; Bando, Y.; Nishida, E.; Hattori, S. Purification of Phosphoproteins by Immobilized Metal Affinity Chromatography and Its Application to Phosphoproteome Analysis. *FEBS J* **2007**, *274* (6), 1576–1587.
- (327) Dubrovskaya, A.; Souchevnytskyi, S. Efficient Enrichment of Intact Phosphorylated Proteins by Modified Immobilized Metal-Affinity Chromatography. *Proteomics* **2005**, *5* (18), 4678–4683.
- (328) Zhou, S.; Bailey, M. J.; Dunn, M. J.; Preedy, V. R.; Emery, P. W. A Quantitative Investigation into the Losses of Proteins at Different Stages of a Two-Dimensional Gel Electrophoresis Procedure. *Proteomics* **2005**, *5* (11), 2739–2747.
- (329) Batth, T. S.; Francavilla, C.; Olsen, J. V. Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *J. Proteome Res.* **2014**, *13* (12), 6176–6186.
- (330) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jørgensen, T. J. D. Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. *Mol Cell Proteomics* **2005**, *4* (7), 873–886.
- (331) Thingholm, T. E.; Larsen, M. R. Phosphopeptide Enrichment by Immobilized Metal Affinity Chromatography. *Methods Mol Biol* **2016**, *1355*, 123–133.
- (332) Carr, S. A.; Huddleston, M. J.; Annan, R. S. Selective Detection and Sequencing of Phosphopeptides at the Femtomole Level by Mass Spectrometry. *Analytical Biochemistry* **1996**, *239* (2), 180–192.

- (333) Ruprecht, B.; Koch, H.; Medard, G.; Mundt, M.; Kuster, B.; Lemeer, S. Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns. *Mol Cell Proteomics* **2015**, *14* (1), 205–215.
- (334) Ruprecht, B.; Koch, H.; Domasinska, P.; Frejno, M.; Kuster, B.; Lemeer, S. Optimized Enrichment of Phosphoproteomes by Fe-IMAC Column Chromatography. *Methods Mol Biol* **2017**, *1550*, 47–60.
- (335) Lai, A. C.-Y.; Tsai, C.-F.; Hsu, C.-C.; Sun, Y.-N.; Chen, Y.-J. Complementary Fe(3+)- and Ti(4+)-Immobilized Metal Ion Affinity Chromatography for Purification of Acidic and Basic Phosphopeptides. *Rapid Commun Mass Spectrom* **2012**, *26* (18), 2186–2194.
- (336) Choi, S.; Kim, J.; Cho, K.; Park, G.; Yoon, J. H.; Park, S.; Yoo, J. S.; Ryu, S. H.; Kim, Y. H.; Kim, J. Sequential Fe3O4/TiO2 Enrichment for Phosphopeptide Analysis by Liquid Chromatography/Tandem Mass Spectrometry. *Rapid Commun Mass Spectrom* **2010**, *24* (10), 1467–1474.
- (337) Tong, J.; Cao, B.; Martyn, G. D.; Krieger, J. R.; Taylor, P.; Yates, B.; Sidhu, S. S.; Li, S. S. C.; Mao, X.; Moran, M. F. Protein-Phosphotyrosine Proteome Profiling by Superbinder-SH2 Domain Affinity Purification Mass Spectrometry, sSH2-AP-MS. *Proteomics* **2017**, *17* (6), 1600360.
- (338) Martyn, G. D.; Veggiani, G.; Sidhu, S. S. Engineering SH2 Domains with Tailored Specificities and Affinities. *Methods Mol Biol* **2023**, *2705*, 307–348.
- (339) Li, J.; Wang, J.; Yan, Y.; Li, N.; Qing, X.; Tuerxun, A.; Guo, X.; Chen, X.; Yang, F. Comprehensive Evaluation of Different TiO. *Cells* **2022**, *11* (13), 2047.
- (340) Spiro, R. G. Protein Glycosylation: Nature, Distribution, Enzymatic Formation, and Disease Implications of Glycopeptide Bonds. *Glycobiology* **2002**, *12* (4), 43R–56R.
- (341) Riley, N. M.; Bertozzi, C. R.; Pitteri, S. J. A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry–Based Glycoproteomics. *Molecular & Cellular Proteomics* **2021**, *20*, 100029.
- (342) Bermudez, A.; Pitteri, S. J. Enrichment of Intact Glycopeptides Using Strong Anion Exchange and Electrostatic Repulsion Hydrophilic Interaction Chromatography. In *Methods in Molecular Biology*; Springer, 2021; pp 107–120. DOI: 10.1007/978-1-0716-1241-5_8.
- (343) Totten, S. M.; Feasley, C. L.; Bermudez, A.; Pitteri, S. J. Parallel Comparison of N-Linked Glycopeptide Enrichment Techniques Reveals Extensive Glycoproteomic Analysis of Plasma Enabled by SAX-ERLIC. *J. Proteome Res.* **2017**, *16* (3), 1249–1260.
- (344) Wohlgemuth, J.; Karas, M.; Eichhorn, T.; Hendriks, R.; Andrecht, S. Quantitative Site-Specific Analysis of Protein Glycosylation by LC-MS Using Different Glycopeptide-Enrichment Strategies. *Analytical Biochemistry* **2009**, *395* (2), 178–188.
- (345) Thaysen-Andersen, M.; Packer, N. H. Advances in LC–MS/MS-Based Glycoproteomics: Getting Closer to System-Wide Site-Specific Mapping of the N- and O-Glycoproteome. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2014**, *1844* (9), 1437–1452.
- (346) Thaysen-Andersen, M.; Larsen, M. R.; Packer, N. H.; Palmisano, G. Structural Analysis of Glycoprotein Sialylation – Part I: Pre-LC-MS Analytical Strategies. *RSC Adv.* **2013**, *3* (45), 22683.
- (347) Gaunitz, S.; Nagy, G.; Pohl, N. L. B.; Novotny, M. V. Recent Advances in the Analysis of Complex Glycoproteins. *Anal. Chem.* **2017**, *89* (1), 389–413.
- (348) Ghosh, R.; Gilda, J. E.; Gomes, A. V. The Necessity of and Strategies for Improving Confidence in the Accuracy of Western Blots. *Expert Rev Proteomics* **2014**, *11* (5), 549–560.
- (349) Zhao, Y.; Muir, T. W.; Kent, S. B.; Fischer, E.; Scardina, J. M.; Chait, B. T. Mapping Protein-Protein Interactions by Affinity-Directed Mass Spectrometry. *Proc Natl Acad Sci U S A* **1996**, *93* (9), 4020–4024.
- (350) Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X.-M.; Polakiewicz, R. D.; Comb, M. J. Immunoaffinity Profiling of Tyrosine Phosphorylation in Cancer Cells. *Nat Biotechnol* **2005**, *23* (1), 94–101.
- (351) Matsuoka, S.; Ballif, B. A.; Smogorzewska, A.; McDonald, E. R.; Hurov, K. E.; Luo, J.; Bakalarski, C. E.; Zhao, Z.; Solimini, N.; Lerenthal, Y.; Shiloh, Y.; Gygi, S. P.; Elledge, S. J. ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage. *Science* **2007**, *316* (5828), 1160–1166.
- (352) Moritz, A.; Li, Y.; Guo, A.; Villén, J.; Wang, Y.; MacNeill, J.; Kornhauser, J.; Sprott, K.; Zhou, J.; Possemato, A.; Ren, J. M.; Hornbeck, P.; Cantley, L. C.; Gygi, S. P.; Rush, J.; Comb, M. J. Akt-RSK-S6 Kinase Signaling Networks Activated by Oncogenic Receptor Tyrosine Kinases. *Sci Signal* **2010**, *3* (136), ra64.
- (353) Emanuele, M. J.; Elia, A. E. H.; Xu, Q.; Thoma, C. R.; Izhar, L.; Leng, Y.; Guo, A.; Chen, Y.-N.; Rush, J.; Hsu, P. W.-C.; Yen, H.-C. S.; Elledge, S. J. Global Identification of Modular Cullin-RING Ligase Substrates. *Cell* **2011**, *147* (2), 459–474.
- (354) Kim, W.; Bennett, E. J.; Huttlin, E. L.; Guo, A.; Li, J.; Possemato, A.; Sowa, M. E.; Rad, R.; Rush, J.; Comb, M. J.; Harper, J. W.; Gygi, S. P. Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. *Mol Cell* **2011**, *44* (2), 325–340.
- (355) Wagner, S. A.; Beli, P.; Weinert, B. T.; Nielsen, M. L.; Cox, J.; Mann, M.; Choudhary, C. A Proteome-Wide, Quantitative Survey of in Vivo Ubiquitylation Sites Reveals Widespread Regulatory Roles. *Mol Cell Proteomics* **2011**, *10* (10), No. M111.013284.
- (356) Choudhary, C.; Kumar, C.; Gnäd, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **2009**, *325* (5942), 834–840.
- (357) Hebert, A. S.; Dittenhafer-Reed, K. E.; Yu, W.; Bailey, D. J.; Selen, E. S.; Boersma, M. D.; Carson, J. J.; Tonelli, M.; Balloon, A. J.; Higbee, A. J.; Westphall, M. S.; Pagliarini, D. J.; Prolla, T. A.; Assadi-Porter, F.; Roy, S.; Denu, J. M.; Coon, J. J. Calorie Restriction and SIRT3 Trigger Global Reprogramming of the Mitochondrial Protein Acetylome. *Mol Cell* **2013**, *49* (1), 186–199.
- (358) Svinikina, T.; Gu, H.; Silva, J. C.; Mertins, P.; Qiao, J.; Fereshetian, S.; Jaffe, J. D.; Kuhn, E.; Udeshi, N. D.; Carr, S. A. Deep, Quantitative Coverage of the Lysine Acetylome Using Novel Anti-Acetyl-Lysine Antibodies and an Optimized Proteomic Workflow. *Mol Cell Proteomics* **2015**, *14* (9), 2429–2440.
- (359) Schilling, B.; Meyer, J. G.; Wei, L.; Ott, M.; Verdin, E. High-Resolution Mass Spectrometry to Identify and Quantify Acetylation Protein Targets. *Methods Mol Biol* **2019**, *1983*, 3–16.
- (360) Basisty, N.; Meyer, J. G.; Wei, L.; Gibson, B. W.; Schilling, B. Simultaneous Quantification of the Acetylome and Succinylome by ‘One-Pot’ Affinity Enrichment. *Proteomics* **2018**, *18* (17), 1800123.
- (361) Guo, A.; Gu, H.; Zhou, J.; Mulhern, D.; Wang, Y.; Lee, K. A.; Yang, V.; Aguiar, M.; Kornhauser, J.; Jia, X.; Ren, J.; Beausoleil, S. A.; Silva, J. C.; Vemulapalli, V.; Bedford, M. T.; Comb, M. J. Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. *Mol Cell Proteomics* **2014**, *13* (1), 372–387.
- (362) Zhao, Y.; Zhang, Y.; Sun, H.; Maroto, R.; Brasier, A. R. Selective Affinity Enrichment of Nitrotyrosine-Containing Peptides for Quantitative Analysis in Complex Samples. *J. Proteome Res.* **2017**, *16* (8), 2983–2992.
- (363) Boersema, P. J.; Foong, L. Y.; Ding, V. M. Y.; Lemeer, S.; van Breukelen, B.; Philp, R.; Boekhorst, J.; Snel, B.; den Hertog, J.; Choo, A. B. H.; Heck, A. J. R. In-Depth Qualitative and Quantitative Profiling of Tyrosine Phosphorylation Using a Combination of Phosphopeptide Immunoaffinity Purification and Stable Isotope Dimethyl Labeling. *Mol Cell Proteomics* **2010**, *9* (1), 84–99.
- (364) Zhang, Y.; Wolf-Yadlin, A.; Ross, P. L.; Pappin, D. J.; Rush, J.; Lauffenburger, D. A.; White, F. M. Time-Resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules. *Mol Cell Proteomics* **2005**, *4* (9), 1240–1250.
- (365) Lu, Q.; Zhang, X.; Liang, T.; Bai, X. O-GlcNAcylation: An Important Post-Translational Modification and a Potential Therapeutic Target for Cancer Therapy. *Mol Med* **2022**, *28* (1), 115.
- (366) Burt, R. A.; Dejanovic, B.; Peckham, H. J.; Lee, K. A.; Li, X.; Unadajela, J. R.; Rao, A.; Malaker, S. A.; Carr, S. A.; Myers, S. A. Novel Antibodies for the Simple and Efficient Enrichment of Native

- O-GlcNAc Modified Peptides. *Mol Cell Proteomics* **2021**, *20*, No. 100167.
- (367) Fu, Q.; Kowalski, M. P.; Mastali, M.; Parker, S. J.; Sobhani, K.; van den Broek, L.; Hunter, C. L.; Van Eyk, J. E. Highly Reproducible Automated Proteomics Sample Preparation Workflow for Quantitative Mass Spectrometry. *J. Proteome Res.* **2018**, *17* (1), 420–428.
- (368) Overmyer, K. A.; Shishkova, E.; Miller, I. J.; Balnis, J.; Bernstein, M. N.; Peters-Clarke, T. M.; Meyer, J. G.; Quan, Q.; Muehlbauer, L. K.; Trujillo, E. A.; He, Y.; Chopra, A.; Chieng, H. C.; Tiwari, A.; Judson, M. A.; Paulson, B.; Brademan, D. R.; Zhu, Y.; Serrano, L. R.; Linke, V.; Drake, L. A.; Adam, A. P.; Schwartz, B. S.; Singer, H. A.; Swanson, S.; Mosher, D. F.; Stewart, R.; Coon, J. J.; Jaitovich, A. Large-Scale Multi-Omic Analysis of COVID-19 Severity. *Cell Systems* **2021**, *12* (1), 23–40.e7.
- (369) Tu, C.; Rudnick, P. A.; Martinez, M. Y.; Cheek, K. L.; Stein, S. E.; Slebos, R. J. C.; Liebler, D. C. Depletion of Abundant Plasma Proteins and Limitations of Plasma Proteomics. *J. Proteome Res.* **2010**, *9* (10), 4982–4991.
- (370) Zolotarjova, N.; Martosella, J.; Nicol, G.; Bailey, J.; Boyes, B. E.; Barrett, W. C. Differences Among Techniques for High-Abundant Protein Depletion. *Proteomics* **2005**, *5* (13), 3304–3313.
- (371) Denizli, A.; Pişkin, E. Dye-Ligand Affinity Systems. *J. Biochem Biophys Methods* **2001**, *49* (1–3), 391–416.
- (372) Andaç, M. Cibacron Blue Immobilized Poly(glycidyl-Methacrylate) Nanobeads for Albumin Removal in Proteome Studies. *Artif Cells Nanomed Biotechnol* **2015**, *43* (2), 133–139.
- (373) Sjöbring, U.; Björck, L.; Kastern, W. Protein G Genes: Structure and Distribution of IgG-Binding and Albumin-Binding Domains. *Mol Microbiol* **1989**, *3* (3), 319–327.
- (374) Moks, T.; Abrahamsén, L.; Nilsson, B.; Hellman, U.; Sjöquist, J.; Uhlén, M. Staphylococcal Protein A Consists of Five IgG-Binding Domains. *Eur J Biochem* **1986**, *156* (3), 637–643.
- (375) Graille, M.; Stura, E. A.; Corper, A. L.; Sutton, B. J.; Taussig, M. J.; Charbonnier, J. B.; Silverman, G. J. Crystal Structure of a Staphylococcus Aureus Protein A Domain Complexed with the Fab Fragment of a Human IgM Antibody: Structural Basis for Recognition of B-Cell Receptors and Superantigen Activity. *Proc Natl Acad Sci U S A* **2000**, *97* (10), 5399–5404.
- (376) Lee, P. Y.; Osman, J.; Low, T. Y.; Jamal, R. Plasma/Serum Proteomics: Depletion Strategies for Reducing High-Abundance Proteins for Biomarker Discovery. *Bioanalysis* **2019**, *11* (19), 1799–1812.
- (377) Garcia, S.; Baldasso, P. A.; Guest, P. C.; Martins-de-Souza, D. Depletion of Highly Abundant Proteins of the Human Blood Plasma: Applications in Proteomics Studies of Psychiatric Disorders. *Methods Mol Biol* **2017**, *1546*, 195–204.
- (378) Guerrier, L.; Righetti, P. G.; Boschetti, E. Reduction of Dynamic Protein Concentration Range of Biological Extracts for the Discovery of Low-Abundance Proteins by Means of Hexapeptide Ligand Library. *Nat Protoc* **2008**, *3* (5), 883–890.
- (379) Righetti, P. G.; Boschetti, E. Sample Treatment Methods Involving Combinatorial Peptide Ligand Libraries for Improved Proteomes Analyses. *Methods Mol Biol* **2015**, *1243*, 55–82.
- (380) Mahn, A.; Ismail, M. Depletion of Highly Abundant Proteins in Blood Plasma by Ammonium Sulfate Precipitation for 2D-PAGE Analysis. *J. Chromatogr B Analyt Technol Biomed Life Sci* **2011**, *879* (30), 3645–3648.
- (381) Liu, G.; Zhao, Y.; Angeles, A.; Hamuro, L. L.; Arnold, M. E.; Shen, J. X. A Novel and Cost Effective Method of Removing Excess Albumin from Plasma/Serum Samples and Its Impacts on LC-MS/MS Bioanalysis of Therapeutic Proteins. *Anal Chem* **2014**, *86* (16), 8336–8343.
- (382) Liu, Z.; Fan, S.; Liu, H.; Yu, J.; Qiao, R.; Zhou, M.; Yang, Y.; Zhou, J.; Xie, P. Enhanced Detection of Low-Abundance Human Plasma Proteins by Integrating Polyethylene Glycol Fractionation and Immunoaffinity Depletion. *PLoS One* **2016**, *11* (11), e0166306.
- (383) Warder, S. E.; Tucker, L. A.; Strelitzer, T. J.; McKeegan, E. M.; Meuth, J. L.; Jung, P. M.; Saraf, A.; Singh, B.; Lai-Zhang, J.; Gagne, G.; Rogers, J. C. Reducing Agent-Mediated Precipitation of High-Abundance Plasma Proteins. *Anal Biochem* **2009**, *387* (2), 184–193.
- (384) Göktürk, I.; Tamahkar, E.; Yilmaz, F.; Denizli, A. Protein Depletion with Bacterial Cellulose Nanofibers. *J. Chromatogr B Analyt Technol Biomed Life Sci* **2018**, *1099*, 1–9.
- (385) Andac, M.; Galaev, I. Y.; Denizli, A. Molecularly Imprinted Poly(hydroxyethyl Methacrylate) Based Cryogel for Albumin Depletion from Human Serum. *Colloids Surf B Biointerfaces* **2013**, *109*, 259–265.
- (386) Tamahkar, E.; Babaç, C.; Kutsal, T.; Pişkin, E.; Denizli, A. Bacterial Cellulose Nanofibers for Albumin Depletion from Human Serum. *Process Biochemistry* **2010**, *45* (10), 1713–1719.
- (387) Blume, J. E.; Manning, W. C.; Troiano, G.; Hornburg, D.; Figa, M.; Hesterberg, L.; Platt, T. L.; Zhao, X.; Cuaresma, R. A.; Everley, P. A.; Ko, M.; Liou, H.; Mahoney, M.; Ferdosi, S.; Elgierari, E. M.; Stolarczyk, C.; Tangeysh, B.; Xia, H.; Benz, R.; Siddiqui, A.; Carr, S. A.; Ma, P.; Langer, R.; Farias, V.; Farokhzad, O. C. Rapid, Deep and Precise Profiling of the Plasma Proteome with Multi-Nanoparticle Protein Corona. *Nat Commun* **2020**, *11* (1), 3662.
- (388) Huang, T.; Wang, J.; Stukalov, A.; Donovan, M. K. R.; Ferdosi, S.; Williamson, L.; Just, S.; Castro, G.; Cantrell, L. S.; Elgierari, E.; Benz, R. W.; Huang, Y.; Motamedchaboki, K.; Hakimi, A.; Arrey, T.; Damoc, E.; Kreimer, S.; Farokhzad, O. C.; Batzoglou, S.; Siddiqui, A.; Van Eyk, J. E.; Hornburg, D. Protein Coronas on Functionalized Nanoparticles Enable Quantitative and Precise Large-Scale Deep Plasma Proteomics. *bioRxiv*, 2023. DOI: [10.1101/2023.08.28.555225](https://doi.org/10.1101/2023.08.28.555225).
- (389) Wu, C. C.; Tsantilas, K. A.; Park, J.; Plubell, D.; Sanders, J. A.; Naicker, P.; Govender, I.; Buthelezi, S.; Stoychev, S.; Jordaan, J.; Merrihew, G.; Huang, E.; Parker, E. D.; Riffle, M.; Hoofnagle, A. N.; Noble, W. S.; Poston, K. L.; Montine, T. J.; MacCoss, M. J. Mag-Net: Rapid Enrichment of Membrane-Bound Particles Enables High Coverage Quantitative Analysis of the Plasma Proteome *bioRxiv*, 2023. DOI: [10.1101/2023.06.10.544439](https://doi.org/10.1101/2023.06.10.544439).
- (390) Huang, T.; Wang, J.; Stukalov, A.; Donovan, M. K. R.; Ferdosi, S.; Williamson, L.; Just, S.; Castro, G.; Cantrell, L. S.; Elgierari, E.; Benz, R. W.; Huang, Y.; Motamedchaboki, K.; Hakimi, A.; Arrey, T.; Damoc, E.; Kreimer, S.; Farokhzad, O. C.; Batzoglou, S.; Siddiqui, A.; Van Eyk, J. E.; Hornburg, D. Protein Coronas on Functionalized Nanoparticles Enable Quantitative and Precise Large-Scale Deep Plasma Proteomics. *bioRxiv*, 2023. DOI: [10.1101/2023.08.28.555225](https://doi.org/10.1101/2023.08.28.555225).
- (391) Blume, J. E.; Manning, W. C.; Troiano, G.; Hornburg, D.; Figa, M.; Hesterberg, L.; Platt, T. L.; Zhao, X.; Cuaresma, R. A.; Everley, P. A.; Ko, M.; Liou, H.; Mahoney, M.; Ferdosi, S.; Elgierari, E. M.; Stolarczyk, C.; Tangeysh, B.; Xia, H.; Benz, R.; Siddiqui, A.; Carr, S. A.; Ma, P.; Langer, R.; Farias, V.; Farokhzad, O. C. Rapid, Deep and Precise Profiling of the Plasma Proteome with Multi-Nanoparticle Protein Corona. *Nat Commun* **2020**, *11* (1), 3662.
- (392) Shen, Y.; Tolić, N.; Masselon, C.; Pasa-Tolić, L.; Camp, D. G.; Hixson, K. K.; Zhao, R.; Anderson, G. A.; Smith, R. D. Ultrasensitive Proteomics Using High-Efficiency on-Line Micro-SPE-nanoLC-nano-ESI MS and MS/MS. *Anal Chem* **2004**, *76* (1), 144–154.
- (393) Magni, F.; Van Der Burgt, Y. E. M.; Chinello, C.; Mainini, V.; Gianazza, E.; Squeo, V.; Deelder, A. M.; Kienle, M. G. Biomarkers Discovery by Peptide and Protein Profiling in Biological Fluids Based on Functionalized Magnetic Beads Purification and Mass Spectrometry. *Blood Transfus* **2010**, No. Suppl 3, s92–7.
- (394) Peter, J. F.; Otto, A. M. Magnetic Particles as Powerful Purification Tool for High Sensitive Mass Spectrometric Screening Procedures. *Proteomics* **2010**, *10* (4), 628–633.
- (395) Ruhaak, L. R.; Huhn, C.; Koeleman, C. A. M.; Deelder, A. M.; Wührer, M. Robust and High-Throughput Sample Preparation for (Semi-)Quantitative Analysis of N-Glycosylation Profiles from Plasma Samples. *Methods Mol Biol* **2012**, *893*, 371–385.
- (396) Mysling, S.; Palmisano, G.; Højrup, P.; Thaysen-Andersen, M. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Anal Chem* **2010**, *82* (13), 5598–5609.

- (397) Luo, Q.; Yue, G.; Valaskovic, G. A.; Gu, Y.; Wu, S.-L.; Karger, B. L. On-Line 1D and 2D Porous Layer Open Tubular/LC-ESI-MS Using 10-Microm-i.d. Poly(styrene-Divinylbenzene) Columns for Ultrasensitive Proteomic Analysis. *Anal Chem* **2007**, *79* (16), 6174–6181.
- (398) Miyazaki, S.; Morisato, K.; Ishizuka, N.; Minakuchi, H.; Shintani, Y.; Furuno, M.; Nakanishi, K. Development of a Monolithic Silica Extraction Tip for the Analysis of Proteins. *J Chromatogr A* **2004**, *1043* (1), 19–25.
- (399) Bladergroen, M. R.; van der Burgt, Y. E. M. Solid-Phase Extraction Strategies to Surmount Body Fluid Sample Complexity in High-Throughput Mass Spectrometry-Based Proteomics. *J Anal Methods Chem* **2015**, *2015*, No. 250131.
- (400) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Anal. Chem.* **2003**, *75* (3), 663–670.
- (401) Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, Encapsulated Proteomic-Sample Processing Applied to Copy-Number Estimation in Eukaryotic Cells. *Nat Methods* **2014**, *11* (3), 319–324.
- (402) Müller, T.; Kalxdorf, M.; Longuespée, R.; Kazdal, D. N.; Stenzinger, A.; Krijgsvel, J. Automated Sample Preparation with SP3 for Low-Input Clinical Proteomics. *Mol Syst Biol* **2020**, *16* (1), No. e9111.
- (403) Li, T.; Wang, W.; Zhao, H.; He, F.; Zhong, K.; Yuan, S.; Wang, Z. National Continuous Surveys on Internal Quality Control for HbA1c in 306 Clinical Laboratories of China from 2012 to 2016: Continual Improvement. *J Clin Lab Anal* **2017**, *31* (5), e22099.
- (404) Abstracts of Presentations at the Association of Clinical Scientists 143. *Ann Clin Lab Sci* **2022**, *52* (3), 511–525.
- (405) Essader, A. S.; Cargile, B. J.; Bundy, J. L.; Stephenson, J. L. A Comparison of Immobilized pH Gradient Isoelectric Focusing and Strong-Cation-Exchange Chromatography as a First Dimension in Shotgun Proteomics. *Proteomics* **2005**, *5* (1), 24–34.
- (406) Manadas, B.; Mendes, V. M.; English, J.; Dunn, M. J. Peptide Fractionation in Proteomics Approaches. *Expert Rev Proteomics* **2010**, *7* (5), 655–663.
- (407) McCalley, D. V. The Challenges of the Analysis of Basic Compounds by High Performance Liquid Chromatography: Some Possible Approaches for Improved Separations. *J Chromatogr A* **2010**, *1217* (6), 858–880.
- (408) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. Two-Dimensional Separation of Peptides Using RP-RP-HPLC System with Different pH in First and Second Separation Dimensions. *J Sep Sci* **2005**, *28* (14), 1694–1703.
- (409) Yang, F.; Shen, Y.; Camp, D. G.; Smith, R. D. High-pH Reversed-Phase Chromatography with Fraction Concatenation for 2D Proteomic Analysis. *Expert Rev Proteomics* **2012**, *9* (2), 129–134.
- (410) Simpson, R. J.; Moritz, R. L.; Nice, E. E.; Grego, B. A High-Performance Liquid Chromatography Procedure for Recovering Subnanomole Amounts of Protein from SDS-Gel Electroeluates for Gas-Phase Sequence Analysis. *Eur J Biochem* **1987**, *165* (1), 21–29.
- (411) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. Hydrophilic Interaction Liquid Chromatography (HILIC) in Proteomics. *Anal Bioanal Chem* **2008**, *391* (1), 151–159.
- (412) Mizukami, Y.; Iwamatsu, A.; Aki, T.; Kimura, M.; Nakamura, K.; Nao, T.; Okusa, T.; Matsuzaki, M.; Yoshida, K.-I.; Kobayashi, S. ERK1/2 Regulates Intracellular ATP Levels Through Alpha-Enolase Expression in Cardiomyocytes Exposed to Ischemic Hypoxia and Reoxygenation. *J Biol Chem* **2004**, *279* (48), 50120–50131.
- (413) Buszewski, B.; Noga, S. Hydrophilic Interaction Liquid Chromatography (HILIC)—a Powerful Separation Technique. *Anal Bioanal Chem* **2012**, *402* (1), 231–247.
- (414) Jandera, P. Stationary Phases for Hydrophilic Interaction Chromatography, Their Characterization and Implementation into Multidimensional Chromatography Concepts. *J Sep Sci* **2008**, *31* (9), 1421–1437.
- (415) Alpert, A. J. Electrostatic Repulsion Hydrophilic Interaction Chromatography for Isocratic Separation of Charged Solutes and Selective Isolation of Phosphopeptides. *Anal. Chem.* **2008**, *80* (1), 62–76.
- (416) Hao, P.; Guo, T.; Li, X.; Adav, S. S.; Yang, J.; Wei, M.; Sze, S. K. Novel Application of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) in Shotgun Proteomics: Comprehensive Profiling of Rat Kidney Proteome. *J. Proteome Res.* **2010**, *9* (7), 3520–3526.
- (417) Hörth, P.; Miller, C. A.; Preckel, T.; Wenz, C. Efficient Fractionation and Improved Protein Identification by Peptide OFFGEL Electrophoresis. *Mol Cell Proteomics* **2006**, *5* (10), 1968–1974.
- (418) Ernoul, E.; Gamelin, E.; Guette, C. Improved Proteome Coverage by Using iTRAQ Labelling and Peptide OFFGEL Fractionation. *Proteome Sci* **2008**, *6*, 27.
- (419) Fraterman, S.; Zeiger, U.; Khurana, T. S.; Rubinstein, N. A.; Wilm, M. Combination of Peptide OFFGEL Fractionation and Label-Free Quantitation Facilitated Proteomics Profiling of Extraocular Muscle. *Proteomics* **2007**, *7* (18), 3404–3416.
- (420) François, I.; Sandra, K.; Sandra, P. Comprehensive Liquid Chromatography: Fundamental Aspects and Practical Considerations—A Review. *Analytica Chimica Acta* **2009**, *641* (1–2), 14–31.
- (421) Annesley, T. M. Ion Suppression in Mass Spectrometry. *Clinical Chemistry* **2003**, *49* (7), 1041–1044.
- (422) Xie, F.; Smith, R. D.; Shen, Y. Advanced Proteomic Liquid Chromatography. *Journal of Chromatography A* **2012**, *1261*, 78–90.
- (423) Zhang, X.; Fang, A.; Riley, C. P.; Wang, M.; Regnier, F. E.; Buck, C. Multi-Dimensional Liquid Chromatography in Proteomics—A Review. *Analytica Chimica Acta* **2010**, *664* (2), 101–113.
- (424) SHI, Y.; XIANG, R.; HORVATH, C.; WILKINS, J. The Role of Liquid Chromatography in Proteomics. *Journal of Chromatography A* **2004**, *1053* (1–2), 27–36.
- (425) Rayleigh, L. XX. <I>On the Equilibrium of Liquid Conducting Masses Charged with Electricity</i>. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1882**, *14* (87), 184–186.
- (426) Luo, Q.; Tang, K.; Yang, F.; Elias, A.; Shen, Y.; Moore, R. J.; Zhao, R.; Hixson, K. K.; Rossie, S. S.; Smith, R. D. More Sensitive and Quantitative Proteomic Measurements Using Very Low Flow Rate Porous Silica Monolithic LC Columns with Electrospray Ionization-Mass Spectrometry. *J. Proteome Res.* **2006**, *5* (5), 1091–1097.
- (427) van Deemter, J. J.; Zuiderweg, F. J.; Klinkenberg, A. Longitudinal Diffusion and Resistance to Mass Transfer as Causes of Nonideality in Chromatography. *Chemical Engineering Science* **1956**, *5* (6), 271–289.
- (428) Midha, M. K.; Kapil, C.; Maes, M.; Baxter, D. H.; Morrone, S. R.; Prokop, T. J.; Moritz, R. L. Vacuum Insulated Probe Heated Electrospray Ionization Source Enhances Microflow Rate Chromatography Signals in the Bruker timsTOF Mass Spectrometer. *J Proteome Res* **2023**, *22* (7), 2525–2537.
- (429) Kreimer, S.; Haghani, A.; Binek, A.; Hauspurg, A.; Seyedmohammad, S.; Rivas, A.; Momenzadeh, A.; Meyer, J. G.; Raedschelders, K.; Van Eyk, J. E. Parallelization with Dual-Trap Single-Column Configuration Maximizes Throughput of Proteomic Analysis. *Anal. Chem.* **2022**, *94* (36), 12452–12460.
- (430) Kreimer, S.; Binek, A.; Chazarin, B.; Cho, J. H.; Haghani, A.; Hutton, A.; Marbán, E.; Mastali, M.; Meyer, J. G.; Mesquita, T.; Song, Y.; Van Eyk, J.; Parker, S. High-Throughput Single-Cell Proteomic Analysis of Organ-Derived Heterogeneous Cell Populations by Nanoflow Dual-Trap Single-Column Liquid Chromatography. *Anal. Chem.* **2023**, *95* (24), 9145–9150.
- (431) Pellett, J.; Lukulay, P.; Mao, Y.; Bowen, W.; Reed, R.; Ma, M.; Munger, R. C.; Dolan, J. W.; Wrisley, L.; Medwid, K.; Toltl, N. P.; Chan, C. C.; Skibic, M.; Biswas, K.; Wells, K. A.; Snyder, L. R. “Orthogonal” Separations for Reversed-Phase Liquid Chromatography. *Journal of Chromatography A* **2006**, *1101* (1–2), 122–135.

- (432) Washburn, M. P.; Wolters, D.; Yates, J. R., III Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat Biotechnol* **2001**, *19* (3), 242–247.
- (433) Wang, Y.; Yang, F.; Gritsenko, M. A.; Wang, Y.; Clauss, T.; Liu, T.; Shen, Y.; Monroe, M. E.; Lopez-Ferrer, D.; Reno, T.; Moore, R. J.; Klemke, R. L.; Camp, D. G., II; Smith, R. D. Reversed-phase Chromatography with Multiple Fraction Concatenation Strategy for Proteome Profiling of Human MCF10A Cells. *Proteomics* **2011**, *11* (10), 2019–2026.
- (434) Kawashima, Y.; Nagai, H.; Konno, R.; Ishikawa, M.; Nakajima, D.; Sato, H.; Nakamura, R.; Furuyashiki, T.; Ohara, O. Single-Shot 10K Proteome Approach: Over 10,000 Protein Identifications by Data-Independent Acquisition-Based Single-Shot Proteomics with Ion Mobility Spectrometry. *J. Proteome Res.* **2022**, *21* (6), 1418–1427.
- (435) Biemann, K.; Gapp, G.; Seibl, J. Application of Mass Spectrometry to Structure Problems. I. Amino Acid Sequence in Peptides. *J. Am. Chem. Soc.* **1959**, *81* (9), 2274–2275.
- (436) Nau, H.; Kelley, J. A.; Biemann, K. Determination of the Amino Acid Sequence of the C-Terminal Cyanogen Bromide Fragment of Actin by Computer-Assisted Gas Chromatography-Mass Spectrometry. *J. Am. Chem. Soc.* **1973**, *95* (21), 7162–7164.
- (437) Hunt, D. F.; Bone, W. M.; Shabanowitz, J.; Rhodes, J.; Ballard, J. M. Sequence Analysis of Oligopeptides by Secondary Ion/Collision Activated Dissociation Mass Spectrometry. *Anal. Chem.* **1981**, *53* (11), 1704–1706.
- (438) Yates, J. R., III The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. *J. Am. Chem. Soc.* **2013**, *135* (5), 1629–1640.
- (439) The Nobel Prize in Chemistry 2002. NobelPrize.org. <https://www.nobelprize.org/prizes/chemistry/2002/summary/> (accessed 2024-05-07).
- (440) Karas, M.; Bachmann, D.; Hillenkamp, F. Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules. *Analytical Chemistry* **1985**, *57* (14), 2935–2939.
- (441) Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988**, *60* (20), 2299–2301.
- (442) The Scientist :: Nobel Prize controversy. <https://web.archive.org/web/20070517202246/http://cmbi.bjmu.edu.cn/news/0212/55.htm> (accessed 2024-05-07).
- (443) Yang, H.; Wan, D.; Song, F.; Liu, Z.; Liu, S. A-Cyano-4-Hydroxycinnamic Acid, Sinapinic Acid, and Ferulic Acid as Matrices and Alkylating Agents for Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometric Analysis of Cysteine-Containing Peptides. *Rapid Commun Mass Spectrom* **2013**, *27* (12), 1410–1412.
- (444) Dreisewerd, K. The Desorption Process in MALDI. *Chem. Rev.* **2003**, *103* (2), 395–426.
- (445) Karas, M.; Bahr, U.; Strupat, K.; Hillenkamp, F.; Tsarbopoulos, A.; Pramanik, B. N. Matrix Dependence of Metastable Fragmentation of Glycoproteins in MALDI TOF Mass Spectrometry. *Anal. Chem.* **1995**, *67* (3), 675–679.
- (446) Buchberger, A. R.; DeLaney, K.; Johnson, J.; Li, L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal. Chem.* **2018**, *90* (1), 240–265.
- (447) Meyer, J. G.; Komives, E. A. Charge State Coalescence During Electrospray Ionization Improves Peptide Identification by Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (8), 1390–1399.
- (448) Kebarle, P.; Verkerk, U. H. Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know Now. *Mass Spectrom. Rev.* **2009**, *28* (6), 898–917.
- (449) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9.
- (450) Griffiths, J. A Brief History of Mass Spectrometry. *Anal. Chem.* **2008**, *80* (15), 5678–5683.
- (451) Klampfl, C. W.; Himmelsbach, M. Direct Ionization Methods in Mass Spectrometry: An Overview. *Analytica Chimica Acta* **2015**, *890*, 44–59.
- (452) Awad, H.; Khamis, M. M.; El-Aneel, A. Mass Spectrometry, Review of the Basics: Ionization. *Applied Spectroscopy Reviews* **2015**, *50* (2), 158–175.
- (453) Li, C.; Chu, S.; Tan, S.; Yin, X.; Jiang, Y.; Dai, X.; Gong, X.; Fang, X.; Tian, D. Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. *Front. Chem.* **2021**, *9*, 813359.
- (454) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual Rev. Anal. Chem.* **2015**, *8* (1), 61–80.
- (455) Glish, G. L.; Burinsky, D. J. Hybrid Mass Spectrometers for Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (2), 161–172.
- (456) Hecht, E. S.; Scigelova, M.; Eliuk, S.; Makarov, A. Fundamentals and Advances of Orbitrap Mass Spectrometry. *Encyclopedia of Analytical Chemistry* **2019**, 1–40.
- (457) Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85* (11), 5288–5296.
- (458) Tsybin, Y. O.; Nagornov, K. O.; Kozhinov, A. N. Advanced Fundamentals in Fourier Transform Mass Spectrometry. In *Fundamentals and Applications of Fourier Transform Mass Spectrometry*; Elsevier, 2019; pp 113–132. DOI: 10.1016/b978-0-12-814013-0.00005-3.
- (459) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol* **2005**, *402*, 148–185.
- (460) Thomson, B. A.; Douglas, D. J.; Corr, J. J.; Hager, J. W.; Jolliffe, C. L. Improved Collisionally Activated Dissociation Efficiency and Mass Resolution on a Triple Quadrupole Mass Spectrometer System. *Anal. Chem.* **1995**, *67* (10), 1696–1704.
- (461) Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected Reaction Monitoring for Quantitative Proteomics: A Tutorial. *Molecular Systems Biology* **2008**, *4* (1), 222.
- (462) Johnson, J. V.; Yost, R. A.; Kelley, P. E.; Bradford, D. C. Tandem-in-Space and Tandem-in-Time Mass Spectrometry: Triple Quadrupoles and Quadrupole Ion Traps. *Anal. Chem.* **1990**, *62* (20), 2162–2172.
- (463) Sleno, L.; Volmer, D. A. Ion Activation Methods for Tandem Mass Spectrometry. *J. Mass Spectrom.* **2004**, *39* (10), 1091–1112.
- (464) Anderson, L.; Hunter, C. L. Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. *Molecular & Cellular Proteomics* **2006**, *5* (4), 573–588.
- (465) Hager, J. W. Recent Trends in Mass Spectrometer Development. *Anal. Bioanal. Chem.* **2004**, *378* (4), 845–850.
- (466) Yost, R. A.; Enke, C. G. Selected Ion Fragmentation with a Tandem Quadrupole Mass Spectrometer. *J. Am. Chem. Soc.* **1978**, *100* (7), 2274–2275.
- (467) Hopfgartner, G.; Varesio, E.; Tschäppät, V.; Grivet, C.; Bourgonne, E.; Leuthold, L. A. Triple Quadrupole Linear Ion Trap Mass Spectrometer for the Analysis of Small Molecules and Macromolecules. *J. Mass Spectrom.* **2004**, *39* (8), 845–855.
- (468) Ma, S.; Zhu, M. Recent Advances in Applications of Liquid Chromatography–Tandem Mass Spectrometry to the Analysis of Reactive Drug Metabolites. *Chemico-Biological Interactions* **2009**, *179* (1), 25–37.
- (469) Hernández, F.; Cervera, M. I.; Portolés, T.; Beltrán, J.; Pitarch, E. The Role of GC-MS/MS with Triple Quadrupole in Pesticide Residue Analysis in Food and the Environment. *Anal. Methods* **2013**, *5* (21), 5875.
- (470) van der Gugten, J. G. Tandem Mass Spectrometry in the Clinical Laboratory: A Tutorial Overview. *Clinical Mass Spectrometry* **2020**, *15*, 36–43.
- (471) Anonymous. Proceedings of the American Physical Society. *Phys. Rev.* **1946**, *69* (11-12), 674–674.

- (472) Wolff, M. M.; Stephens, W. E. A Pulsed Mass Spectrometer with Time Dispersion. *Review of Scientific Instruments* **1953**, *24* (8), 616–617.
- (473) Wiley, W. C.; McLaren, I. H. Time-of-Flight Mass Spectrometer with Improved Resolution. *Review of Scientific Instruments* **1955**, *26* (12), 1150–1157.
- (474) Sandow, J. J.; Infusini, G.; Dagley, L. F.; Larsen, R.; Webb, A. I. Simplified High-Throughput Methods for Deep Proteome Analysis on the timsTOF Pro. *bioRxiv* **2019**, DOI: [10.1101/657908](https://doi.org/10.1101/657908).
- (475) Aballo, T. J.; Roberts, D. S.; Melby, J. A.; Buck, K. M.; Brown, K. A.; Ge, Y. Ultrafast and Reproducible Proteomics from Small Amounts of Heart Tissue Enabled by Azo and timsTOF Pro. *J. Proteome Res.* **2021**, *20* (8), 4203–4211.
- (476) Pan, S.; Zhang, H.; Rush, J.; Eng, J.; Zhang, N.; Patterson, D.; Comb, M. J.; Aebersold, R. High Throughput Proteome Screening for Biomarker Detection. *Molecular & Cellular Proteomics* **2005**, *4* (2), 182–190.
- (477) Wörner, T. P.; Shamorkina, T. M.; Snijder, J.; Heck, A. J. R. Mass Spectrometry-Based Structural Virology. *Anal. Chem.* **2021**, *93* (1), 620–640.
- (478) Snijder, J.; Heck, A. J. R. Analytical Approaches for Size and Mass Analysis of Large Protein Assemblies. *Annual Rev. Anal. Chem.* **2014**, *7* (1), 43–64.
- (479) Comby-Zerbino, C.; Dagany, X.; Chirot, F.; Dugourd, P.; Antoine, R. The Emergence of Mass Spectrometry for Characterizing Nanomaterials. Atomically Precise Nanoclusters and Beyond. *Mater. Adv.* **2021**, *2* (15), 4896–4913.
- (480) Kingdon, R. H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Phys. Rev.* **1923**, *21*, 408.
- (481) Knight, R. D. Storage of ions from laser-produced plasmas. *Appl. Phys. Lett.* **1981**, *38*, 221–223.
- (482) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72* (6), 1156–1162.
- (483) Makarov, A.; Scigelova, M. Coupling Liquid Chromatography to Orbitrap Mass Spectrometry. *Journal of Chromatography A* **2010**, *1217* (25), 3938–3945.
- (484) Crutchfield, C. A.; Thomas, S. N.; Sokoll, L. J.; Chan, D. W. Advances in Mass Spectrometry-Based Clinical Biomarker Discovery. *Clin. Proteom.* **2016**, *13* (1), 1.
- (485) Meyer, J. G.; Mukkamalla, S.; Steen, H.; Nesvizhskii, A. I.; Gibson, B. W.; Schilling, B. PIQED: Automated Identification and Quantification of Protein Modifications from DIA-MS Data. *Nat. Methods* **2017**, *14* (7), 646–647.
- (486) Momenzadeh, A.; Jiang, Y.; Kreimer, S.; Teigen, L. E.; Zepeda, C. S.; Haghighi, A.; Mastali, M.; Song, Y.; Hutton, A.; Parker, S. J.; Van Eyk, J. E.; Sundberg, C. W.; Meyer, J. G. A Complete Workflow for High Throughput Human Single Skeletal Muscle Fiber Proteomics. *J. Am. Soc. Mass Spectrom.* **2023**, *34* (9), 1858–1867.
- (487) Smits, A. H.; Vermeulen, M. Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends in Biotechnology* **2016**, *34* (10), 825–834.
- (488) Sun, J.; Liu, X. R.; Li, S.; He, P.; Li, W.; Gross, M. L. Nanoparticles and Photochemistry for Native-Like Transmembrane Protein Footprinting. *Nat. Commun.* **2021**, *12* (1), 7270.
- (489) Sun, J.; Li, S.; Li, W.; Gross, M. L. Carbocation Footprinting of Soluble and Transmembrane Proteins. *Anal. Chem.* **2021**, *93* (39), 13101–13105.
- (490) Comisarow, M. B.; Marshall, A. G. Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chemical Physics Letters* **1974**, *25* (2), 282–283.
- (491) Cho, Y.; Ahmed, A.; Islam, A.; Kim, S. Developments in FT-ICR MS Instrumentation, Ionization Techniques, and Data Interpretation Methods for Petroleomics. *Mass Spectrometry Reviews* **2015**, *34* (2), 248–263.
- (492) Jiang, B.; Kuang, B. Y.; Liang, Y.; Zhang, J.; Huang, X. H. H.; Xu, C.; Yu, J. Z.; Shi, Q. Molecular Composition of Urban Organic Aerosols on Clear and Hazy Days in Beijing: A Comparative Study Using FT-ICR MS. *Environ. Chem.* **2016**, *13* (5), 888.
- (493) Bogdanov, B.; Smith, R. D. Proteomics by FTICR Mass Spectrometry: Top down and Bottom Up. *Mass Spectrometry Reviews* **2005**, *24* (2), 168–200.
- (494) Motoyama, A.; Yates, J. R., III. Multidimensional LC Separations in Shotgun Proteomics. *Anal. Chem.* **2008**, *80* (19), 7187–7193.
- (495) Chen, G.; Pramanik, B. N. Application of LC/MS to Proteomics Studies: Current Status and Future Prospects. *Drug Discovery Today* **2009**, *14* (9–10), 465–471.
- (496) May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal. Chem.* **2015**, *87* (3), 1422–1436.
- (497) Dodds, J. N.; Baker, E. S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (11), 2185–2195.
- (498) Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M. A.; Raether, O.; Mann, M. Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **2015**, *14* (12), 5378–5387.
- (499) Ridgeway, M. E.; Bleiholder, C.; Mann, M.; Park, M. A. Trends in Trapped Ion Mobility – Mass Spectrometry Instrumentation. *TrAC Trends in Analytical Chemistry* **2019**, *116*, 324–331.
- (500) Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M. A Cyclic Ion Mobility-Mass Spectrometry System. *Anal. Chem.* **2019**, *91* (13), 8564–8573.
- (501) Odenkirk, M. T.; Baker, E. S. Utilizing Drift Tube Ion Mobility Spectrometry for the Evaluation of Metabolites and Xenobiotics. In *Methods in Molecular Biology*; Springer, 2019; pp 35–54. DOI: [10.1007/978-1-0716-0030-6_2](https://doi.org/10.1007/978-1-0716-0030-6_2).
- (502) Jiang, W.; Chung, N. A.; May, J. C.; McLean, J. A.; Robinson, R. A. S. Ion Mobility–Mass Spectrometry. *Encyclopedia of Analytical Chemistry* **2019**, 1–34.
- (503) Swearingen, K. E.; Moritz, R. L. High-Field Asymmetric Waveform Ion Mobility Spectrometry for Mass Spectrometry-Based Proteomics. *Expert Review of Proteomics* **2012**, *9* (5), 505–517.
- (504) Hale, O. J.; Illes-Toth, E.; Mize, T. H.; Cooper, H. J. High-Field Asymmetric Waveform Ion Mobility Spectrometry and Native Mass Spectrometry: Analysis of Intact Protein Assemblies and Protein Complexes. *Anal. Chem.* **2020**, *92* (10), 6811–6816.
- (505) Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J.-J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal. Chem.* **2018**, *90* (15), 9529–9537.
- (506) GUEVREMENT, R. High-Field Asymmetric Waveform Ion Mobility Spectrometry: A New Tool for Mass Spectrometry. *Journal of Chromatography A* **2004**, *1058* (1–2), 3–19.
- (507) Ridgeway, M. E.; Lubeck, M.; Jordens, J.; Mann, M.; Park, M. A. Trapped Ion Mobility Spectrometry: A Short Review. *International Journal of Mass Spectrometry* **2018**, *425*, 22–35.
- (508) Meier, F.; Park, M. A.; Mann, M. Trapped Ion Mobility Spectrometry and Parallel Accumulation–Serial Fragmentation in Proteomics. *Molecular & Cellular Proteomics* **2021**, *20*, 100138.
- (509) Tolmachev, A. V.; Webb, I. K.; Ibrahim, Y. M.; Garimella, S. V. B.; Zhang, X.; Anderson, G. A.; Smith, R. D. Characterization of Ion Dynamics in Structures for Lossless Ion Manipulations. *Anal. Chem.* **2014**, *86* (18), 9162–9168.
- (510) Deng, L.; Webb, I. K.; Garimella, S. V. B.; Hamid, A. M.; Zheng, X.; Norheim, R. V.; Prost, S. A.; Anderson, G. A.; Sandoval, J. A.; Baker, E. S.; Ibrahim, Y. M.; Smith, R. D. Serpentine Ultralong Path with Extended Routing (SUPER) High Resolution Traveling Wave Ion Mobility-MS Using Structures for Lossless Ion Manipulations. *Anal. Chem.* **2017**, *89* (8), 4628–4634.
- (511) Arndt, J. R.; Wormwood Moser, K. L.; Van Aken, G.; Doyle, R. M.; Talamantes, T.; DeBord, D.; Maxon, L.; Stafford, G.; Fjeldsted, J.; Miller, B.; Sherman, M. High-Resolution Ion-Mobility-Enabled

Peptide Mapping for High-Throughput Critical Quality Attribute Monitoring. *J. Am. Soc. Mass Spectrom.* **2021**, 32 (8), 2019–2032.

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

(513) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., III Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, 113 (4), 2343–2394.

(514) Yu, L.; Xiong, Y.-M.; Polfer, N. C. Periodicity of Monoisotopic Mass Isomers and Isobars in Proteomics. *Anal. Chem.* **2011**, 83 (20), 8019–8023.

(515) Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. Protein Sequencing by Tandem Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, 83 (17), 6233–6237.

(516) Ivanov, M. V.; Bubis, J. A.; Gorshkov, V.; Tarasova, I. A.; Levitsky, L. I.; Lobas, A. A.; Solovyeva, E. M.; Pridatchenko, M. L.; Kjeldsen, F.; Gorshkov, M. V. DirectMS1: MS/MS-Free Identification of 1000 Proteins of Cellular Proteomes in 5 Minutes. *Anal. Chem.* **2020**, 92 (6), 4326–4333.

(517) Jaitly, N.; Monroe, M. E.; Petyuk, V. A.; Clauss, T. R. W.; Adkins, J. N.; Smith, R. D. Robust Algorithm for Alignment of Liquid Chromatography–Mass Spectrometry Analyses in an Accurate Mass and Time Tag Data Analysis Pipeline. *Anal. Chem.* **2006**, 78 (21), 7397–7409.

(518) Meier, F.; Geyer, P. E.; Virreira Winter, S.; Cox, J.; Mann, M. BoxCar Acquisition Method Enables Single-Shot Proteomics at a Depth of 10,000 Proteins in 100 Minutes. *Nat Methods* **2018**, 15 (6), 440–448.

(519) Roepstorff, P.; Fohlman, J. Letter to the Editors. *Biol. Mass Spectrom.* **1984**, 11 (11), 601–601.

(520) McCloskey, J. A. *Mass spectrometry*; Methods in enzymology; Academic Press: San Diego, 1990.

(521) Goeringer, D. E.; McLuckey, S. A. Evolution of Ion Internal Energy During Collisional Excitation in the Paul Ion Trap: A Stochastic Approach. *The Journal of Chemical Physics* **1996**, 104 (6), 2214–2221.

(522) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-Energy C-Trap Dissociation for Peptide Modification Analysis. *Nat Methods* **2007**, 4 (9), 709–712.

(523) Hohmann, L. J.; Eng, J. K.; Gemmill, A.; Klimek, J.; Vitek, O.; Reid, G. E.; Martin, D. B. Quantification of the Compositional Information Provided by Immonium Ions on a Quadrupole-Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2008**, 80 (14), 5596–5606.

(524) Nolting, D.; Malek, R.; Makarov, A. Ion Traps in Modern Mass Spectrometry. *Mass Spectrometry Reviews* **2019**, 38 (2), 150–168.

(525) Kovacic, I.; Rand, R.; Mohamed Sah, S. Mathieu's Equation and Its Generalizations: Overview of Stability Charts and Their Features. *Applied Mechanics Reviews* **2018**, 70 (2), 020802.

(526) March, R. E. Ion Trap Mass Spectrometers. In *Encyclopedia of Spectroscopy and Spectrometry*; Elsevier, 1999; pp 1000–1009. DOI: 10.1006/rwsp.2000.0143.

(527) Palumbo, A. M.; Tepe, J. J.; Reid, G. E. Mechanistic Insights into the Multistage Gas-Phase Fragmentation Behavior of Phosphoserine- and Phosphothreonine-Containing Peptides. *J. Proteome Res.* **2008**, 7 (2), 771–779.

(528) Schroeder, M. J.; Shabanowitz, J.; Schwartz, J. C.; Hunt, D. F.; Coon, J. J. A Neutral Loss Activation Method for Improved Phosphopeptide Sequence Analysis by Quadrupole Ion Trap Mass Spectrometry. *Anal. Chem.* **2004**, 76 (13), 3590–3598.

(529) Wysocki, V. H.; Tsapralis, G.; Smith, L. L.; Breci, L. A. Mobile and Localized Protons: A Framework for Understanding Peptide Dissociation. *J Mass Spectrom* **2000**, 35 (12), 1399–1406.

(530) Coon, J. J. Collisions or Electrons? Protein Sequence Analysis in the 21st Century. *Anal. Chem.* **2009**, 81 (9), 3208–3215.

(531) Udeshi, N. D.; Compton, P. D.; Shabanowitz, J.; Hunt, D. F.; Rose, K. L. Methods for Analyzing Peptides and Proteins on a Chromatographic Timescale by Electron-Transfer Dissociation Mass Spectrometry. *Nat Protoc* **2008**, 3 (11), 1709–1717.

(532) Sarbu, M.; Ghiulai, R. M.; Zamfir, A. D. Recent Developments and Applications of Electron Transfer Dissociation Mass Spectrometry in Proteomics. *Amino Acids* **2014**, 46 (7), 1625–1634.

(533) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, 120 (13), 3265–3266.

(534) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101 (26), 9528–9533.

(535) Chen, X.; Hao, C. Where Does the Electron Go? Electron Distribution and Reactivity of Peptide Cation Radicals Formed by Electron Transfer in the Gas Phase. *J. Am. Chem. Soc.* **2008**, 130 (27), 8818–8833.

(536) Syrtstad, E. A.; Tureček, F. Toward a General Mechanism of Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2005**, 16 (2), 208–224.

(537) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* **2018**, 90 (1), 40–64.

(538) Lermyte, F.; Valkenburg, D.; Loo, J. A.; Sobott, F. Radical Solutions: Principles and Application of Electron-based Dissociation in Mass Spectrometry-based Analysis of Protein Structure. *Mass Spectrometry Reviews* **2018**, 37 (6), 750–771.

(539) Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J.-H.; Begley, T. P.; McLafferty, F. W. Top Down Characterization of Larger Proteins (45 kDa) by Electron Capture Dissociation Mass Spectrometry. *Journal of the American Chemical Society* **2002**, 124 (4), 672–678.

(540) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteom Form Characterization. *Analytical Chemistry* **2015**, 87 (8), 4152–4158.

(541) Schürch, S. Characterization of Nucleic Acids by Tandem Mass Spectrometry - The Second Decade (2004–2013): From DNA to RNA and Modified Sequences. *Mass Spectrometry Reviews* **2016**, 35 (4), 483–523.

(542) Peters-Clarke, T. M.; Quan, Q.; Brademan, D. R.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. Ribonucleic Acid Sequence Characterization by Negative Electron Transfer Dissociation Mass Spectrometry. *Anal. Chem.* **2020**, 92 (6), 4436–4444.

(543) Pimentel, E. B.; Peters-Clarke, T. M.; Coon, J. J.; Martell, J. D. DNA-Scaffolded Synergistic Catalysis. *J. Am. Chem. Soc.* **2021**, 143 (50), 21402–21409.

(544) Taucher, M.; Breuker, K. Characterization of Modified RNA by Top-Down Mass Spectrometry. *Angewandte Chemie* **2012**, 124 (45), 11451–11454.

(545) Crittenden, C. M.; Lanzillotti, M. B.; Chen, B. Top-Down Mass Spectrometry of Synthetic Single Guide Ribonucleic Acids Enabled by Facile Sample Clean-Up. *Anal. Chem.* **2023**, 95 (6), 3180–3186.

(546) Merrifield, J. L.; Pimentel, E. B.; Peters-Clarke, T. M.; Nesbitt, D. J.; Coon, J. J.; Martell, J. D. DNA-Compatible Copper/TEMPO Oxidation for DNA-Encoded Libraries. *Bioconjugate Chem.* **2023**, 34 (8), 1380–1386.

(547) Shaw, J. B.; Malhan, N.; Vasil'ev, Y. V.; Lopez, N. I.; Makarov, A.; Beckman, J. S.; Voinov, V. G. Sequencing Grade Tandem Mass Spectrometry for Top-Down Proteomics Using Hybrid Electron Capture Dissociation Methods in a Benchtop Orbitrap Mass Spectrometer. *Anal. Chem.* **2018**, 90 (18), 10819–10827.

(548) Fort, K. L.; Cramer, C. N.; Voinov, V. G.; Vasil'ev, Y. V.; Lopez, N. I.; Beckman, J. S.; Heck, A. J. R. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *J. Proteome Res.* **2018**, 17 (2), 926–933.

(549) Baba, T.; Ryumin, P.; Duchoslav, E.; Chen, K.; Chelur, A.; Loyd, B.; Chernushevich, I. Dissociation of Biomolecules by an Intense Low-Energy Electron Beam in a High Sensitivity Time-of-

- Flight Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2021**, 32 (8), 1964–1975.
- (550) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Molecular & Cellular Proteomics* **2007**, 6 (11), 1942–1951.
- (551) Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. Electron-Transfer Ion/Ion Reactions of Doubly Protonated Peptides: Effect of Elevated Bath Gas Temperature. *Anal. Chem.* **2005**, 77 (17), 5662–5669.
- (552) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors. *Anal. Chem.* **2007**, 79 (2), 477–485.
- (553) Campbell, J. L.; Hager, J. W.; Le Blanc, J. C. Y. On Performing Simultaneous Electron Transfer Dissociation and Collision-Induced Dissociation on Multiply Protonated Peptides in a Linear Ion Trap. *J. Am. Soc. Mass Spectrom.* **2009**, 20 (9), 1672–1683.
- (554) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2012**, 84 (22), 9668–9673.
- (555) Mikhailov, V. A.; Cooper, H. J. Activated Ion Electron Capture Dissociation (AI ECD) of Proteins: Synchronization of Infrared and Electron Irradiation with Ion Magnetron Motion. *J. Am. Soc. Mass Spectrom.* **2009**, 20 (5), 763–771.
- (556) Horn, D. M.; Ge, Y.; McLafferty, F. W. Activated Ion Electron Capture Dissociation for Mass Spectral Sequencing of Larger (42 kDa) Proteins. *Anal. Chem.* **2000**, 72 (20), 4778–4784.
- (557) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal. Chem.* **2015**, 87 (14), 7109–7116.
- (558) Ledvina, A. R.; McAlister, G. C.; Gardner, M. W.; Smith, S. I.; Madsen, J. A.; Schwartz, J. C.; Stafford, G. C.; Syka, J. E. P.; Brodbelt, J. S.; Coon, J. J. Infrared Photoactivation Reduces Peptide Folding and Hydrogen-Atom Migration Following ETD Tandem Mass Spectrometry. *Angew Chem Int Ed* **2009**, 48 (45), 8526–8528.
- (559) Peters-Clarke, T. M.; Schauer, K. L.; Riley, N. M.; Lodge, J. M.; Westphall, M. S.; Coon, J. J. Optical Fiber-Enabled Photoactivation of Peptides and Proteins. *Anal. Chem.* **2020**, 92 (18), 12363–12370.
- (560) Peters-Clarke, T. M.; Riley, N. M.; Westphall, M. S.; Coon, J. J. Practical Effects of Intramolecular Hydrogen Rearrangement in Electron Transfer Dissociation-Based Proteomics. *J. Am. Soc. Mass Spectrom.* **2022**, 33 (1), 100–110.
- (561) Cannon, J. R.; Holden, D. D.; Brodbelt, J. S. Hybridizing Ultraviolet Photodissociation with Electron Transfer Dissociation for Intact Protein Characterization. *Anal. Chem.* **2014**, 86 (21), 10970–10977.
- (562) Kjeldsen, F.; Haselmann, K. F.; Budnik, B. A.; Jensen, F.; Zubarev, R. A. Dissociative Capture of Hot (3–13 eV) Electrons by Polypeptide Polycations: An Efficient Process Accompanied by Secondary Fragmentation. *Chemical Physics Letters* **2002**, 356 (3), 201–206.
- (563) Fung, Y. M. E.; Adams, C. M.; Zubarev, R. A. Electron Ionization Dissociation of Singly and Multiply Charged Peptides. *J. Am. Chem. Soc.* **2009**, 131 (29), 9977–9985.
- (564) Riley, N. M.; Westphall, M. S.; Hebert, A. S.; Coon, J. J. Implementation of Activated Ion Electron Transfer Dissociation on a Quadrupole-Orbitrap-Linear Ion Trap Hybrid Mass Spectrometer. *Anal. Chem.* **2017**, 89 (12), 6358–6366.
- (565) Macias, L. A.; Santos, I. C.; Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2020**, 92 (1), 227–251.
- (566) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, 43 (8), 2757–2783.
- (567) Brodbelt, J. S.; Morrison, L. J.; Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **2020**, 120 (7), 3328–3380.
- (568) Maitre, P.; Scuderi, D.; Corinti, D.; Chiavarino, B.; Crestoni, M. E.; Fornarini, S. Applications of Infrared Multiple Photon Dissociation (IRMPD) to the Detection of Posttranslational Modifications. *Chem. Rev.* **2020**, 120 (7), 3261–3295.
- (569) Crowe, M. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation (IRMPD) and Collisionally Activated Dissociation of Peptides in a Quadrupole Ion Trap with Selective IRMPD of Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2004**, 15 (11), 1581–1592.
- (570) Thompson, M. S.; Cui, W.; Reilly, J. P. Fragmentation of Singly Charged Peptide Ions by Photodissociation at $\lambda=157$ Nm. *Angew Chem Int Ed* **2004**, 43 (36), 4791–4794.
- (571) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 Nm and Its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, 9 (8), 4205–4214.
- (572) Fornelli, L.; Srzentić, K.; Toby, T. K.; Doubleday, P. F.; Huguet, R.; Mullen, C.; Melani, R. D.; dos Santos Seckler, H.; DeHart, C. J.; Weisbrod, C. R.; Durbin, K. R.; Greer, J. B.; Early, B. P.; Fellers, R. T.; Zabrouskov, V.; Thomas, P. M.; Compton, P. D.; Kelleher, N. L. Thorough Performance Evaluation of 213 Nm Ultraviolet Photodissociation for Top-down Proteomics. *Molecular & Cellular Proteomics* **2020**, 19 (2), 405–420.
- (573) Yeh, G. K.; Sun, Q.; Meneses, C.; Julian, R. R. Rapid Peptide Fragmentation Without Electrons, Collisions, Infrared Radiation, or Native Chromophores. *J. Am. Soc. Mass Spectrom.* **2009**, 20 (3), 385–393.
- (574) Park, S.; Ahn, W.; Lee, S.; Han, S. Y.; Rhee, B. K.; Oh, H. B. Ultraviolet Photodissociation at 266 Nm of Phosphorylated Peptide Cations. *Rapid Comm Mass Spectrometry* **2009**, 23 (23), 3609–3620.
- (575) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, 79 (20), 7883–7892.
- (576) Ly, T.; Julian, R. R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, 130 (1), 351–358.
- (577) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments Towards Applications for Mass-Spectrometry-Based Proteomics. *Angew Chem Int Ed* **2009**, 48 (39), 7130–7137.
- (578) Julian, R. R. The Mechanism Behind Top-Down UVPD Experiments: Making Sense of Apparent Contradictions. *J. Am. Soc. Mass Spectrom.* **2017**, 28 (9), 1823–1826.
- (579) Brodbelt, J. S. Deciphering Combinatorial Post-Translational Modifications by Top-down Mass Spectrometry. *Current Opinion in Chemical Biology* **2022**, 70, No. 102180.
- (580) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, 135 (34), 12646–12651.
- (581) Kolbowski, L.; Belsom, A.; Rappsilber, J. Ultraviolet Photodissociation of Tryptic Peptide Backbones at 213 Nm. *J. Am. Soc. Mass Spectrom.* **2020**, 31 (6), 1282–1290.
- (582) Lanzillotti, M.; Brodbelt, J. S. Comparison of Top-Down Protein Fragmentation Induced by 213 and 193 Nm UVPD. *J. Am. Soc. Mass Spectrom.* **2023**, 34 (2), 279–285.
- (583) Davis, M. T.; Stahl, D. C.; Swiderek, K. M.; Lee, T. D. Capillary Liquid Chromatography/Mass Spectrometry for Peptide and Protein Characterization. *Methods* **1994**, 6 (3), 304–314.
- (584) Yates, J. R.; Eng, J. K.; McCormack, A. L.; Schieltz, D. Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. *Anal. Chem.* **1995**, 67 (8), 1426–1436.
- (585) Stahl, D. C.; Swiderek, K. M.; Davis, M. T.; Lee, T. D. Data-Controlled Automation of Liquid Chromatography/Tandem Mass Spectrometry Analysis of Peptide Mixtures. *J. Am. Soc. Mass Spectrom.* **1996**, 7 (6), 532–540.

- (586) Hoopmann, M. R.; Merrihew, G. E.; von Haller, P. D.; MacCoss, M. J. Post Analysis Data Acquisition for the Iterative MS/MS Sampling of Proteomics Mixtures. *J. Proteome Res.* **2009**, *8* (4), 1870–1875.
- (587) Picotti, P.; Aebersold, R.; Domon, B. The Implications of Proteolytic Background for Shotgun Proteomics. *Molecular & Cellular Proteomics* **2007**, *6* (9), 1589–1598.
- (588) Meyer, J. G.; Schilling, B. Clinical Applications of Quantitative Proteomics Using Targeted and Untargeted Data-Independent Acquisition Techniques. *Expert Review of Proteomics* **2017**, *14* (5), 419–429.
- (589) Colangelo, C. M.; Chung, L.; Bruce, C.; Cheung, K.-H. Review of Software Tools for Design and Analysis of Large Scale MRM Proteomic Datasets. *Methods* **2013**, *61* (3), 287–298.
- (590) Kuhn, E.; Wu, J.; Karl, J.; Liao, H.; Zolg, W.; Guild, B. Quantification of C-reactive Protein in the Serum of Patients with Rheumatoid Arthritis Using Multiple Reaction Monitoring Mass Spectrometry and ¹³C-labeled Peptide Standards. *Proteomics* **2004**, *4* (4), 1175–1186.
- (591) Anderson, N. L.; Anderson, N. G.; Haines, L. R.; Hardie, D. B.; Olafson, R. W.; Pearson, T. W. Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res.* **2004**, *3* (2), 235–244.
- (592) Vidova, V.; Spacil, Z. A Review on Mass Spectrometry-Based Quantitative Proteomics: Targeted and Data Independent Acquisition. *Analytica Chimica Acta* **2017**, *964*, 7–23.
- (593) Lemoine, J.; Fortin, T.; Salvador, A.; Jaffuel, A.; Charrier, J.-P.; Choquet-Kastylevsky, G. The Current Status of Clinical Proteomics and the Use of MRM and MRM³ for Biomarker Validation. *Expert Review of Molecular Diagnostics* **2012**, *12* (4), 333–342.
- (594) Stahl-Zeng, J.; Lange, V.; Ossola, R.; Eckhardt, K.; Krek, W.; Aebersold, R.; Domon, B. High Sensitivity Detection of Plasma Proteins by Multiple Reaction Monitoring of N-Glycosites. *Molecular & Cellular Proteomics* **2007**, *6* (10), 1809–1817.
- (595) Picotti, P.; Lam, H.; Campbell, D.; Deutsch, E. W.; Mirzaei, H.; Ranish, J.; Domon, B.; Aebersold, R. A Database of Mass Spectrometric Assays for the Yeast Proteome. *Nat Methods* **2008**, *5* (11), 913–914.
- (596) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Molecular & Cellular Proteomics* **2012**, *11* (11), 1475–1488.
- (597) Ronsein, G. E.; Pamir, N.; von Haller, P. D.; Kim, D. S.; Oda, M. N.; Jarvik, G. P.; Vaisar, T.; Heinecke, J. W. Parallel Reaction Monitoring (PRM) and Selected Reaction Monitoring (SRM) Exhibit Comparable Linearity, Dynamic Range and Precision for Targeted Quantitative HDL Proteomics. *Journal of Proteomics* **2015**, *113*, 388–399.
- (598) Purvine, S.; Eppel, J.; Yi, E. C.; Goodlett, D. R. Shotgun Collision-induced Dissociation of Peptides Using a Time of Flight Mass Analyzer. *Proteomics* **2003**, *3* (6), 847–850.
- (599) Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R., III Automated Approach for Quantitative Analysis of Complex Peptide Mixtures from Tandem Mass Spectra. *Nat Methods* **2004**, *1* (1), 39–45.
- (600) Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. UPLC/MS^E; a New Approach for Generating Molecular Fragment Information for Biomarker Structure Elucidation. *Rapid Comm Mass Spectrometry* **2006**, *20* (13), 1989–1994.
- (601) Panchaud, A.; Scherl, A.; Shaffer, S. A.; von Haller, P. D.; Kulasekara, H. D.; Miller, S. I.; Goodlett, D. R. Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal. Chem.* **2009**, *81* (15), 6481–6488.
- (602) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-Ion Fragmentation. *Molecular & Cellular Proteomics* **2010**, *9* (10), 2252–2261.
- (603) Carvalho, P. C.; Han, X.; Xu, T.; Cociorva, D.; Carvalho, M. da G.; Barbosa, V. C.; Yates, J. R., III XDIA: Improving on the Label-Free Data-Independent Analysis. *Bioinformatics* **2010**, *26* (6), 847–848.
- (604) Wong, J. W.; Schwahn, A. B.; Downard, K. M. ETISEQ – an Algorithm for Automated Elution Time Ion Sequencing of Concurrently Fragmented Peptides for Mass Spectrometry-Based Proteomics. *BMC Bioinformatics* **2009**, *10* (1), 244.
- (605) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics* **2012**, *11* (6), O111.016717.
- (606) Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.; MacLean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.; Kellmann, M.; Zabrouskov, V.; Wu, C. C.; MacCoss, M. J. Multiplexed MS/MS for Improved Data-Independent Acquisition. *Nat Methods* **2013**, *10* (8), 744–746.
- (607) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Xuan, Y.; Sondermann, J.; Schmidt, M.; Gomez-Varela, D.; Reiter, L. Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Molecular & Cellular Proteomics* **2017**, *16* (12), 2296–2309.
- (608) Meier, F.; Brunner, A.-D.; Frank, M.; Ha, A.; Bludau, I.; Voytk, E.; Kaspar-Schoenefeld, S.; Lubeck, M.; Raether, O.; Bache, N.; Aebersold, R.; Collins, B. C.; Röst, H. L.; Mann, M. diaPASEF: Parallel Accumulation–Serial Fragmentation Combined with Data-Independent Acquisition. *Nat Methods* **2020**, *17* (12), 1229–1236.
- (609) Jedrychowski, M. P.; Huttlin, E. L.; Haas, W.; Sowa, M. E.; Rad, R.; Gygi, S. P. Evaluation of HCD- and CID-Type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics. *Molecular & Cellular Proteomics* **2011**, *10* (12), M111.009910.
- (610) Nagaraj, N.; D’Souza, R. C. J.; Cox, J.; Olsen, J. V.; Mann, M. Feasibility of Large-Scale Phosphoproteomics with Higher Energy Collisional Dissociation Fragmentation. *J. Proteome Res.* **2010**, *9* (12), 6786–6794.
- (611) Riley, N. M.; Hebert, A. S.; Dürnberger, G.; Stanek, F.; Mechtler, K.; Westphall, M. S.; Coon, J. J. Phosphoproteomics with Activated Ion Electron Transfer Dissociation. *Anal. Chem.* **2017**, *89* (12), 6367–6376.
- (612) Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. F. M.; Mechtler, K.; Heck, A. J. R.; Mohammed, S. Unambiguous Phosphosite Localization Using Electron-Transfer/Higher-Energy Collision Dissociation (ET_hCD). *J. Proteome Res.* **2013**, *12* (3), 1520–1525.
- (613) Palumbo, A. M.; Reid, G. E. Evaluation of Gas-Phase Rearrangement and Competing Fragmentation Reactions on Protein Phosphorylation Site Assignment Using Collision Induced Dissociation-MS/MS and MS³. *Anal. Chem.* **2008**, *80* (24), 9735–9747.
- (614) Cui, L.; Reid, G. E. Examining Factors That Influence Erroneous Phosphorylation Site Localization via Competing Fragmentation and Rearrangement Reactions During Ion Trap <Scp>CID</Scp>-<scp>MS</Scp>/<Scp>MS</Scp> and -<Scp>MS</Scp>³. *Proteomics* **2013**, *13* (6), 964–973.
- (615) Mischerikow, N.; Altelaar, A. F. M.; Navarro, J. D.; Mohammed, S.; Heck, A. J. R. Comparative Assessment of Site Assignments in CID and Electron Transfer Dissociation Spectra of Phosphopeptides Discloses Limited Relocation of Phosphate Groups. *Molecular & Cellular Proteomics* **2010**, *9* (10), 2140–2148.
- (616) Kelstrup, C. D.; Hekmat, O.; Francavilla, C.; Olsen, J. V. Pinpointing Phosphorylation Sites: Quantitative Filtering and a Novel Site-Specific x-Ion Fragment. *J. Proteome Res.* **2011**, *10* (7), 2937–2948.
- (617) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villén, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-Scale Characterization of HeLa Cell Nuclear Phosphoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (33), 12130–12135.

- (618) Villén, J.; Beausoleil, S. A.; Gygi, S. P. Evaluation of the Utility of Neutral-loss-dependent MS3 Strategies in Large-scale Phosphorylation Analysis. *Proteomics* **2008**, *8* (21), 4444–4452.
- (619) Parker, B. L.; Yang, G.; Humphrey, S. J.; Chaudhuri, R.; Ma, X.; Peterman, S.; James, D. E. Targeted Phosphoproteomics of Insulin Signaling Using Data-Independent Acquisition Mass Spectrometry. *Sci. Signal.* **2015**, *8* (380), rs6.
- (620) Martinez-Val, A.; Bekker-Jensen, D. B.; Høgrebe, A.; Olsen, J. V. Data Processing and Analysis for DIA-Based Phosphoproteomics Using Spectronaut. In *Methods in Molecular Biology*; Springer, 2021; pp 95–107. DOI: 10.1007/978-1-0716-1641-3_6.
- (621) Lawrence, R. T.; Searle, B. C.; Llovet, A.; Villén, J. Plug-and-Play Analysis of the Human Phosphoproteome by Targeted High-Resolution Mass Spectrometry. *Nat Methods* **2016**, *13* (5), 431–434.
- (622) Høgrebe, A.; von Stechow, L.; Bekker-Jensen, D. B.; Weinert, B. T.; Kelstrup, C. D.; Olsen, J. V. Benchmarking Common Quantification Strategies for Large-Scale Phosphoproteomics. *Nat Commun* **2018**, *9* (1), 1045.
- (623) Singh, C.; Zampronia, C. G.; Creese, A. J.; Cooper, H. J. Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of <i>N</i>-Linked Glycoproteins. *J. Proteome Res.* **2012**, *11* (9), 4517–4525.
- (624) Nesvizhskii, A. I.; Aebersold, R. Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Mol Cell Proteomics* **2005**, *4* (10), 1419–1440.
- (625) Audain, E.; Uszkoreit, J.; Sachsenberg, T.; Pfeuffer, J.; Liang, X.; Hermjakob, H.; Sanchez, A.; Eisenacher, M.; Reinert, K.; Tabb, D. L.; Kohlbacher, O.; Perez-Riverol, Y. In-Depth Analysis of Protein Inference Algorithms Using Multiple Search Engines and Well-Defined Metrics. *Journal of Proteomics* **2017**, *150*, 170–182.
- (626) Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W. H.; Römpf, A.; Neumann, S.; Pizarro, A. D.; Montecchi-Palazzi, L.; Tasman, N.; Coleman, M.; Reisinger, F.; Souda, P.; Hermjakob, H.; Binz, P.-A.; Deutsch, E. W. mzML—a Community Standard for Mass Spectrometry Data. *Molecular & Cellular Proteomics* **2011**, *10* (1), R110.000133.
- (627) Deutsch, E. W. Mass Spectrometer Output File Format mzML. *Methods Mol Biol* **2010**, *604*, 319–331.
- (628) Deutsch, E. W. File Formats Commonly Used in Mass Spectrometry Proteomics. *Molecular & Cellular Proteomics* **2012**, *11* (12), 1612–1621.
- (629) Verheggen, K.; Ræder, H.; Berven, F. S.; Martens, L.; Barsnes, H.; Vaudel, M. Anatomy and Evolution of Database Search Engines—a Central Component of Mass Spectrometry Based Proteomic Workflows. *Mass Spectrometry Reviews* **2020**, *39* (3), 292–306.
- (630) 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.
- (631) 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.
- (632) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20* (18), 3551–3567.
- (633) Kim, S.; Pevzner, P. A. MS-GF+ Makes Progress Towards a Universal Database Search Tool for Proteomics. *Nat Commun* **2014**, *5* (1), 5277.
- (634) Craig, R.; Beavis, R. C. A Method for Reducing the Time Required to Match Protein Sequences with Tandem Mass Spectra. *Rapid Comm Mass Spectrometry* **2003**, *17* (20), 2310–2316.
- (635) Craig, R.; Beavis, R. C. TANDEM: Matching Proteins with Tandem Mass Spectra. *Bioinformatics* **2004**, *20* (9), 1466–1467.
- (636) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: An Open-source <Scp>MS</Scp>/<Scp>MS</Scp> Sequence Database Search Tool. *Proteomics* **2013**, *13* (1), 22–24.
- (637) 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. *Bioinformatics* **2010**, *26* (7), 966–968.
- (638) Tran, N. H.; Qiao, R.; Xin, L.; Chen, X.; Liu, C.; Zhang, X.; Shan, B.; Ghodsi, A.; Li, M. Deep Learning Enables de Novo Peptide Sequencing from Data-Independent-Acquisition Mass Spectrometry. *Nat Methods* **2019**, *16* (1), 63–66.
- (639) Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics* **2014**, *13* (9), 2513–2526.
- (640) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. *Nat Protoc* **2016**, *11* (12), 2301–2319.
- (641) Zimmer, J. S. D.; Monroe, M. E.; Qian, W.; Smith, R. D. Advances in Proteomics Data Analysis and Display Using an Accurate Mass and Time Tag Approach. *Mass Spectrometry Reviews* **2006**, *25* (3), 450–482.
- (642) Andreev, V. P.; Li, L.; Cao, L.; Gu, Y.; Rejtar, T.; Wu, S.-L.; Karger, B. L. A New Algorithm Using Cross-Assignment for Label-Free Quantitation with LC–LTQ-FT MS. *J. Proteome Res.* **2007**, *6* (6), 2186–2194.
- (643) Tsou, C.-C.; Tsai, C.-F.; Tsui, Y.-H.; Sudhir, P.-R.; Wang, Y.-T.; Chen, Y.-J.; Chen, J.-Y.; Sung, T.-Y.; Hsu, W.-L. IDEAL-Q, an Automated Tool for Label-Free Quantitation Analysis Using an Efficient Peptide Alignment Approach and Spectral Data Validation. *Molecular & Cellular Proteomics* **2010**, *9* (1), 131–144.
- (644) Mueller, L. N.; Rinner, O.; Schmidt, A.; Letarte, S.; Bodenmiller, B.; Brusniak, M.; Vitek, O.; Aebersold, R.; Müller, M. <i>SuperHirn</i> — a Novel Tool for High Resolution LC-MS-based Peptide/Protein Profiling. *Proteomics* **2007**, *7* (19), 3470–3480.
- (645) Lim, M. Y.; Paulo, J. A.; Gygi, S. P. Evaluating False Transfer Rates from the Match-Between-Runs Algorithm with a Two-Proteome Model. *J. Proteome Res.* **2019**, *18* (11), 4020–4026.
- (646) Yu, F.; Haynes, S. E.; Nesvizhskii, A. I. IonQuant Enables Accurate and Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs. *Molecular & Cellular Proteomics* **2021**, *20*, 100077.
- (647) Gessulat, S.; Schmidt, T.; Zolg, D. P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; Reimer, U.; Ehrlich, H.-C.; Aiche, S.; Kuster, B.; Wilhelm, M. Prosit: Proteome-Wide Prediction of Peptide Tandem Mass Spectra by Deep Learning. *Nat Methods* **2019**, *16* (6), 509–518.
- (648) Wang, M.; Wang, J.; Carver, J.; Pullman, B. S.; Cha, S. W.; Bandeira, N. Assembling the Community-Scale Discoverable Human Proteome. *Cell Syst* **2018**, *7* (4), 412–421.e5.
- (649) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R. OpenSWATH Enables Automated, Targeted Analysis of Data-Independent Acquisition MS Data. *Nat Biotechnol* **2014**, *32* (3), 219–223.
- (650) Cranney, C. W.; Meyer, J. G. CsoDIAQ Software for Direct Infusion Shotgun Proteome Analysis. *Anal. Chem.* **2021**, *93* (36), 12312–12319.
- (651) 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.
- (652) Tsou, C.-C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.-C.; Nesvizhskii, A. I. DIA-Umpire: Comprehensive Computational Framework for Data-Independent Acquisition Proteomics. *Nat Methods* **2015**, *12* (3), 258–264.
- (653) Ting, Y. S.; Egerton, J. D.; Bollinger, J. G.; Searle, B. C.; Payne, S. H.; Noble, W. S.; MacCoss, M. J. PECAN: Library-Free

Peptide Detection for Data-Independent Acquisition Tandem Mass Spectrometry Data. *Nat Methods* **2017**, *14* (9), 903–908.

(654) Searle, B. C.; Pino, L. K.; Egerton, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J. Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry. *Nat Commun* **2018**, *9* (1), 5128.

(655) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Miladinović, S. M.; Cheng, L.-Y.; Messner, S.; Ehrenberger, T.; Zanotelli, V.; Butscheid, Y.; Escher, C.; Vitek, O.; Rinner, O.; Reiter, L. Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Molecular & Cellular Proteomics* **2015**, *14* (5), 1400–1410.

(656) Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *J. Proteome Res.* **2008**, *7* (1), 29–34.

(657) Elias, J. E.; Gygi, S. P. Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry. *Nat Methods* **2007**, *4* (3), 207–214.

(658) Aggarwal, S.; Yadav, A. K. False Discovery Rate Estimation in Proteomics. In *Methods in Molecular Biology*; Springer: New York, 2016; pp 119–128. DOI: 10.1007/978-1-4939-3106-4_7.

(659) Levitsky, L. I.; Ivanov, M. V.; Lobas, A. A.; Gorshkov, M. V. Unbiased False Discovery Rate Estimation for Shotgun Proteomics Based on the Target-Decoy Approach. *J. Proteome Res.* **2017**, *16* (2), 393–397.

(660) Vaudel, M.; Burkhart, J. M.; Zahedi, R. P.; Oveland, E.; Berven, F. S.; Sickmann, A.; Martens, L.; Barsnes, H. PeptideShaker Enables Reanalysis of MS-Derived Proteomics Data Sets. *Nat Biotechnol* **2015**, *33* (1), 22–24.

(661) Farag, Y. M.; Horro, C.; Vaudel, M.; Barsnes, H. PeptideShaker Online: A User-Friendly Web-Based Framework for the Identification of Mass Spectrometry-Based Proteomics Data. *J. Proteome Res.* **2021**, *20* (12), 5419–5423.

(662) Keller, A.; Eng, J.; Zhang, N.; Li, X.; Aebersold, R. A Uniform Proteomics MS/MS Analysis Platform Utilizing Open XML File Formats. *Mol Syst Biol* **2005**, *1*, No. 2005.0017.

(663) Pedrioli, P. G. A. Trans-Proteomic Pipeline: A Pipeline for Proteomic Analysis. *Methods Mol Biol* **2010**, *604*, 213–238.

(664) Keller, A.; Shteynberg, D. Software Pipeline and Data Analysis for MS/MS Proteomics: The Trans-Proteomic Pipeline. *Methods Mol Biol* **2011**, *694*, 169–189.

(665) Shteynberg, D.; Deutsch, E. W.; Lam, H.; Eng, J. K.; Sun, Z.; Tasman, N.; Mendoza, L.; Moritz, R. L.; Aebersold, R.; Nesvizhskii, A. I. iProphet: Multi-Level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. *Mol Cell Proteomics* **2011**, *10* (12), No. M111.007690.

(666) Slagel, J.; Mendoza, L.; Shteynberg, D.; Deutsch, E. W.; Moritz, R. L. Processing Shotgun Proteomics Data on the Amazon Cloud with the Trans-Proteomic Pipeline. *Mol Cell Proteomics* **2015**, *14* (2), 399–404.

(667) Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Slagel, J.; Sun, Z.; Moritz, R. L. Trans-Proteomic Pipeline, a Standardized Data Processing Pipeline for Large-Scale Reproducible Proteomics Informatics. *Proteomics Clin Appl* **2015**, *9* (7-8), 745–754.

(668) Shteynberg, D.; Mendoza, L.; Hoopmann, M. R.; Sun, Z.; Schmidt, F.; Deutsch, E. W.; Moritz, R. L. reSpect: Software for Identification of High and Low Abundance Ion Species in Chimeric Tandem Mass Spectra. *J Am Soc Mass Spectrom* **2015**, *26* (11), 1837–1847.

(669) Hoopmann, M. R.; Winget, J. M.; Mendoza, L.; Moritz, R. L. StPeter: Seamless Label-Free Quantification with the Trans-Proteomic Pipeline. *J Proteome Res* **2018**, *17* (3), 1314–1320.

(670) Shteynberg, D. D.; Deutsch, E. W.; Campbell, D. S.; Hoopmann, M. R.; Kusebauch, U.; Lee, D.; Mendoza, L.; Midha, M. K.; Sun, Z.; Whetton, A. D.; Moritz, R. L. PTMProphet: Fast and Accurate Mass Modification Localization for the Trans-Proteomic Pipeline. *J Proteome Res* **2019**, *18* (12), 4262–4272.

(671) Deutsch, E. W.; Mendoza, L.; Shteynberg, D. D.; Hoopmann, M. R.; Sun, Z.; Eng, J. K.; Moritz, R. L. Trans-Proteomic Pipeline: Robust Mass Spectrometry-Based Proteomics Data Analysis Suite. *J Proteome Res* **2023**, *22* (2), 615–624.

(672) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal Chem* **2002**, *74* (20), 5383–5392.

(673) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Anal Chem* **2003**, *75* (17), 4646–4658.

(674) auf dem Keller, U.; Overall, C. M. CLIPPER: An Add-on to the Trans-Proteomic Pipeline for the Automated Analysis of TAILS N-Terminomics Data. *Biol Chem* **2012**, *393* (12), 1477–1483.

(675) Mitchell, C. J.; Kim, M.-S.; Na, C. H.; Pandey, A. PyQuant: A Versatile Framework for Analysis of Quantitative Mass Spectrometry Data. *Mol Cell Proteomics* **2016**, *15* (8), 2829–2838.

(676) Suni, V.; Suomi, T.; Tsubosaka, T.; Imanishi, S. Y.; Elo, L. L.; Corthals, G. L. SimPhospho: A Software Tool Enabling Confident Phosphosite Assignment. *Bioinformatics* **2018**, *34* (15), 2690–2692.

(677) Chen, C.-T.; Ko, C.-L.; Choong, W.-K.; Wang, J.-H.; Hsu, W.-L.; Sung, T.-Y. WinProphet: A User-Friendly Pipeline Management System for Proteomics Data Analysis Based on Trans-Proteomic Pipeline. *Anal Chem* **2019**, *91* (15), 9403–9406.

(678) Winkler, R. ProtyQuant: Comparing Label-Free Shotgun Proteomics Datasets Using Accumulated Peptide Probabilities. *J Proteomics* **2021**, *230*, No. 103985.

(679) He, S.; Chakraborty, R.; Ranganathan, S. Metaproteomic Analysis of an Oral Squamous Cell Carcinoma Dataset Suggests Diagnostic Potential of the Mycobiome. *Int J Mol Sci* **2023**, *24* (2), 1050.

(680) Lam, H.; Deutsch, E. W.; Eddes, J. S.; Eng, J. K.; Stein, S. E.; Aebersold, R. Building Consensus Spectral Libraries for Peptide Identification in Proteomics. *Nat Methods* **2008**, *5* (10), 873–875.

(681) mzSpecLib – HUPO Proteomics Standards Initiative. <https://www.psides.info/mzspeclib> (accessed 2024-05-07).

(682) Ma, B. Novor: Real-Time Peptide de Novo Sequencing Software. *J Am Soc Mass Spectrom* **2015**, *26* (11), 1885–1894.

(683) Yilmaz, M.; Fondrie, W. E.; Bittremieux, W.; Oh, S.; Noble, W. S. De Novo Mass Spectrometry Peptide Sequencing with a Transformer Model. *Proc. Mach. Learning Res.* **2022**, *1*.

(684) Riffle, M.; Hoopmann, M. R.; Jaschob, D.; Zhong, G.; Moritz, R. L.; MacCoss, M. J.; Davis, T. N.; Isoherranen, N.; Zelter, A. Discovery and Visualization of Uncharacterized Drug-Protein Adducts Using Mass Spectrometry. *Anal Chem* **2022**, *94* (8), 3501–3509.

(685) Martens, L.; et al. Quality Control in Proteomics. *Proteomics* **2011**, *11* (6), 1015–1016.

(686) Köcher, T.; Pichler, P.; Swart, R.; Mechtler, K. Quality Control in LC-MS/MS. *Proteomics* **2011**, *11* (6), 1026–1030.

(687) Bittremieux, W.; Tabb, D. L.; Impens, F.; Staes, A.; Timmerman, E.; Martens, L.; Laukens, K. Quality Control in Mass Spectrometry-based Proteomics. *Mass Spectrometry Reviews* **2018**, *37* (5), 697–711.

(688) Strauss, M. T.; Bludau, I.; Zeng, W.-F.; Voytik, E.; Ammar, C.; Schessner, J.; Ilango, R.; Gill, M.; Meier, F.; Willems, S.; Mann, M. AlphaPept, a Modern and Open Framework for MS-Based Proteomics. *Nat. Commun.* **2024**, *21*, 2168.

(689) Trachsel, C.; Panse, C.; Kockmann, T.; Wolski, W. E.; Grossmann, J.; Schlapbach, R. rawDiag: An R Package Supporting Rational LC-MS Method Optimization for Bottom-up Proteomics. *J. Proteome Res.* **2018**, *17* (8), 2908–2914.

(690) Kockmann, T.; Panse, C. The Rawrr R Package: Direct Access to Orbitrap Data and Beyond. *J. Proteome Res.* **2021**, *20* (4), 2028–2034.

(691) Morgenstern, D.; Barzilay, R.; Levin, Y. RawBeans: A Simple, Vendor-Independent, Raw-Data Quality-Control Tool. *J. Proteome Res.* **2021**, *20* (4), 2098–2104.

(692) Pichler, P.; Mazanek, M.; Dusberger, F.; Weinböck, L.; Huber, C. G.; Stingl, C.; Luidert, T. M.; Straube, W. L.; Köcher, T.

- Mechtler, K. SIMPATIQCO: A Server-Based Software Suite Which Facilitates Monitoring the Time Course of LC–MS Performance Metrics on Orbitrap Instruments. *J. Proteome Res.* **2012**, *11* (11), 5540–5547.
- (693) Stanfill, B. A.; Nakayasu, E. S.; Bramer, L. M.; Thompson, A. M.; Ansong, C. K.; Clauss, T. R.; Gritsenko, M. A.; Monroe, M. E.; Moore, R. J.; Orton, D. J.; Piehowski, P. D.; Schepmoes, A. A.; Smith, R. D.; Webb-Robertson, B.-J. M.; Metz, T. O. Quality Control Analysis in Real-Time (QC-ART): A Tool for Real-Time Quality Control Assessment of Mass Spectrometry-Based Proteomics Data. *Molecular & Cellular Proteomics* **2018**, *17* (9), 1824–1836.
- (694) Scheltema, R. A.; Mann, M. SprayQc: A Real-Time LC–MS/MS Quality Monitoring System To Maximize Uptime Using Off the Shelf Components. *J. Proteome Res.* **2012**, *11* (6), 3458–3466.
- (695) Taylor, R. M.; Dance, J.; Taylor, R. J.; Prince, J. T. MetriCulator: Quality Assessment for Mass Spectrometry-Based Proteomics. *Bioinformatics* **2013**, *29* (22), 2948–2949.
- (696) Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weissner, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.-C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.; Choudhary, J. S.; Malmström, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O. OpenMS: A Flexible Open-Source Software Platform for Mass Spectrometry Data Analysis. *Nat Methods* **2016**, *13* (9), 741–748.
- (697) 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. *Bioinformatics* **2014**, *30* (17), 2524–2526.
- (698) Dogu, E.; Mohammad-Taheri, S.; Abbatiello, S. E.; Bereman, M. S.; MacLean, B.; Schilling, B.; Vitek, O. MSstatsQC: Longitudinal System Suitability Monitoring and Quality Control for Targeted Proteomic Experiments. *Molecular & Cellular Proteomics* **2017**, *16* (7), 1335–1347.
- (699) Bielow, C.; Mastrobuoni, G.; Kempa, S. Proteomics Quality Control: Quality Control Software for MaxQuant Results. *J. Proteome Res.* **2016**, *15* (3), 777–787.
- (700) Quast, J.-P.; Schuster, D.; Picotti, P. Protti: An R Package for Comprehensive Data Analysis of Peptide- and Protein-Centric Bottom-up Proteomics Data. *Bioinformatics Advances* **2022**, *2* (1), vbab041.
- (701) Lualdi, M.; Fasano, M. Statistical Analysis of Proteomics Data: A Review on Feature Selection. *Journal of Proteomics* **2019**, *198*, 18–26.
- (702) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Prote)omics Data. *Nat Methods* **2016**, *13* (9), 731–740.
- (703) Dubois, E.; Galindo, A. N.; Dayon, L.; Cominetti, O. Assessing Normalization Methods in Mass Spectrometry-Based Proteome Profiling of Clinical Samples. *Biosystems* **2022**, *215*–216, No. 104661.
- (704) Callister, S. J.; Barry, R. C.; Adkins, J. N.; Johnson, E. T.; Qian, W.; Webb-Robertson, B.-J. M.; Smith, R. D.; Lipton, M. S. Normalization Approaches for Removing Systematic Biases Associated with Mass Spectrometry and Label-Free Proteomics. *J. Proteome Res.* **2006**, *5* (2), 277–286.
- (705) Välikangas, T.; Suomi, T.; Elo, L. L. A Systematic Evaluation of Normalization Methods in Quantitative Label-Free Proteomics. *Brief Bioinform* **2016**, No. bbw095.
- (706) Karpievitch, Y. V.; Dabney, A. R.; Smith, R. D. Normalization and Missing Value Imputation for Label-Free LC-MS Analysis. *BMC Bioinformatics* **2012**, *13* (S16), S5.
- (707) Chawade, A.; Alexandersson, E.; Levander, F. Normalyzer: A Tool for Rapid Evaluation of Normalization Methods for Omics Data Sets. *J. Proteome Res.* **2014**, *13* (6), 3114–3120.
- (708) Molania, R.; Foroutan, M.; Gagnon-Bartsch, J. A.; Gandolfo, L. C.; Jain, A.; Sinha, A.; Olshansky, G.; Dobrovic, A.; Papenfuss, A. T.; Speed, T. P. Removing Unwanted Variation from Large-Scale RNA Sequencing Data with PRPS. *Nat Biotechnol* **2023**, *41* (1), 82–95.
- (709) Čuklina, J.; Lee, C. H.; Williams, E. G.; Sajic, T.; Collins, B. C.; Rodríguez Martínez, M.; Sharma, V. S.; Wendt, F.; Goetze, S.; Keele, G. R.; Wollscheid, B.; Aebersold, R.; Pedrioli, P. G. A. Diagnostics and Correction of Batch Effects in Large-scale Proteomic Studies: A Tutorial. *Molecular Systems Biology* **2021**, *17* (8), e10240.
- (710) Evans, C.; Hardin, J.; Stoebe, D. M. Selecting Between-Sample RNA-Seq Normalization Methods from the Perspective of Their Assumptions. *Briefings in Bioinformatics* **2018**, *19* (5), 776–792.
- (711) Bolstad, B. M.; Irizarry, R. A.; Åstrand, M.; Speed, T. P. A Comparison of Normalization Methods for High Density oligonucleotide Array Data Based on Variance and Bias. *Bioinformatics* **2003**, *19* (2), 185–193.
- (712) Risso, D.; Ngai, J.; Speed, T. P.; Dudoit, S. Normalization of RNA-Seq Data Using Factor Analysis of Control Genes or Samples. *Nat Biotechnol* **2014**, *32* (9), 896–902.
- (713) Wang, M.; Jiang, L.; Jian, R.; Chan, J. Y.; Liu, Q.; Snyder, M. P.; Tang, H. RobNorm: Model-Based Robust Normalization Method for Labeled Quantitative Mass Spectrometry Proteomics Data. *Bioinformatics* **2021**, *37* (6), 815–821.
- (714) Karpievitch, Y. V.; Taverner, T.; Adkins, J. N.; Callister, S. J.; Anderson, G. A.; Smith, R. D.; Dabney, A. R. Normalization of Peak Intensities in Bottom-up MS-Based Proteomics Using Singular Value Decomposition. *Bioinformatics* **2009**, *25* (19), 2573–2580.
- (715) Fan, S.; Kind, T.; Cajka, T.; Hazen, S. L.; Tang, W. H. W.; Kaddurah-Daouk, R.; Irvin, M. R.; Arnett, D. K.; Barupal, D. K.; Fiehn, O. Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. *Anal. Chem.* **2019**, *91* (5), 3590–3596.
- (716) Kulima, K.; Nilsson, A.; Scholz, B.; Rossbach, U. L.; Fälth, M.; Andrén, P. E. Development and Evaluation of Normalization Methods for Label-Free Relative Quantification of Endogenous Peptides. *Molecular & Cellular Proteomics* **2009**, *8* (10), 2285–2295.
- (717) Willforss, J.; Chawade, A.; Levander, F. NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. *J. Proteome Res.* **2019**, *18* (2), 732–740.
- (718) Krümer, E.; Bludau, I.; Strauss, M. T.; Mann, M. AlphaPeptStats: An Open-Source Python Package for Automated and Scalable Statistical Analysis of Mass Spectrometry-Based Proteomics. *Bioinformatics* **2023**, *39* (8), btad461.
- (719) Karpievitch, Y.; Stanley, J.; Taverner, T.; Huang, J.; Adkins, J. N.; Ansong, C.; Heffron, F.; Metz, T. O.; Qian, W.-J.; Yoon, H.; Smith, R. D.; Dabney, A. R. A Statistical Framework for Protein Quantitation in Bottom-up MS-Based Proteomics. *Bioinformatics* **2009**, *25* (16), 2028–2034.
- (720) Luo, R.; Colangelo, C. M.; Sessa, W. C.; Zhao, H. Bayesian Analysis of iTRAQ Data with Nonrandom Missingness: Identification of Differentially Expressed Proteins. *Stat Biosci* **2009**, *1* (2), 228–245.
- (721) Stekhoven, D. J.; Bühlmann, P. MissForest—Non-Parametric Missing Value Imputation for Mixed-Type Data. *Bioinformatics* **2011**, *28* (1), 112–118.
- (722) Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R. B. Missing Value Estimation Methods for DNA Microarrays. *Bioinformatics* **2001**, *17* (6), 520–525.
- (723) Stacklies, W.; Redestig, H.; Scholz, M.; Walther, D.; Selbig, J. pcaMethods—a Bioconductor Package Providing PCA Methods for Incomplete Data. *Bioinformatics* **2007**, *23* (9), 1164–1167.
- (724) Wei, R.; Wang, J.; Su, M.; Jia, E.; Chen, S.; Chen, T.; Ni, Y. Missing Value Imputation Approach for Mass Spectrometry-Based Metabolomics Data. *Sci Rep* **2018**, *8* (1), 663.
- (725) Kokla, M.; Virtanen, J.; Kolehmainen, M.; Paananen, J.; Hanhineva, K. Random Forest-Based Imputation Outperforms Other Methods for Imputing LC-MS Metabolomics Data: A Comparative Study. *BMC Bioinformatics* **2019**, *20* (1), 663.

- (726) Jin, L.; Bi, Y.; Hu, C.; Qu, J.; Shen, S.; Wang, X.; Tian, Y. A Comparative Study of Evaluating Missing Value Imputation Methods in Label-Free Proteomics. *Sci Rep* **2021**, *11* (1), 1760.
- (727) Goh, W. W. B.; Wang, W.; Wong, L. Why Batch Effects Matter in Omics Data, and How to Avoid Them. *Trends in Biotechnology* **2017**, *35* (6), 498–507.
- (728) Leek, J. T.; Scharpf, R. B.; Bravo, H. C.; Simcha, D.; Langmead, B.; Johnson, W. E.; Geman, D.; Baggerly, K.; Irizarry, R. A. Tackling the Widespread and Critical Impact of Batch Effects in High-Throughput Data. *Nat Rev Genet* **2010**, *11* (10), 733–739.
- (729) Hu, J. The Importance of Experimental Design in Proteomic Mass Spectrometry Experiments: Some Cautionary Tales. *Briefings in Functional Genomics and Proteomics* **2005**, *3* (4), 322–331.
- (730) Johnson, W. E.; Li, C.; Rabinovic, A. Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics* **2007**, *8* (1), 118–127.
- (731) Chen, C.; Grennan, K.; Badner, J.; Zhang, D.; Gershon, E.; Jin, L.; Liu, C. Removing Batch Effects in Analysis of Expression Microarray Data: An Evaluation of Six Batch Adjustment Methods. *PLoS ONE* **2011**, *6* (2), e17238.
- (732) Gagnon-Bartsch, J. A.; Speed, T. P. Using Control Genes to Correct for Unwanted Variation in Microarray Data. *Biostatistics* **2012**, *13* (3), 539–552.
- (733) Haghverdi, L.; Lun, A. T. L.; Morgan, M. D.; Marioni, J. C. Batch Effects in Single-Cell RNA-Sequencing Data Are Corrected by Matching Mutual Nearest Neighbors. *Nat Biotechnol* **2018**, *36* (5), 421–427.
- (734) Hie, B.; Bryson, B.; Berger, B. Efficient Integration of Heterogeneous Single-Cell Transcriptomes Using Scanorama. *Nat Biotechnol* **2019**, *37* (6), 685–691.
- (735) Schessner, J. P.; Voytik, E.; Bludau, I. A Practical Guide to Interpreting and Generating Bottom-up Proteomics Data Visualizations. *Proteomics* **2022**, *22* (8), 2100103.
- (736) Mann, M.; Kumar, C.; Zeng, W.-F.; Strauss, M. T. Artificial Intelligence for Proteomics and Biomarker Discovery. *Cell Systems* **2021**, *12* (8), 759–770.
- (737) Swan, A. L.; Mobasheri, A.; Allaway, D.; Liddell, S.; Bacardit, J. Application of Machine Learning to Proteomics Data: Classification and Biomarker Identification in Postgenomics Biology. *OMICS: A Journal of Integrative Biology* **2013**, *17* (12), 595–610.
- (738) Hernández, B.; Pennington, S. R.; Parnell, A. C. Bayesian Methods for Proteomic Biomarker Development. *EuPA Open Proteomics* **2015**, *9*, 54–64.
- (739) Kuschner, K. W.; Malyarenko, D. I.; Cooke, W. E.; Cazares, L. H.; Semmes, O.; Tracy, E. R. A Bayesian Network Approach to Feature Selection in Mass Spectrometry Data. *BMC Bioinformatics* **2010**, *11* (1), 177.
- (740) Desaire, H.; Go, E. P.; Hua, D. Advances, Obstacles, and Opportunities for Machine Learning in Proteomics. *Cell Reports Physical Science* **2022**, *3* (10), No. 101069.
- (741) Hicks, S. A.; Strümke, I.; Thambawita, V.; Hammou, M.; Riegler, M. A.; Halvorsen, P.; Parasa, S. On Evaluation Metrics for Medical Applications of Artificial Intelligence. *Sci Rep* **2022**, *12* (1), 5979.
- (742) Khan, M. J.; Desaire, H.; Lopez, O. L.; Kamboh, M. I.; Robinson, R. A. S. Why Inclusion Matters for Alzheimer's Disease Biomarker Discovery in Plasma. *JAD* **2021**, *79* (3), 1327–1344.
- (743) Desaire, H. How (Not) to Generate a Highly Predictive Biomarker Panel Using Machine Learning. *J. Proteome Res.* **2022**, *21* (9), 2071–2074.
- (744) Ransohoff, D. F. Rules of Evidence for Cancer Molecular-Marker Discovery and Validation. *Nat Rev Cancer* **2004**, *4* (4), 309–314.
- (745) Kim, M.; Hwang, K.-B. An Empirical Evaluation of Sampling Methods for the Classification of Imbalanced Data. *PLoS ONE* **2022**, *17* (7), e0271260.
- (746) van den Goorbergh, R.; van Smeden, M.; Timmerman, D.; Van Calster, B. The Harm of Class Imbalance Corrections for Risk Prediction Models: Illustration and Simulation Using Logistic Regression. *Journal of the American Medical Informatics Association* **2022**, *29* (9), 1525–1534.
- (747) Movahedi, F.; Antaki, J. F. Limitation of ROC in Evaluation of Classifiers for Imbalanced Data. *The Journal of Heart and Lung Transplantation* **2021**, *40* (4), S413.
- (748) Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative Than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10* (3), e0118432.
- (749) Holt, C.; Yandell, M. MAKER2: An Annotation Pipeline and Genome-Database Management Tool for Second-Generation Genome Projects. *BMC Bioinformatics* **2011**, *12*, 491.
- (750) The NCBI Eukaryotic Genome Annotation Pipeline. https://www.ncbi.nlm.nih.gov/refseq/annotation_euk/process/ (accessed 2024-05-07).
- (751) UniProt. <https://www.uniprot.org/help/biocuration> (accessed 2024-05-07).
- (752) https://www.psdev.info/sites/default/files/2018-03/MIAPE_MSI_1.1.pdf.
- (753) Martínez-Bartolomé, S.; Deutsch, E. W.; Binz, P.-A.; Jones, A. R.; Eisenacher, M.; Mayer, G.; Campos, A.; Canals, F.; Bech-Serra, J.-J.; Carrascal, M.; Gay, M.; Paradelo, A.; Navajas, R.; Marcella, M.; Hernández, M. L.; Gutiérrez-Blázquez, M. D.; Velarde, L. F. C.; Aloria, K.; Beaskoetxea, J.; Medina-Aunon, J. A.; Albar, J. P. Guidelines for Reporting Quantitative Mass Spectrometry Based Experiments in Proteomics. *J Proteomics* **2013**, *95*, 84–88.
- (754) Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L.-S. L. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res* **2004**, *32*, D115.
- (755) Lussi, Y. C.; Magrane, M.; Martin, M. J.; Orchard, S. Searching and Navigating UniProt Databases. *Current Protocols* **2023**, *3* (3), 1.
- (756) Bowler-Barnett, E. H.; Fan, J.; Luo, J.; Magrane, M.; Martin, M. J.; Orchard, S. UniProt and Mass Spectrometry-Based Proteomics—A 2-Way Working Relationship. *Molecular & Cellular Proteomics* **2023**, *22* (8), 100591.
- (757) Bateman, A.; Martin, M.-J.; Orchard, S.; Magrane, M.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E. H.; Britto, R.; Bye-A-Jee, H.; Cukura, A.; Denny, P.; Dogan, T.; Ebenezer, T.; Fan, J.; Garmiri, P.; da Costa Gonzales, L. J.; Hatton-Ellis, E.; Hussein, A.; Ignatchenko, A.; Insana, G.; Ishtiaq, R.; Joshi, V.; Jyothi, D.; Kandasamy, S.; Lock, A.; Luciani, A.; Lugalic, M.; Luo, J.; Lussi, Y.; MacDougall, A.; Madeira, F.; Mahmoudy, M.; Mishra, A.; Moulang, K.; Nightingale, A.; Pundir, S.; Qi, G.; Raj, S.; Raposo, P.; Rice, D. L.; Saidi, R.; Santos, R.; Speretta, E.; Stephenson, J.; Tototo, P.; Turner, E.; Tyagi, N.; Vasudev, P.; Warner, K.; Watkins, X.; Zaru, R.; Zellner, H.; Bridge, A. J.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A. H.; Axelsen, K. B.; Bansal, P.; Baratin, D.; Batista Neto, T. M.; Blatter, M.-C.; Bolleman, J. T.; Boutet, E.; Breuza, L.; Gil, B. C.; Casals-Casas, C.; Echioukh, K. C.; Coudert, E.; Cucho, B.; de Castro, E.; Estreicher, A.; Famiglietti, M. L.; Feuerhahn, M.; Gasteiger, E.; Gaudet, P.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz, N.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Kerhornou, A.; Le Mercier, P.; Lieberherr, D.; Masson, P.; Morgat, A.; Muthukrishnan, V.; Paesano, S.; Pedruzzi, I.; Pilboud, S.; Pourcel, L.; Poux, S.; Pozzato, M.; Pruess, M.; Redaschi, N.; Rivoire, C.; Sigrist, C. J. A.; Sonesson, K.; Sundaram, S.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C.; Chen, Y.; Huang, H.; Laiho, K.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Q.; Wang, Y.; Zhang, J. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **2022**, *51* (D1), D523–D531.
- (758) UniProt. https://www.uniprot.org/help/sequence_origin (accessed 2024-05-07).
- (759) UniProt. https://www.uniprot.org/help/manual_curation (accessed 2024-05-07).
- (760) Wilmarth, P. Pwilmart/Fasta_utilities, 2023. https://github.com/pwilmart/fasta_utilities (accessed 2024-05-07).
- (761) Shortreed, M. R.; Wenger, C. D.; Frey, B. L.; Sheynkman, G. M.; Scaff, M.; Keller, M. P.; Attie, A. D.; Smith, L. M. Global

Identification of Protein Post-Translational Modifications in a Single-Pass Database Search. *J Proteome Res* **2015**, *14* (11), 4714–4720.

(762) *Datasets*. NCBI. <https://www.ncbi.nlm.nih.gov/datasets> (accessed 2024-05-07).

(763) *Eukaryotic genomes annotated at NCBI*. https://www.ncbi.nlm.nih.gov/refseq/annotation_euk/all/ (accessed 2024-05-07).

(764) *Prokaryotic RefSeq Genomes*. <https://www.ncbi.nlm.nih.gov/refseq/about/prokaryotes/> (accessed 2024-05-07).

(765) *Eukaryotic RefSeq Genome Annotation Status*. https://www.ncbi.nlm.nih.gov/refseq/annotation_euk/status/ (accessed 2024-05-07).

(766) *Training & Tutorials - Site Guide - NCBI*. <https://www.ncbi.nlm.nih.gov/guide/training-tutorials/> (accessed 2024-05-07).

(767) Heck, M.; Neely, B. A. Proteomics in Non-Model Organisms: A New Analytical Frontier. *J Proteome Res* **2020**, *19* (9), 3595–3606.

(768) *Ensembl genome browser 111*. <https://useast.ensembl.org/index.html> (accessed 2024-05-07).

(769) *Ensembl Genomes*. <http://ensemblgenomes.org/> (accessed 2024-05-07).

(770) Keller, B. O.; Sui, J.; Young, A. B.; Whittall, R. M. Interferences and Contaminants Encountered in Modern Mass Spectrometry. *Anal Chim Acta* **2008**, *627* (1), 71–81.

(771) *cRAP protein sequences*. <https://www.thegpm.org/crap/> (accessed 2024-05-07).

(772) da Veiga Leprevost, F.; Haynes, S. E.; Avtonomov, D. M.; Chang, H.-Y.; Shanmugam, A. K.; Mellacheruvu, D.; Kong, A. T.; Nesvizhskii, A. I. Philosopher: A Versatile Toolkit for Shotgun Proteomics Data Analysis. *Nat Methods* **2020**, *17* (9), 869–870.

(773) Frankenfield, A. M.; Ni, J.; Ahmed, M.; Hao, L. Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics. *J. Proteome Res.* **2022**, *21*, 2104–2113.

(774) Bunkenborg, J.; García, G. E.; Paz, M. I. P.; Andersen, J. S.; Molina, H. The Minotaur Proteome: Avoiding Cross-Species Identifications Deriving from Bovine Serum in Cell Culture Models. *Proteomics* **2010**, *10* (16), 3040–3044.

(775) Lehrich, B. M.; Liang, Y.; Fiandaca, M. S. Foetal Bovine Serum Influence on in Vitro Extracellular Vesicle Analyses. *J Extracell Vesicles* **2021**, *10* (3), No. e12061.

(776) Mellacheruvu, D.; Wright, Z.; Couzens, A. L.; Lambert, J.-P.; St-Denis, N. A.; 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.; Goudreau, 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.

(777) Shin, J.; Kwon, Y.; Lee, S.; Na, S.; Hong, E. Y.; Ju, S.; Jung, H.-G.; Kaushal, P.; Shin, S.; Back, J. H.; Choi, S. Y.; Kim, E. H.; Lee, S. J.; Park, Y. E.; Ahn, H.-S.; Ahn, Y.; Kabir, M. H.; Park, S.-J.; Yang, W. S.; Yeom, J.; Bang, O. Y.; Ha, C.-W.; Lee, J.-W.; Kang, U.-B.; Kim, H.-J.; Park, K.-S.; Lee, J. E.; Lee, J. E.; Kim, J. Y.; Kim, K. P.; Kim, Y.; Hirano, H.; Yi, E. C.; Cho, J.-Y.; Paek, E.; Lee, C. Common Repository of FBS Proteins (cRFP) To Be Added to a Search Database for Mass Spectrometric Analysis of Cell Secretome. *J Proteome Res* **2019**, *18* (10), 3800–3806.

(778) Lin, A.; Plubell, D. L.; Keich, U.; Noble, W. S. Accurately Assigning Peptides to Spectra When Only a Subset of Peptides Are Relevant. *J Proteome Res* **2021**, *20* (8), 4153–4164.

(779) Keich, U.; Tamura, K.; Noble, W. S. Averaging Strategy To Reduce Variability in Target-Decoy Estimates of False Discovery Rate. *J Proteome Res* **2019**, *18* (2), 585–593.

(780) Noble, W. S. Mass Spectrometrists Should Search Only for Peptides They Care About. *Nat Methods* **2015**, *12* (7), 605–608.

(781) Kumar, D.; Yadav, A. K.; Dash, D. Choosing an Optimal Database for Protein Identification from Tandem Mass Spectrometry Data. *Methods Mol Biol* **2017**, *1549*, 17–29.

(782) Jones, P. PRIDE: A Public Repository of Protein and Peptide Identifications for the Proteomics Community. *Nucleic Acids Research* **2006**, *34* (90001), D659–D663.

(783) Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S.; Brazma, A.; Vizcaino, J. The PRIDE Database Resources in 2022: A Hub for Mass Spectrometry-Based Proteomics Evidences. *Nucleic Acids Research* **2021**, *50* (D1), D543–D552.

(784) *Chorus - Home*. <https://chorusproject.org/pages/index.html> (accessed 2024-05-07).

(785) Deutsch, E. W.; Eng, J. K.; Zhang, H.; King, N. L.; Nesvizhskii, A. I.; Lin, B.; Lee, H.; Yi, E. C.; Ossola, R.; Aebersold, R. Human Plasma PeptideAtlas. *Proteomics* **2005**, *5* (13), 3497–3500.

(786) Desiere, F.; Deutsch, E. W.; King, N. L.; Nesvizhskii, A. I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S. N.; Aebersold, R. The PeptideAtlas Project. *Nucleic Acids Res* **2006**, *34* (Database issue), D655–8.

(787) Farrah, T.; Deutsch, E. W.; Kreisberg, R.; Sun, Z.; Campbell, D. S.; Mendoza, L.; Kusebauch, U.; Brusniak, M.-Y.; Hüttenhain, R.; Schiess, R.; Selevsek, N.; Aebersold, R.; Moritz, R. L. PASSEL: The PeptideAtlas SRMexperiment Library. *Proteomics* **2012**, *12* (8), 1170–1175.

(788) Kusebauch, U.; Deutsch, E. W.; Campbell, D. S.; Sun, Z.; Farrah, T.; Moritz, R. L. Using PeptideAtlas, SRMATlas, and PASSEL: Comprehensive Resources for Discovery and Targeted Proteomics. *Curr Protoc Bioinformatics* **2014**, *46*, 13.25.1–13.25.28.

(789) Kusebauch, U.; Campbell, D. S.; Deutsch, E. W.; Chu, C. S.; Spicer, D. A.; Brusniak, M.-Y.; Slagel, J.; Sun, Z.; Stevens, J.; Grimes, B.; Shteynberg, D.; Hoopmann, M. R.; Blattmann, P.; Ratushny, A. V.; Rinner, O.; Picotti, P.; Carapito, C.; Huang, C.-Y.; Kapousouz, M.; Lam, H.; Tran, T.; Demir, E.; Aitchison, J. D.; Sander, C.; Hood, L.; Aebersold, R.; Moritz, R. L. Human SRMATlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* **2016**, *166* (3), 766–778.

(790) Deutsch, E. W.; Orchard, S.; Binz, P.-A.; Bittremieux, W.; Eisenacher, M.; Hermjakob, H.; Kawano, S.; Lam, H.; Mayer, G.; Menschaert, G.; Perez-Riverol, Y.; Salek, R. M.; Tabb, D. L.; Tenzer, S.; Vizcaino, J. A.; Walzer, M.; Jones, A. R. Proteomics Standards Initiative: Fifteen Years of Progress and Future Work. *J. Proteome Res.* **2017**, *16* (12), 4288–4298.

(791) Zolg, D. P.; Wilhelm, M.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Delanghe, B.; Bailey, D. J.; Gessulat, S.; Ehrlich, H.-C.; Weininger, M.; Yu, P.; Schlegl, J.; Kramer, K.; Schmidt, T.; Kusebauch, U.; Deutsch, E. W.; Aebersold, R.; Moritz, R. L.; Wenschuh, H.; Moehring, T.; Aiche, S.; Huhmer, A.; Reimer, U.; Kuster, B. Building ProteomeTools Based on a Complete Synthetic Human Proteome. *Nat Methods* **2017**, *14* (3), 259–262.

(792) Ma, J.; Chen, T.; Wu, S.; Yang, C.; Bai, M.; Shu, K.; Li, K.; Zhang, G.; Jin, Z.; He, F.; Hermjakob, H.; Zhu, Y. iProX: An Integrated Proteome Resource. *Nucleic Acids Research* **2018**, *47* (D1), D1211–D1217.

(793) Sharma, V.; Eckels, J.; Taylor, G. K.; Shulman, N. J.; Stergachis, A. B.; Joyner, S. A.; Yan, P.; Whiteaker, J. R.; Halusa, G. N.; Schilling, B.; Gibson, B. W.; Colangelo, C. M.; Paulovich, A. G.; Carr, S. A.; Jaffe, J. D.; MacCoss, M. J.; MacLean, B. Panorama: A Targeted Proteomics Knowledge Base. *J. Proteome Res.* **2014**, *13* (9), 4205–4210.

(794) Low, T. Y.; Syafruddin, S. E.; Mohtar, M. A.; Vellaichamy, A.; A Rahman, N. S.; Pung, Y.-F.; Tan, C. S. H. Recent Progress in Mass Spectrometry-Based Strategies for Elucidating Protein-Protein Interactions. *Cell Mol Life Sci* **2021**, *78* (13), 5325–5339.

(795) Del Toro, N.; Shrivastava, A.; Ragueneau, E.; Meldal, B.; Combe, C.; Barrera, E.; Perfetto, L.; How, K.; Ratan, P.; Shirodkar, G.; Lu, O.; Mészáros, B.; Watkins, X.; Pundir, S.; Licata, L.; Iannuccelli, M.; Pellegrini, M.; Martin, M. J.; Panni, S.; Duesbury, M. J.

- Vallet, S. D.; Rappsilber, J.; Ricard-Blum, S.; Cesareni, G.; Salwinski, L.; Orchard, S.; Porras, P.; Panneerselvam, K.; Hermjakob, H. The IntAct Database: Efficient Access to Fine-Grained Molecular Interaction Data. *Nucleic Acids Res* **2022**, *50* (D1), D648–D653.
- (796) Lees, J. G.; Heriche, J. K.; Morilla, I.; Ranea, J. A.; Orengo, C. A. Systematic Computational Prediction of Protein Interaction Networks. *Phys Biol* **2011**, *8* (3), 035008.
- (797) Szklarczyk, D.; Gable, A. L.; Nastou, K. C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N. T.; Legeay, M.; Fang, T.; Bork, P.; Jensen, L. J.; von Mering, C. The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets. *Nucleic Acids Res* **2021**, *49* (D1), D605–D612.
- (798) Warde-Farley, D.; Donaldson, S. L.; Comes, O.; Zuberi, K.; Badrawi, R.; Chao, P.; Franz, M.; Grouios, C.; Kazi, F.; Lopes, C. T.; Maitland, A.; Mostafavi, S.; Montojo, J.; Shao, Q.; Wright, G.; Bader, G. D.; Morris, Q. The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function. *Nucleic Acids Research* **2010**, *38* (suppl 2), W214–W220.
- (799) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13* (11), 2498–2504.
- (800) Jalili, M.; Salehzadeh-Yazdi, A.; Gupta, S.; Wolkenhauer, O.; Yaghmaie, M.; Resendis-Antonio, O.; Alimoghaddam, K. Evolution of Centrality Measurements for the Detection of Essential Proteins in Biological Networks. *Front Physiol* **2016**, *7*, 375.
- (801) Ashtiani, M.; Salehzadeh-Yazdi, A.; Razaghi-Moghadam, Z.; Hennig, H.; Wolkenhauer, O.; Mirzaie, M.; Jafari, M. A Systematic Survey of Centrality Measures for Protein-Protein Interaction Networks. *BMC Syst Biol* **2018**, *12* (1), 80.
- (802) He, X.; Zhang, J. Why Do Hubs Tend to Be Essential in Protein Networks? *PLoS Genet* **2006**, *2* (6), e88.
- (803) Hahn, M. W.; Kern, A. D. Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks. *Mol Biol Evol* **2004**, *22* (4), 803–806.
- (804) EMBL-EBI. Closeness centrality | Network analysis of protein interaction data. <https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/building-and-analysing-ppins/topological-ppin-analysis/closeness-centrality/> (accessed 2024-05-07).
- (805) Yu, H.; Kim, P. M.; Sprecher, E.; Trifonov, V.; Gerstein, M. The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. *PLoS Comput Biol* **2007**, *3* (4), e59.
- (806) Li, X.; Wu, M.; Kwok, C.-K.; Ng, S.-K. Computational Approaches for Detecting Protein Complexes from Protein Interaction Networks: A Survey. *BMC Genomics* **2010**, *11* Suppl 1 (Suppl 1), S3.
- (807) Brohée, S.; van Helden, J. Evaluation of Clustering Algorithms for Protein-Protein Interaction Networks. *BMC Bioinformatics* **2006**, *7*, 488.
- (808) Bader, G. D.; Hogue, C. W. V. An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks. *BMC Bioinformatics* **2003**, *4*, 2.
- (809) Piovesan, D.; Giolito, M.; Ferrari, C.; Tosatto, S. C. E. Protein Function Prediction Using Guilty by Association from Interaction Networks. *Amino Acids* **2015**, *47* (12), 2583–2592.
- (810) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13* (11), 2498–2504.
- (811) Doncheva, N. T.; Morris, J. H.; Holze, H.; Kirsch, R.; Nastou, K. C.; Cuesta-Astro, Y.; Rattei, T.; Szklarczyk, D.; von Mering, C.; Jensen, L. J. Cytoscape stringApp 2.0: Analysis and Visualization of Heterogeneous Biological Networks. *J Proteome Res* **2022**, *22* (2), 637–646.
- (812) *Protein Bioinformatics*; Wu, C. H., Arighi, C. N., Ross, K. E., Eds.; Springer: New York, 2017. DOI: 10.1007/978-1-4939-6783-4.
- (813) Halder, A.; Verma, A.; Biswas, D.; Srivastava, S. Recent Advances in Mass-Spectrometry Based Proteomics Software, Tools and Databases. *Drug Discovery Today: Technologies* **2021**, *39*, 69–79.
- (814) Deutsch, E. W.; Lam, H.; Aebersold, R. Data Analysis and Bioinformatics Tools for Tandem Mass Spectrometry in Proteomics. *Physiological Genomics* **2008**, *33* (1), 18–25.
- (815) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat Genet* **2000**, *25* (1), 25–29.
- (816) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **2000**, *28* (1), 27–30.
- (817) Jassal, B.; Matthews, L.; Viteri, G.; Gong, C.; Lorente, P.; Fabregat, A.; Sidiropoulos, K.; Cook, J.; Gillespie, M.; Haw, R.; Loney, F.; May, B.; Milacic, M.; Rothfels, K.; Sevilla, C.; Shamovsky, V.; Shorsler, S.; Varusai, T.; Weiser, J.; Wu, G.; Stein, L.; Hermjakob, H.; D'Eustachio, P. The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **2020**, *48* (D1), D498–D503.
- (818) Hornbeck, P. V.; Kornhauser, J. M.; Latham, V.; Murray, B.; Nandhikonda, V.; Nord, A.; Skrzypek, E.; Wheeler, T.; Zhang, B.; Gnad, F. 15 Years of PhosphoSitePlus®: Integrating Post-Translationally Modified Sites, Disease Variants and Isoforms. *Nucleic Acids Res* **2019**, *47* (D1), D433–D441.
- (819) Selevsek, N.; Caiment, F.; Nudischer, R.; Gmuender, H.; Agarkova, I.; Atkinson, F. L.; Bachmann, I.; Baier, V.; Barel, G.; Bauer, C.; Boerno, S.; Bosc, N.; Clayton, O.; Cordes, H.; Deeb, S.; Gotta, S.; Guye, P.; Hersey, A.; Hunter, F. M. I.; Kunz, L.; Lewalle, A.; Lienhard, M.; Merken, J.; Minguet, J.; Oliveira, B.; Pluess, C.; Sarkans, U.; Schrooders, Y.; Schuchhardt, J.; Smit, I.; Thiel, C.; Timmermann, B.; Verheijen, M.; Wittenberger, T.; Wolski, W.; Zerck, A.; Heymans, S.; Kuelper, L.; Roth, A.; Schlappbach, R.; Niederer, S.; Herwig, R.; Kleinjans, J. Network Integration and Modelling of Dynamic Drug Responses at Multi-Omics Levels. *Commun Biol* **2020**, *3* (1), 573.
- (820) Gillespie, M.; Jassal, B.; Stephan, R.; Milacic, M.; Rothfels, K.; Senff-Ribeiro, A.; Griss, J.; Sevilla, C.; Matthews, L.; Gong, C.; Deng, C.; Varusai, T.; Ragueneau, E.; Haider, Y.; May, B.; Shamovsky, V.; Weiser, J.; Brunson, T.; Sanati, N.; Beckman, L.; Shao, X.; Fabregat, A.; Sidiropoulos, K.; Murillo, J.; Viteri, G.; Cook, J.; Shorsler, S.; Bader, G.; Demir, E.; Sander, C.; Haw, R.; Wu, G.; Stein, L.; Hermjakob, H.; D'Eustachio, P. The Reactome Pathway Knowledgebase 2022. *Nucleic Acids Res* **2022**, *50* (D1), D687–D692.
- (821) Laukens, K.; Naulaerts, S.; Berghe, W. V. Bioinformatics Approaches for the Functional Interpretation of Protein Lists: From Ontology Term Enrichment to Network Analysis. *Proteomics* **2015**, *15* (5-6), 981–996.
- (822) The Gene Ontology Consortium. The Gene Ontology Resource: Enriching a GOLD Mine. *Nucleic Acids Res* **2021**, *49* (D1), D325–D334.
- (823) Wijesooriya, K.; Jadaan, S. A.; Perera, K. L.; Kaur, T.; Ziemann, M. Urgent Need for Consistent Standards in Functional Enrichment Analysis. *PLoS Comput Biol* **2022**, *18* (3), e1009935.
- (824) Maleki, F.; Ovens, K.; Hogan, D. J.; Kusalik, A. J. Gene Set Analysis: Challenges, Opportunities, and Future Research. *Front Genet* **2020**, *11*, 654.
- (825) Fisher, R. A. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society* **1922**, *85* (1), 87.
- (826) Wijesooriya, K.; Jadaan, S. A.; Perera, K. L.; Kaur, T.; Ziemann, M. Urgent Need for Consistent Standards in Functional Enrichment Analysis. *PLoS Comput Biol* **2022**, *18* (3), e1009935.
- (827) Chen, E. Y.; Tan, C. M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G. V.; Clark, N. R.; Ma'ayan, A. Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool. *BMC Bioinformatics* **2013**, *14*, 128.
- (828) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. Gene Set Enrichment Analysis: A

Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (43), 15545–15550.

(829) Sherman, B. T.; Hao, M.; Qiu, J.; Jiao, X.; Baseler, M. W.; Lane, H. C.; Imamichi, T.; Chang, W. DAVID: A Web Server for Functional Enrichment Analysis and Functional Annotation of Gene Lists (2021 Update). *Nucleic Acids Res* **2022**, *50* (W1), W216–W221.

(830) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc Natl Acad Sci U S A* **2005**, *102* (43), 15545–15550.

(831) Mitrea, C.; Taghavi, Z.; Bokanizad, B.; Hanoudi, S.; Tagett, R.; Donato, M.; Voichița, C.; Drăghici, S. Methods and Approaches in the Topology-Based Analysis of Biological Pathways. *Front Physiol* **2013**, *4*, 278.

(832) Masuda, N. Faculty Opinions Recommendation of Gene Co-Expression Analysis for Functional Classification and Gene-Disease Predictions. *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature* **2021**, DOI: [10.3410/f727201931.793589848](https://doi.org/10.3410/f727201931.793589848).

(833) Pardo-Díaz, J.; Bozhilova, L. V.; Beguerisse-Díaz, M.; Poole, P. S.; Deane, C. M.; Reinert, G. Robust Gene Coexpression Networks Using Signed Distance Correlation. *Bioinformatics* **2021**, 1982.

(834) Fortunato, S. Community Detection in Graphs. *Physics Reports* **2010**, *486* (3–5), 75–174.

(835) Gunning, P. W.; Hardeman, E. C. Fundamental Differences. *eLife* **2018**, *7*, No. e34477.

(836) Bludau, I.; Frank, M.; Dörig, C.; Cai, Y.; Heusel, M.; Rosenberger, G.; Picotti, P.; Collins, B. C.; Röst, H.; Aebersold, R. Systematic Detection of Functional Proteoform Groups from Bottom-up Proteomic Datasets. *Nat. Commun.* **2021**, 3810.

(837) Dermitt, M.; Peters-Clarke, T. M.; Shishkova, E.; Meyer, J. G. Peptide Correlation Analysis (PeCorA) Reveals Differential Proteoform Regulation. *J Proteome Res* **2020**, *20* (4), 1972–1980.

(838) Sun, P. D.; Foster, C. E.; Boyington, J. C. Overview of Protein Structural and Functional Folds. *Curr Protoc Protein Sci* **2004**, No. 1, Unit 17.1.

(839) Xu, J. Distance-Based Protein Folding Powered by Deep Learning. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (34), 16856–16865.

(840) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, 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.

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

(842) Oates, M. E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M. J.; Xue, B.; Dosztányi, Z.; Uversky, V. N.; Obradovic, Z.; Kurgan, L.; Dunker, A. K.; Gough, J. D²P²: Database of Disordered Protein Predictions. *Nucleic Acids Res* **2012**, *41* (Database issue), D508–16.

(843) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr Comput Aided Drug Des* **2011**, *7* (2), 146–157.

(844) Gelpi, J.; Hospital, A.; Goñi, R.; Orozco, M. Molecular Dynamics Simulations: Advances and Applications. *AABC* **2015**, 37.

(845) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J Mol Biol* **1990**, *215* (3), 403–410.

(846) Unsal, S.; Atas, H.; Albayrak, M.; Turhan, K.; Acar, A. C.; Doğan, T. Learning Functional Properties of Proteins with Language Models. *Nat Mach Intell* **2022**, *4* (3), 227–245.

(847) Bepler, T.; Berger, B. Learning the Protein Language: Evolution, Structure, and Function. *Cell Systems* **2021**, *12* (6), 654–669.e3.

(848) Mateus, A.; Kurzawa, N.; Becher, I.; Sridharan, S.; Helm, D.; Stein, F.; Typas, A.; Savitski, M. M. Thermal Proteome Profiling for Interrogating Protein Interactions. *Mol Syst Biol* **2020**, *16* (3), No. e9232.

(849) Siebenmorgen, T.; Zacharias, M. Computational Prediction of Protein–Protein Binding Affinities. *WIREs Comput Mol Sci* **2019**, *10* (3), e1448.

(850) Luo, H.; Lin, Y.; Gao, F.; Zhang, C.-T.; Zhang, R. DEG 10, an Update of the Database of Essential Genes That Includes Both Protein-Coding Genes and Noncoding Genomic Elements: Table 1. *Nucl. Acids Res.* **2013**, *42* (D1), D574–D580.

(851) Funk, L.; Su, K.-C.; Ly, J.; Feldman, D.; Singh, A.; Moodie, B.; Blainey, P. C.; Cheeseman, I. M. The Phenotypic Landscape of Essential Human Genes. *Cell* **2022**, *185* (24), 4634–4653.e22.

(852) Stepanenko, A. A.; Heng, H. H. Transient and Stable Vector Transfection: Pitfalls, Off-Target Effects, Artifacts. *Reviews in Mutation Research* **2017**, *773*, 91–103.

(853) Boettcher, M.; McManus, Michael, T. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular Cell* **2015**, *58* (4), 575–585.

(854) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337* (6096), 816–821.

(855) Barman, A.; Deb, B.; Chakraborty, S. A Glance at Genome Editing with CRISPR–Cas9 Technology. *Curr Genet* **2019**, *66* (3), 447–462.

(856) Rezazade Bazaz, M.; Dehghani, H. From DNA Break Repair Pathways to CRISPR/Cas-Mediated Gene Knock-in Methods. *Life Sciences* **2022**, *295*, No. 120409.

(857) Li, G.; Li, X.; Zhuang, S.; Wang, L.; Zhu, Y.; Chen, Y.; Sun, W.; Wu, Z.; Zhou, Z.; Chen, J.; Huang, X.; Wang, J.; Li, D.; Li, W.; Wang, H.; Wei, W. Gene Editing and Its Applications in Biomedicine. *Sci. China Life Sci.* **2022**, *65* (4), 660–700.

(858) Saber Sichani, A.; Ranjbar, M.; Baneshi, M.; Torabi Zadeh, F.; Fallahi, J. A Review on Advanced CRISPR-Based Genome-Editing Tools: Base Editing and Prime Editing. *Mol Biotechnol* **2022**, *65* (6), 849–860.

(859) Boti, M. A.; Athanasopoulou, K.; Adamopoulos, P. G.; Sideris, D. C.; Scorilas, A. Recent Advances in Genome-Engineering Strategies. *Genes* **2023**, *14* (1), 129.

(860) Naeem, M.; Majeed, S.; Hoque, M. Z.; Ahmad, I. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR–Cas-Mediated Genome Editing. *Cells* **2020**, *9* (7), 1608.

(861) Prozzillo, Y.; Fattorini, G.; Santopietro, M. V.; Suglia, L.; Ruggiero, A.; Ferreri, D.; Messina, G. Targeted Protein Degradation Tools: Overview and Future Perspectives. *Biology* **2020**, *9* (12), 421.

(862) T. Das, A.; Tenenbaum, L.; Berkhout, B. Tet-On Systems For Doxycycline-Inducible Gene Expression. *CGT* **2016**, *16* (3), 156–167.

(863) Kallunki; Barisic; Jäättelä; Liu. How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells* **2019**, *8* (8), 796.

(864) Spitzer, J.; Landthaler, M.; Tuschl, T. Rapid Creation of Stable Mammalian Cell Lines for Regulated Expression of Proteins Using the Gateway® Recombination Cloning Technology and FLP-In T-Rex® Lines. In *Methods in Enzymology*; Elsevier, 2013; pp 99–124. DOI: [10.1016/b978-0-12-418687-3.00008-2](https://doi.org/10.1016/b978-0-12-418687-3.00008-2).

(865) Ward, R. J.; Alvarez-Curto, E.; Milligan, G. Using the FLP-In T-Rex System to Regulate GPCR Expression. In *Methods in Molecular*

Biology; Humana Press, 2011; pp 21–37. DOI: 10.1007/978-1-61779-126-0_2.

(866) Handler, D. C.; Pascovici, D.; Mirzaei, M.; Gupta, V.; Salekdeh, G. H.; Haynes, P. A. The Art of Validating Quantitative Proteomics Data. *Proteomics* **2018**, 18 (23), e1800222.

(867) Fritzschke, S.; Springer, S. Pulse-Chase Analysis for Studying Protein Synthesis and Maturation. *CP Protein Science* **2014**, 78 (1), 30.3.1.

(868) Sin, C.; Chiarugi, D.; Valleriani, A. Degradation Parameters from Pulse-Chase Experiments. *PLoS ONE* **2016**, 11 (5), e0155028.

(869) Vallejo, D. D.; Rojas Ramírez, C.; Parson, K. F.; Han, Y.; Gadkari, V. V.; Ruotolo, B. T. Mass Spectrometry Methods for Measuring Protein Stability. *Chem. Rev.* **2022**, 122 (8), 7690–7719.

(870) Gadkar, V. J.; Filion, M. New Developments in Quantitative Real-Time Polymerase Chain Reaction Technology. *Current Issues in Molecular Biology* **2014**, 1.

(871) Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nat Rev Genet* **2009**, 10 (1), 57–63.

(872) Pedelacq, J.-D.; Cabantous, S. Development and Applications of Superfolder and Split Fluorescent Protein Detection Systems in Biology. *IJMS* **2019**, 20 (14), 3479.

(873) Lehming, N. Analysis of Protein-Protein Proximities Using the Split-Ubiquitin System. *Briefings in Functional Genomics and Proteomics* **2002**, 1 (3), 230–238.

(874) Azad, T.; Tashakor, A.; Hosseinkhani, S. Split-Luciferase Complementary Assay: Applications, Recent Developments, and Future Perspectives. *Anal Bioanal Chem* **2014**, 406 (23), 5541–5560.

(875) Cesaratto, F.; Burrone, O. R.; Petris, G. Tobacco Etch Virus Protease: A Shortcut Across Biotechnologies. *Journal of Biotechnology* **2016**, 231, 239–249.

(876) Hinton, S. R.; Corpuz, E. L. S.; McFarlane Holman, K. L.; Meyer, S. C. A Split β -Lactamase Sensor for the Detection of DNA Modification by Cisplatin and Ruthenium-Based Chemotherapeutic Drugs. *Journal of Inorganic Biochemistry* **2022**, 236, No. 111986.

(877) Wang, T.; Yang, N.; Liang, C.; Xu, H.; An, Y.; Xiao, S.; Zheng, M.; Liu, L.; Wang, G.; Nie, L. Detecting Protein-Protein Interaction Based on Protein Fragment Complementation Assay. *CPPS* **2020**, 21 (6), 598–610.

(878) Pichlerova, K.; Hanes, J. Technologies for the Identification and Validation of Protein-Protein Interactions. *gpb* **2021**, 40 (06), 495–522.

(879) Fields, S.; Song, O. A Novel Genetic System to Detect Protein-Protein Interactions. *Nature* **1989**, 340 (6230), 245–246.

(880) Brückner, A.; Polge, C.; Lentze, N.; Auerbach, D.; Schlattner, U. Yeast Two-Hybrid, a Powerful Tool for Systems Biology. *IJMS* **2009**, 10 (6), 2763–2788.

(881) Patrício, D.; Fardilha, M. The Mammalian Two-Hybrid System as a Powerful Tool for High-Throughput Drug Screening. *Drug Discovery Today* **2020**, 25 (4), 764–771.

(882) Dong, Y.; Yang, J.; Ye, W.; Wang, Y.; Ye, C.; Weng, D.; Gao, H.; Zhang, F.; Xu, Z.; Lei, Y. Isolation of Endogenously Assembled RNA-Protein Complexes Using Affinity Purification Based on Streptavidin Aptamer S1. *IJMS* **2015**, 16 (9), 22456–22472.

(883) Algar, W. R.; Hildebrandt, N.; Vogel, S. S.; Medintz, I. L. FRET as a Biomolecular Research Tool—Understanding Its Potential While Avoiding Pitfalls. *Nat Methods* **2019**, 16 (9), 815–829.

(884) Shekhawat, S. S.; Ghosh, I. Split-Protein Systems: Beyond Binary Protein-Protein Interactions. *Current Opinion in Chemical Biology* **2011**, 15 (6), 789–797.

(885) Wang, P.; Yang, Y.; Hong, T.; Zhu, G. Proximity Ligation Assay: An Ultrasensitive Method for Protein Quantification and Its Applications in Pathogen Detection. *Appl Microbiol Biotechnol* **2021**, 105 (3), 923–935.

(886) Shrestha, D.; Jenei, A.; Nagy, P.; Vereb, G.; Szöllösi, J. Understanding FRET as a Research Tool for Cellular Studies. *IJMS* **2015**, 16 (12), 6718–6756.

(887) Pflieger, K. D. G.; Eidne, K. A. Illuminating Insights into Protein-Protein Interactions Using Bioluminescence Resonance Energy Transfer (BRET). *Nat Methods* **2006**, 3 (3), 165–174.

(888) Wu, Y.; Jiang, T. Developments in FRET- and BRET-Based Biosensors. *Micromachines* **2022**, 13 (10), 1789.

(889) Martinez-Moro, M.; Di Silvio, D.; Moya, S. E. Fluorescence Correlation Spectroscopy as a Tool for the Study of the Intracellular Dynamics and Biological Fate of Protein Corona. *Biophysical Chemistry* **2019**, 253, No. 106218.

(890) Zafra, F.; Piniella, D. Proximity Labeling Methods for Proteomic Analysis of Membrane Proteins. *Journal of Proteomics* **2022**, 264, No. 104620.

(891) Che, Y.; Khavari, P. A. Research Techniques Made Simple: Emerging Methods to Elucidate Protein Interactions Through Spatial Proximity. *Journal of Investigative Dermatology* **2017**, 137 (12), e197–e203.

(892) Koos, B.; Andersson, L.; Clausson, C.-M.; Grannas, K.; Klaesson, A.; Cane, G.; Söderberg, O. Analysis of Protein Interactions in Situ by Proximity Ligation Assays. *Current Topics in Microbiology and Immunology* **2013**, 111–126.

(893) Sarnowski, C. P.; Bikaki, M.; Leitner, A. Cross-Linking and Mass Spectrometry as a Tool for Studying the Structural Biology of Ribonucleoproteins. *Structure* **2022**, 30 (4), 441–461.

(894) Steinmetz, B.; Smok, I.; Bikaki, M.; Leitner, A. Protein-RNA Interactions: From Mass Spectrometry to Drug Discovery. *Essays in Biochemistry* **2023**, 67 (2), 175–186.

(895) Belsom, A.; Rappsilber, J. Anatomy of a Crosslinker. *Current Opinion in Chemical Biology* **2021**, 60, 39–46.

(896) Drescher, D. G.; Selvakumar, D.; Drescher, M. J. Analysis of Protein Interactions by Surface Plasmon Resonance. In *Protein-Protein Interactions in Human Disease, Part A*; Elsevier, 2018; pp 1–30.

(897) Himmelstein, D. S.; Rubineti, V.; Slochower, D. R.; Hu, D.; Malladi, V. S.; Greene, C. S.; Gitter, A. Open Collaborative Writing with Manubot. *PLoS Comput Biol* **2019**, 15 (6), e1007128.