REVIEW

Resolving protein interactions and complexes by affinity purification followed by label-based quantitative mass spectrometry

Laura Trinkle-Mulcahy[1,2]

[1] Department of Cellular & Molecular Medicine, University of Ottawa, Ottawa, ON, Canada
[2] Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada

Label-based quantitative mass spectrometry analysis of affinity purified complexes, with its built-in negative controls and relative ease of use, is an increasingly popular choice for defining protein–protein interactions and multiprotein complexes. This approach, which differentially labels proteins/peptides from two or more populations and combines them prior to analysis, permits direct comparison of a protein pulldown (e.g. affinity purified tagged protein) to that of a control pulldown (e.g. affinity purified tag alone) in a single mass spectrometry (MS) run, thus avoiding the variability inherent in separate runs. The use of quantitative techniques has been driven in large part by significant improvements in the resolution and sensitivity of high-end mass spectrometers. Importantly, the availability of commercial reagents and open source identification/quantification software has made these powerful techniques accessible to nonspecialists. Benefits and drawbacks of the most popular labeling-based approaches are discussed here, and key steps/strategies for the use of labeling in quantitative immunoprecipitation experiments detailed.

Received: August 19, 2011
Revised: November 15, 2011
Accepted: December 19, 2011

Keywords:
Affinity purification / Animal proteomics / Interactome / Isotopes / Quantitative Proteomics / SILAC

1 Introduction

Most biological processes involve the action and regulation of multiprotein complexes, and thus a key goal in cell biology is the characterization of these complexes through the reliable identification of protein interaction partners. This can be done in a low-throughput fashion to characterize the functional interactome of a single protein of interest [1–3] or in a high-throughput fashion to map global interactomes [4–9]. The current method of choice for interactome analysis is affinity purification followed by mass spectrometry (AP-MS; see [10] and [11] for review). The major strength of this technique is that it can resolve entire multiprotein complexes in a single experiment, identifying both stable and dynamic interactors [12, 13]. Results can then be integrated with data from yeast two-hybrid screens (which resolve binary interactions) and RNAi screens (which resolve the functional consequences of protein knockdown).

While continuing improvement in the sensitivity and resolution of the mass spectrometric technology for protein identification permits the identification of ever larger numbers of proteins in immunoaffinity and pull-down experiments [14, 15], these expanding lists include, in addition to bona fide interaction partners, increased numbers of contaminant proteins, including environmental proteins (e.g. keratins and serum proteins) and proteins that bind nonspecifically to the affinity matrix. The latter have been estimated to account for up to 95% of identified proteins in an AP-MS experiment [13, 16]. The cost and time required to validate putative interaction partners to confirm their physiological relevance highlights

Correspondence: Laura Trinkle-Mulcahy, Department of Cellular and Molecular Medicine, Faculty of Medicine, University of Ottawa, 451 Smyth Road, Ottawa, ON K1H 8M5, Canada
E-mail: ltrinkle@uottawa.ca
Fax: +1-613-562-5636

Abbreviations: AP, affinity purification; **GFP**, Green Fluorescent Protein; **IP**, Immunoprecipitation; **QUICK**, Quantitative Immunoprecipitation Combined with Knockdown; **TAP**, tandem affinity purification

Table 1. Epitope tags commonly used for AP-MS experiments. TEV and PreScission protease sites are included in LAP and TAP tags where indicated.

Affinity tag	Description	Approximate size	Purification strategy	Related references
FLAG	Peptide (DYKDDDDK)	~1 kDa	FLAG antibody (e.g. M2) conjugated to beads	[87]
GFP	Aequoria victoria protein	~ 27 kDa	GFP antibody conjugated to beads (e.g. GFP-Trap®)	[13]
Glutathione-S-transferase (GST)	Enzyme	~26 kDa	Glutathione-coated beads	[88]
Haemagglutinin (HA)	Peptide (YPYDVPDYA)	~1 kDa	HA antibody conjugated to beads	[89]
Hexahistidine 6xHis	(HHHHHH)	<1 kDa	Ni sepharose	[90]
Localization and affinity (LAP)	GFP-TEV-S peptide (KETAAAKFERQH-MDS) or 6xHis-PreScission-GFP	~28-30 kDa	GFP antibody conjugated to beads and S protein agarose (or Ni-NTA agarose)	[25]
Strep•Tag® II	Peptide (WSHPQFEK)	~1 kDa	Strep•Tactin® (engineered streptavidin)-coated beads	[91]
Tandem affinity purification (TAP)	Protein A-TEV-Calmodulin binding peptide (CBP)	~21 kDa	IgG-conjugated to beads and calmodulin conjugated to beads	[22]

the importance of focusing resources on those subsets of potential interactions with a high probability of biological significance. Higher stringency purification methods, such as increasing the salt/detergent concentrations in buffers or multistep purification/elution protocols, can help to overcome the problem of nonspecific binding but can also lead to a loss of low affinity and low abundance specific partner proteins. The most effective strategy is thus one that preserves all specific protein–protein interaction events while clearly distinguishing them from background nonspecific binding events.

We and others have demonstrated that a quantitative MS approach based on the use of isotopic labeling of proteins or peptides can help to distinguish which of the many proteins identified in a pull-down or immunoprecipitation (IP) experiment represent specific binding. This is done by the inclusion of a negative control, which provides a background of contaminant proteins that bind nonspecifically to the affinity matrix and/or the fusion tag, against which proteins that bind specifically to the protein of interest clearly stand out (for review see [12, 17–19]). Proteins can thus be purified under lower stringency conditions, which preserves more specific interactions.

1.1 Epitope tagging

Although endogenous proteins in their native settings under physiological conditions would appear to be the ideal "bait" for mapping interactomes via quantitative AP-MS experiments, and indeed have been utilized successfully in numerous screens, including a recent large-scale study based on 3290 affinity purifications of endogenous proteins from human cell nuclear extracts [4], interpretation of data sets can be complicated by cross-reactivity of the antibodies. Computational methods for dealing with this include calculating the reciprocity of interactions, i.e. proteins in the same complexes should have overlapping interactomes [2,4]. A methodological way around this is to compare an IP of an endogenous protein under normal conditions to the same IP under conditions in which levels of the target protein have been significantly reduced by RNAi. This technique has been termed QUICK (Quantitative Immunoprecipitation Combined with Knockdown), and can be used to highlight interaction partners that are specific to the protein of interest [16].

An alternate approach to endogenous pulldowns is the exogenous expression of epitope-tagged recombinant proteins that can be efficiently recovered from cell extracts using affinity matrices or well-characterized antibodies (for Review see [20, 21]). This is currently the most popular method for AP-MS studies, and Table 1 lists several commonly used epitope tags, along with their respective sizes and purification strategies. Importantly, because the same high affinity matrix or antibody is used to isolate all bait complexes, nonspecific contaminant proteins are more readily distinguishable by their appearance in multiple data sets. It should be noted that although the tandem affinity purification (TAP) tag was originally designed for high stringency purification [22], it can

be also be used for single step purifications with lower stringencies, to preserve more interactions and facilitate large-scale comparative studies [23].

Similarly, although the green fluorescent protein (GFP) originally revolutionized cell biology as a tagging method for in vivo live imaging, minimal nonspecific binding and the availability of highly efficient affinity reagents has led to its adoption as an affinity tag, permitting the direct comparison of imaging and proteomics data [13, 24–26]. Affinity purification of GFP-tagged bait proteins was recently paired with the equally powerful BAC (bacterial artificial chromosome) TransgenOmics approach [27] in a method dubbed QUBIC (Quantitative BAC-green fluorescent protein interactomics). This technique, which can be carried out on single baits or automated for high-throughput analyses of multiple baits, avoids complications associated with protein overexpression by affinity purifying GFP-tagged proteins expressed in human cells under the control of endogenous regulatory elements [5, 28].

In addition to mapping protein–protein interactions, quantitative AP-MS has also been extended to the analysis of protein binding partners for synthetic peptides [29] synthetic DNA oligonucleotides [30] and tagged RNA [31, 32].

1.2 Labeling strategies

A large number of differential labeling techniques have been developed over the years and applied to both whole proteome analyses and quantitative AP-MS studies. These range from metabolic approaches that label proteins in vivo through incorporation of isotopic amino acids, to chemical labeling of proteins following release from cells, to labeling of peptides during or after proteolytic digestion (for review see [33, 34]). Although expense can be an issue with any type of labeling experiment, particularly if costly commercial kits are used, the quantitative nature and built-in negative control mean that fewer replicates need to be run, thus reducing MS analysis costs. A single quantitative AP-MS experiment in which control and experimental samples are combined can be carried out based on a relatively small number of MS runs [5, 13], which is a significant savings in service costs compared to standard nonquantitative approaches in which dozens of bands (and their corresponding locations in control lanes) are cut out of gels and digested for subsequent analysis. In addition, there is now a wealth of free, open source software available for the analysis of the large, high-resolution MS data sets generated in quantitative proteomics experiments (for review see [35, 36]), including two comprehensive suites that provide single environments for performing all or most steps in the workflow: the TransProteomics Pipeline (TPP; http://tools.proteomecenter.org/) developed at the Seattle Proteome Center [37] and MaxQuant (http:// maxquant.org), developed at the Max Planck Institute in Martinsried, Germany [38, 39]. Both provide extensive details about the software, guides for first-time users, lists of frequently asked questions, sample data sets to practice analysis and user groups for ongoing support from the developers and from other users. Taken together, this extends quantitative AP-MS for interactome analysis to the nonspecialist laboratory.

All other steps similarly require only standard laboratory equipment or readily available reagents and only knowledge of common biochemical procedures. Costs per pull-down are very low. QUBIC does require access to high resolution MS equipment coupled to high performance LC. However, such equipment is increasingly accessible, and the MS analyses themselves are relatively standard.

Three of the most commonly used labeling techniques for AP-MS are ICAT (Isotope-Coded Affinity Tags), iTRAQ (Isobaric Tag for Relative and Absolute Quantification) and SILAC (Stable Istope Labeling with Amino Acids in cell Culture). ICAT is an isotopic labeling technique in which cysteines in proteins are reacted with chemical labels carrying different isotope-coded linker regions (e.g. light ^{12}C ICAT 0 and heavy ^{13}C ICAT 9) and a biotin tag to purify labeled peptides (Fig. 1A; for review see [40]). In an AP-MS experiment, separate control and experimental IPs are carried out, proteins eluted and labeled with the appropriate ICAT tag and then mixed for proteolytic digestion. A further affinity purification step is used to selectively capture cystein-containing peptides, and the biotin tag can then be cleaved off if desired. During LC-MS/MS, differentially labeled peptides will elute at the same time from the HPLC, and the isotopic linker regions confer mass differences upon them (i.e. a shift of 9 Da for peptides containing ICAT 9-labeled cysteine) that allow comparison of relative levels of heavy and light peptides.

Strengths of the ICAT approach are its applicability to samples from any source, the upstream labeling of whole proteins prior to digestion, which helps to reduce the technical variability introduced by sample handling, and the selective enrichment and analysis of labeled peptides, which reduces sample complexity. A weakness is that labeling is limited to cysteine residues, which will reduce sequence coverage for all proteins and preclude identification of proteins that contain no cysteine residues (estimated to be 8% of the yeast proteome and <5% of the human proteome [33]).

Examples of the use of ICAT for quantitative AP-MS include the mapping of components of the RNA Polymerase II preinitiation complex [41], identification of proteins that interact with MafK transcription factors during erythroid differentiation [42] and identification of transcription factors that bind a novel MCK promoter element important for promoter activity in cardiac and skeletal muscle [43].

iTRAQ is another chemical labeling technique, albeit one that tags N-termini and lysine residues in peptides following proteolytic digestion (Fig. 1B; for review see [44]). The tags are isobaric, and thus differentially labeled peptides cannot be distinguished in the MS spectrum. Quantification is instead based on reporter ions that are released upon fractionation of the tag during MS/MS.

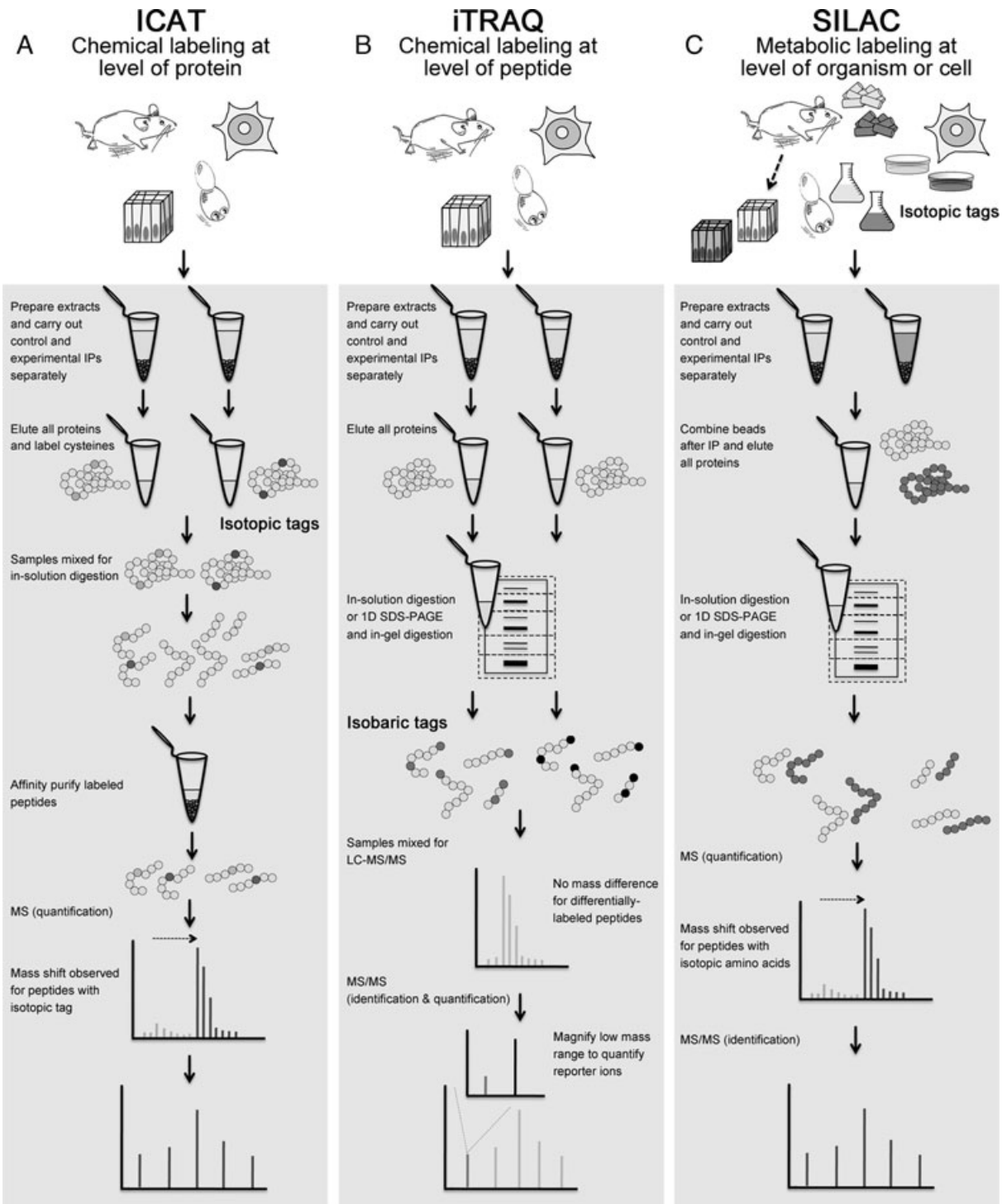


Figure 1. Standard workflows for three of the most common labeling techniques used in AP-MS experiments. (A) For ICAT, affinity purifications of bait and control are carried out separately from cell extracts and the proteins differentially labeled by the chemical addition of isotopic tags to cysteine residues. The samples are then mixed and a second affinity purification step carried out to enrich cysteine-containing peptides for LC-MS/MS analysis. (B) For iTRAQ, affinity purifications of bait and control are carried out separately from cell extracts, proteins eluted and proteolyzed and the resulting peptides differentially labeled by the chemical addition of isobaric tags to N-termini and Lysine residues. The samples are then mixed for LC-MS/MS analysis. (C) For SILAC, whole organisms or cultured cells are labeled metabolically by feeding them isotopic variants of essential amino acids over time such that every protein is labeled. Affinity purifications of bait and control (from differentially labeled cell extracts) are carried out separately and combined at the elution step, upstream of digestion and LC-MS/MS analysis.

Strengths of the iTRAQ approach are its applicability to samples from any source, the ability to label most peptides in the sample and the multiplex options that allow comparison of up to eight samples in a single experiment without any increase in sample complexity due to the isobaric nature of the tags. Weaknesses include the fact that labeling is carried out further downstream in the workflow, which introduces more experimental variability, the fact that iTRAQ ratios can not be determined over the complete LC peak in the MS spectrum but only for single peptides selected for MS/MS, and the requirement of instrumentation capable of measuring fragments in the low m/z range. In addition, because iTRAQ relies on the use of common reporter ions, the accuracy of quantification can be affected by co-eluting labeled peptides, particularly in complex samples.

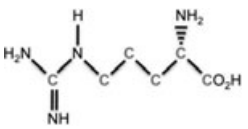
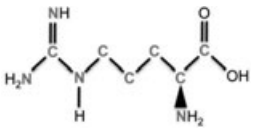
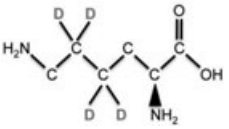
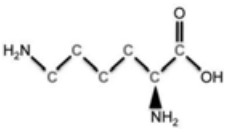
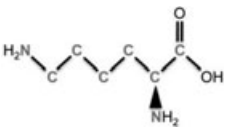
Examples of the use of iTRAQ for quantitative AP-MS include identification of protein–protein interactions and phosphorylation sites in a single experiment [45], identification of matrix metalloproteinase (MMP) substrates in cell culture [46], mapping of in vivo brain interactomes of amyloid precursor proteins [47] and assessment of the binding of protein kinase inhibitors to their targets [48].

SILAC is a metabolic labeling technique in which isotopic versions of essential amino acids are incorporated over time via normal protein turnover. With complete incorporation, all proteins are labeled, and by choosing the appropriate combination of amino acid and proteolytic enzyme, a significant number of peptides can be quantified. Trypsin is a commonly used protease that cuts after Arg and Lys residues, and thus most SILAC experiments are based on the use of isotopic Arg and/or Lys (see Table 2). The use of both Arg and Lys, in combination with trypsin proteolysis, ensures that nearly all peptides will be quantifiable as they will by definition terminate in an Arg or Lys residue. It is not always possible to use both, however, and in that event the preferred single isotope is normally Lys, as it can be used in combination with the proteolytic enzyme LysC, which only cuts after Lys residues. Isotopic Arg can be used on its own, even in combination with trypsin, however the number of quantifiable (i.e. Arg-containing) peptides will be significantly reduced.

In addition to the standard 2-plex (Arg0Lys0 versus Arg6Lys4) and 3-plex (Arg0Lys0 versus Arg6Lys4 versus Arg10Lys8) experiments enabled by the isotopes listed in Table 2, the use of Arg17 (L-Arginine HCl $^{13}C_6$,$^{15}N_4$,D7) and Arg4 (L-Arginine HCl $^{15}N_4$) has also extended SILAC to 4-plex (Arg6Lys4, Arg10Lys8, and Arg17Lys17) and 5-plex (Arg4, Arg6, Arg10, Arg17) comparisons [49]. Arg6Lys6 has also been used for 2-plex, however this complicates analysis as the mass shift is identical for both amino acids.

Strengths of the SILAC approach include its ease of use, which make it an attractive option for nonspecialists, and in vivo incorporation of the label, which permits earlier mixing of samples and thus minimizes variability introduced by handling errors. This method is extremely accurate because all peptides are labeled and processing of proteins occurs after samples have been combined. Like ICAT (and in contrast to iTRAQ), quantification occurs at the MS1 level and ratios can be determined over the complete LC peak. A major weakness is the inability to extend it to AP-MS for samples that cannot be labeled metabolically, such as human tissues, although techniques based on the use of culture-derived isotope tags as internal standards have extended the use of SILAC to whole proteome comparisons in these systems [50–52]). There are also issues with conversion of heavy Arg to heavy Pro in vivo, which can complicate data analysis. These have been addressed experimentally by providing limiting levels of Arg [53] or additional unlabeled Proline [54, 55], or by including isotopic Arg in the light condition to permit formation of heavy Pro at the same rate in both conditions as an internal correction [56]. Computational methods to distinguish and correct for heavy Pro-containing peptides have also been

Table 2. Isotopic amino acids commonly used for SILAC labeling experiments

Label	Chemical formula	Structure	Mass shift (Da)
Arg6 (R6)	L-Arginine:HCl ($^{13}C_6$) CIL #CLM-2265 Sigma #643440		6
Arg10 (R10)	L-Arginine:HCl ($^{13}C_6$, $^{15}N_4$) CIL #CNLM-539 Sigma #608033		10 10
Lys4 (K4)	L-Lysine:HCl (4,4,5,5-D4) CIL #DLM2640 Sigma #616192		4
Lys6 (K6)	L-Lysine : 2HCl ($^{13}C_6$) CIL #CLM-2247 Sigma #643459		6
Lys8 (K8)	L-Lysine:2HCl ($^{13}C_6$, $^{15}N_2$) CIL #CNLM-291 Sigma #608041		8

Where indicated, environmental (^{12}C) carbon is substituted with its ^{13}C isotope, environmental (^{14}N) nitrogen with its ^{15}N isotope and environmental (^{1}H) hydrogen with its ^{2}H (deuterium, D) isotope. Mass shifts are calculated accordingly. Catalogue numbers for Cambridge Isotope Laboratories (CIL) and Sigma are noted below each isotope. Arg- and Lys-free mammalian cell culture media and fetal bovine serum dialyzed at 10 kDa to remove amino acids are available from a variety of commercial sources, providing a more cost effective alternative to commercially available SILAC labeling kits.

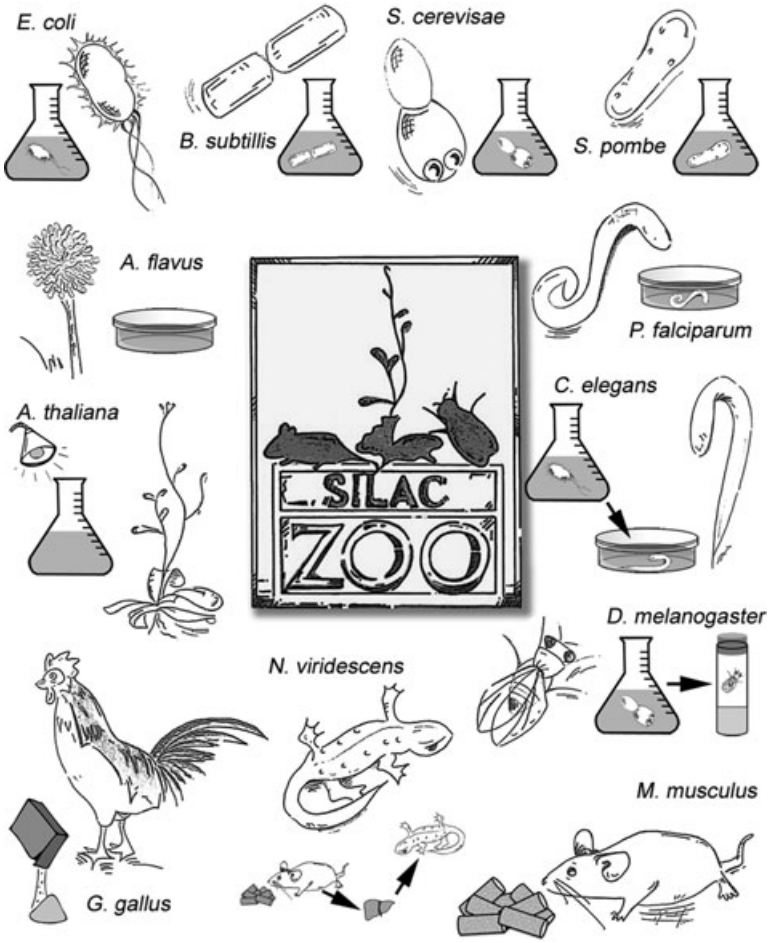


Figure 2. SILAC Zoo. Originally optimized for cultured mammalian cells, SILAC has since been applied to a wide range of organisms. Although not all have been labeled to complete incorporation (e.g. *G. gallus, N. viridescens*), the combination of careful experimental design with partial incorporation permits their use in quantitative proteomics experiments. Where indicated, SILAC encoding of certain organisms is achieved by providing them with an encoded food source. *A. thaliani* has been encoded at the level of cultured cells. Artwork by Michèle Prévost (University of Ottawa).

developed [57–59]. In organisms that permit it, the genes involved in Arg catabolism have even been deleted or knocked down to preclude Arg to Pro conversion [8, 60].

Examples of the use of SILAC for quantitative AP-MS include mapping of the interactomes of specific histone modification patterns [61], comparison of the interactomes of wild type and a cancer-associated mutant of PTEN [62], identification of SUMO target proteins [63], and identification of isoform-specific binding partners for PP1 phosphatase [26]. In a method dubbed QTAX (quantitative analysis of tandem affinity purified in vivo cross-linked protein complexes), SILAC was combined with in vivo formaldehyde cross-linking to decipher the yeast 26S proteasome interactome [64]. When it was noted that dynamic interactors can exchange during affinity purification (resulting in a 1:1 ratio of the heavy:light forms), the SILAC approach was further extended to map both stable and dynamic components of complexes such as the human 26S proteasome, TATA-binding protein (TBP) transcription complexes and the COP9 signalosome [65–67]. This was done by comparing the protein enrichment profiles for experiments in which the samples were mixed before purification (only stable interactors appeared enriched) to those for experiments in which the samples were mixed after purification (both stable and dynamic interactors were clearly enriched).

1.3 SILAC Zoo

Although originally adapted for quantitative analysis of mammalian cells in culture [68], SILAC has since been extended to a wide range of organisms, dubbed the "SILAC Zoo" ([69]; for a comprehensive review of metabolic labeling, see [70]). As shown in Fig. 2, labeling is simple in most cases, involving substitution of isotopic amino acids in culture media or food sources. The only requirement is that the organism is auxotrophic for those particular amino acids, i.e. cannot synthesize them. This can be achieved experimentally by generating auxotrophic strains using genetic manipulation, as has been done in yeast to compensate for the fact that they can synthesize all amino acids. For human cells, Arg and Lys are essential amino acids that must be provided in culture, and thus the use of Arg/Lys-deficient media coupled with add-back of either light or heavy isotopic forms of the amino acids can be used to label them to completion over several cell passages. As previously noted, however, the potential for

conversion of isotopic Arg to isotopic Pro can complicate data analysis.

Direct labeling in culture has been achieved for prokaryotes such as *Escherichia coli* [71] and *Bacillus subtilis* [72], fungi such as the budding yeast *Saccharomyces cerevisae* [73], the fission yeast *Schizosaccharomyces pombe* [8] and the mold *Aspergillus flavus* [74], the protozoan parasite *Plasmodium falciparum* [75] and plant cell cultures derived from *Arabidopsis thaliani* [76, 77]. Direct labeling by feeding has also been achieved for the mouse *Mus musculus* [78], which extends the SILAC approach to the powerful mouse knockout technique, permitting direct comparison of knockout strains to parental strains by differential labeling of all proteins in every tissue and organ in the entire body. It is generally more cost effective to label the parental mouse strain with the isotopic amino acid, so that tissues can be used as internal standards for all knockout strains derived from it. The chicken *Gallus gallus* has also been labeled via an isotopic diet, in this case $^{2}H_{8}$Val [79]. Although labeling was not carried out to completion, it was stable over an extended period and utilized to quantify protein turnover.

Other higher organisms have labeled in a secondary manner, by providing labeled organisms as their food source. This is the case for the nematode *Caenorhabditis elegans*, which was fed a diet of labeled *E. coli* [60] and the fruit fly *Drosophila melanogaster*, which incorporated the isotopes through a diet of labeled *S. cerevisae* [69]. The newt *Notophthalmus viridescens* was partially labeled in a pulsed SILAC approach to study protein turnover in regenerating tails, via a diet of livers extracted from SILAC-encoded mice [80].

As demonstrated by the size and variety of the SILAC zoo, this relatively simple differential labeling approach can now be incorporated, with careful experimental design and consideration of important caveats such as amino acid conversion, into a wide range of quantitative AP-MS workflows.

2 Data analysis

2.1 Thresholding

As noted earlier, a wide range of software options exist for the identification and quantification of differential labeling-based AP-MS data sets [35, 36]. Whichever one is chosen, the end result is a list of ratios that reflect the relative abundance of each identified protein in the experimental condition versus the control condition. Plotting the frequency distribution of these ratios is an important first step in judging the success of a quantitative AP-MS experiment. As shown in the sample data set presented in Fig. 3A, a large number of proteins distribute in a peak over a log ratio of 0 (i.e. a heavy:light ratio of 1:1), which is expected, as most will be binding nonspecifically to the affinity matrix. This distribution also confirms that the two samples were mixed equally. If it is shifted to the left or the right, this indicates that variability between the samples has been introduced, either at the label incorporation stage or at some point during processing (e.g. unequal mixing). A small shift can be compensated for to some extent by normalizing to the mean ratio, and some analysis programs have this function built-in, however the best approach is to attempt to minimize this upstream experimental variability in the first place with careful experimental design and implementation.

The next step is to look at the outlying proteins. For metabolic labeling approaches, a large negative log ratio indicates an environmental contaminant such as keratin, which is only present in the light form. At the other extreme, the bait should in theory have the highest log ratio, as it is the protein that has been enriched. Interaction partners present in stoichiometric amounts should also fall near this ratio, although stoichiometry cannot be accurately predicted from these data. That leaves a collection of putative interaction partners with ratios that fall in between the bait protein and the tail of the contaminant peak. An arbitrary threshold can be chosen visually, as shown in Fig. 3A. The more stringent the threshold, the more confidence there is in the putative interactors with ratios above it. However, this brings with it the increased risk of overlooking genuine interactors that fall below the threshold. Figure 3B demonstrates how the inclusion of protein abundance information for the same data set can help to identify proteins that are both enriched in the IP (H:L >1) and relatively abundant (summed peptide intensities normalized for MW). The bait protein is thus expected to fall in the top right corner of the graph, and putative interactors are identified as lying outwith the cluster of low-medium abundance proteins that bind the affinity matrix nonspecifically (Fig. 3B, hashed circle). Highly abundant contaminants are readily identified by the fact that they are not enriched in the experimental IP. This particular data set also highlights the importance of analyzing common contaminants in the context of the experiment rather than filtering them out. Ribosomal proteins are highly abundant, and constituents of both the large 60S subunit (RPLs) and the small 40S subunit (RPSs) bind nonspecifically to affinity matrices [13]. Both subsets normally cluster with other contaminant proteins at a ratio of 1:1, however in this example the bait protein is a known component of the pre-60S processing complex. The consequence of this is that a large number of RPLs co-purify with the bait and are clearly enriched above background, while most RPSs remain in the background contaminant cluster (Fig. 3B).

Choosing a threshold in an intensity-dependent manner also highlights another important point, namely that the accuracy of quantification increases with the number of quantifiable peptides detected, and hence more abundant proteins can be quantified more accurately than less abundant proteins. That leads to higher confidence in the identification of high abundance interactors even if their protein ratios are close to or below threshold. For lower abundance proteins, particularly those with high variance in their protein ratios, manual inspection of chromatograms is recommended as an additional analysis step. The accuracy of quantification can be affected by many factors, including co-eluting off-target

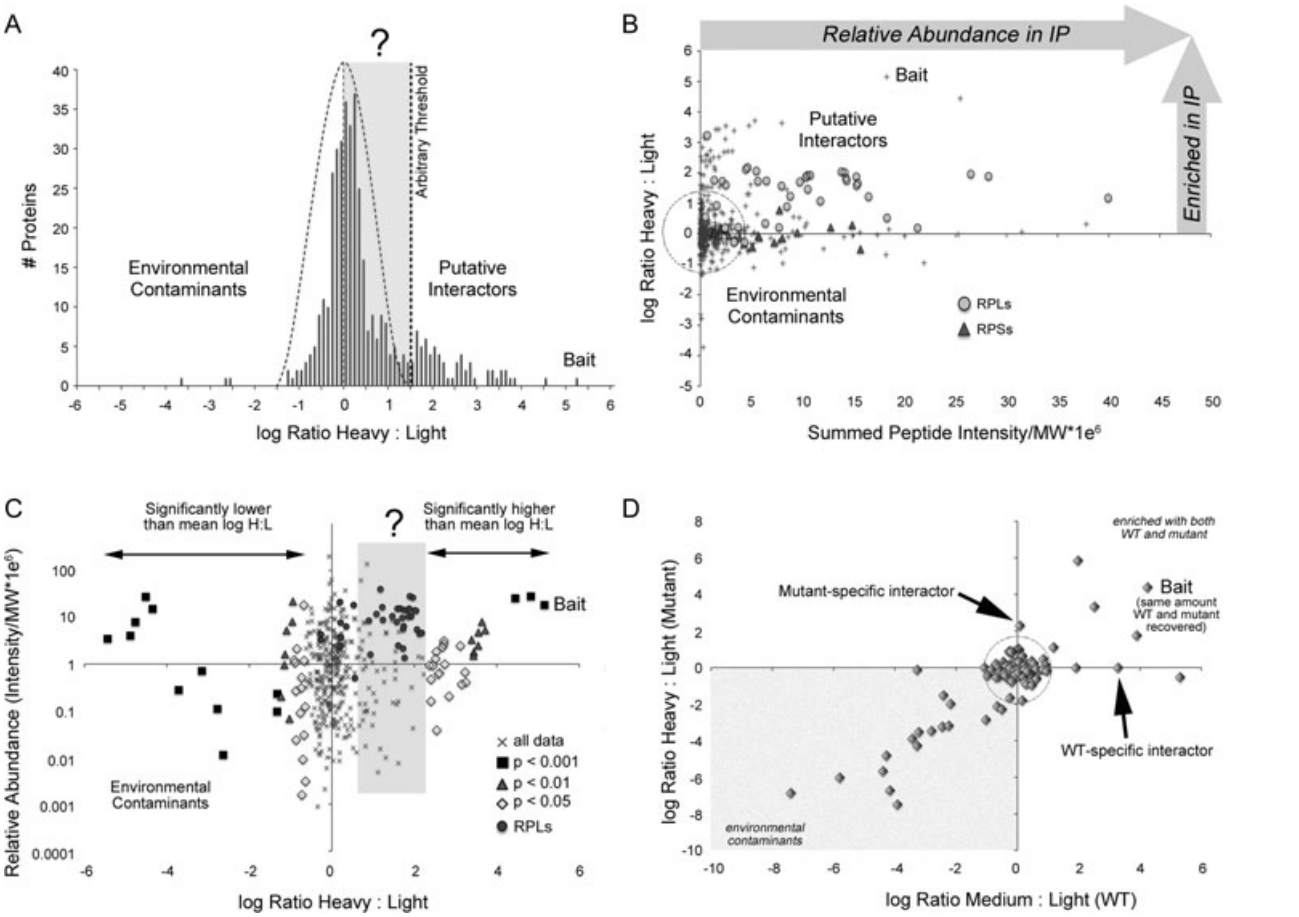


Figure 3. Analysis of labeling-based quantitative AP-MS data. (A) Distribution of experimental : control ratios in a typical differential labeling AP-MS experiment. When control and experimental samples are encoded with light and heavy labels, respectively, as shown here environmental contaminants will display heavy:light ratios <1 (i.e. log heavy:light ratios <0), as they occur only in the light form. The bait protein should be highly enriched, with one of the highest log heavy:light ratios. Putative interactors will also be detected at log heavy:light ratios >0, however it is often difficult to define a threshold above which a protein is a likely real hit. Nonspecific contaminants that are enriched equally in both the control and experimental samples show a Gaussian distribution over a heavy:light ratio of 1:1 (log ratio 0). An arbitrary threshold can be set above this, however a significant number of potential interactors are left buried just below threshold (gray box). (B) Further information is provided when the same dataset is plotted as log heavy:light ratio versus summed peptide intensity (normalized for MW). The bait protein should be both enriched and abundant in the IP, as should putative interactors. Environmental contaminants fall below the line (negative log ratios), while the majority of proteins cluster around a log ratio 0 (hashed circle). Although ribosomal proteins are common contaminants in AP-MS, in this case the 40S subunit proteins (RPSs) remain with the contaminants while the 60S subunit proteins (RPLs) are enriched, demonstrating the specificity of this pulldown of a core member of pre-60S subunit processing complexes. (C) The bait protein and putative interactors can also be identified by their significant variance from mean log heavy:light ratios, as calculated using MaxQuant software in this case, for a chosen confidence threshold value. This leaves a significant number of putative interactors that fall below a confidence value $p < 0.05$, including the RPL proteins (known interactors for this particular bait protein) described in Panel (B). (D) Encoding cells/organisms with three differential labels (light, medium and heavy) permits the direct comparison of a wild-type (WT) bait protein IP (medium) and mutant bait protein IP (heavy) in a single experiment. By plotting log ratio heavy:light (enriched with mutant over control) versus log ratio medium:light (enriched with WT over control), it can be seen that for the same amount of recovered bait, both mutant-specific and WT-specific interactors are identified. Environmental contaminants fall in the bottom left corner of the graph, while the affinity matrix contaminants cluster over log ratios of 0 in both conditions (hashed circle).

peptides, low signal:noise and errors in the quantification algorithms themselves,

A more stringent approach to thresholding AP-MS data sets is the calculation of "significance" values, which measure the deviation from the bulk of distribution and thus highlight those protein ratios that are outliers in comparison to the distribution of all protein ratios. When intensity is taken into account, the statistical spread of highly abundant proteins is much more focused than for less abundant proteins, and thus a protein with a ratio close to threshold may or may not be judged to vary significantly from the mean, depending on its abundance.

In Fig. 3C, proteins with ratios that vary significantly from the mean are highlighted, for three different confidence

values. In this case the RPL proteins still fall below threshold, although they are genuine members of the complex. A large number of other proteins also fall in the same range and it is likely that many are genuine interactors. Thus, although quantitative AP-MS can identify some interactors with high confidence, this technique also benefits from additional filtering approaches that permit mining of the data below threshold. This can include gene ontology analysis to highlight subthreshold interactors that are functionally relevant to the bait protein.

Another powerful application of quantitative AP-MS is the direct comparison of the interactomes of 2 or more proteins in the same experiment. In the example shown in Fig. 3D, a SILAC IP approach was used to pulldown both a wild type (from an R6K4-encoded extract) and mutant (from an R10K8-encoded extract) version of a tagged protein, compared to a pulldown of the tag alone (from an R0K0-encoded extract). Although the heavy (mutant):medium (wild type) ratios can be calculated directly and plotted versus intensity (not shown here), in this case the relative enrichments above background are compared directly by plotting heavy:light versus medium:light ratios. Common contaminants cluster in the center around a log ratio of 0 (hashed circle), and the bait proteins are shown to have been enriched to the same extent in the two IPs (based on ratios calculated for those peptides shared by both the wild-type and mutant proteins). Shared interactors similarly cluster near the bait, while mutant-specific and wild-type specific interactors can be clearly identified by their selective enrichment in one of the two pulldowns (Fig. 3D, arrows). With the proper controls and experimental design, multiplexed label-based AP-MS can also be extended to the comparison of the interactome of the same protein over time (e.g. throughout the cell cycle), or in response to cellular perturbation.

2.2 Mining below the threshold

Although labeling strategies provide great help in separating specific from nonspecific interactors, we have shown here that specific interactions cannot always be unambiguously determined, particularly near the threshold level where signal-to-noise ratios are close to background. AP-MS can detect protein–protein interactions of different strengths and stoichiometries, and also indirect interactions occurring via inclusion in the same multiprotein complex. Therefore, a broad range of enrichment profiles is expected in every experiment. A variety of approaches have been developed to help identify these genuine interactors, including methods to increase the signal, i.e. the abundance of purified protein complexes, and methods of reducing or filtering out the noise, i.e. proteins that bind nonspecifically to the affinity matrix, tag, and/or antibody.

The first filter that can be applied to any data set is the list of proteins that are there by accident or through unavoidable contamination, such as proteins utilized in the workflow (e.g. serum proteins, trypsin), proteins added through dust or physical contact (e.g. keratins) and MW standards carried over from gels. The Common Repository of Adventitious Proteins (cRAP) published by the Global Proteome machine currently comprises a downloadable list of 111 proteins that fit these categories, while the Max Planck Institute provides a downloadable list of 262 common contaminants.

Once environmental contaminants have been removed, the remainder of the background contaminants are presumed to bind nonspecifically to the affinity matrix or pellet with it during processing, possibly due to aggregation or precipitation out of the extract. As increasingly large numbers of AP-MS data sets are generated and shared in common repositories, more specific filters will be developed to mine the subthreshold. One example is the "bead proteome" that we generated for the protein G-sepharose matrix, which was based on a systematic assessment of proteins observed to bind in a ~1:1 ratio in 27 independent quantitative IP experiments [13]. The data were further annotated to compare common contaminants observed with both HeLa and U2OS cell extracts, and also cytoplasmic versus nucleoplasmic versus whole cell extracts. This static proteome has since been extended to a "protein frequency library", which is a dynamic tool that can be filtered for specific experimental parameters to generate a customized, objective library to discriminate between contaminants and specifically bound proteins ([81]; http://peptracker.com/datavisual/). Other computational methods based on confidence assessment and probabilistic scoring include Decontaminator [82], Significance Analysis of Interactome (SAINT; [83]), and the use of applied expectation maximization algorithms [84]. For high-throughput data sets, the topology of the protein–protein interaction network also helps to assess hits, as true interactors should have at least partially overlapping interactomes.

In an attempt to objectively determine whether the choice of affinity matrix impacts the amount and/or type of background contaminants, we also carried out a direct comparison of the bead proteomes of Protein G-conjugated agarose, Sepharose, and magnetic beads (Pierce #20398, GE Healthcare #17-0618-01, and Invitrogen #100-03D) [13]. Nonspecific protein binding was observed for all three matrices after either short (30 min) or long (18 h) incubation with cytoplasmic and nuclear extracts. A similar distribution of classes of contaminating proteins was observed at both time points, although the levels of protein binding increased after longer incubation, particularly for nuclear extracts (Fig. 4C–D). Some interesting difference were apparent in the relative performance of Sepharose and magnetic beads when incubated with either nuclear or cytoplasmic extracts, however a large number of proteins were enriched equally with both matrices. Aggregation/precipitation and thus inclusion in the proteome at the pelleting step may explain a subset of the Sepharose-enriched contaminants, as this list did show some overlap with proteins identified in loose and packed pellet fractions from ultracentrifuged nuclear extracts [23].

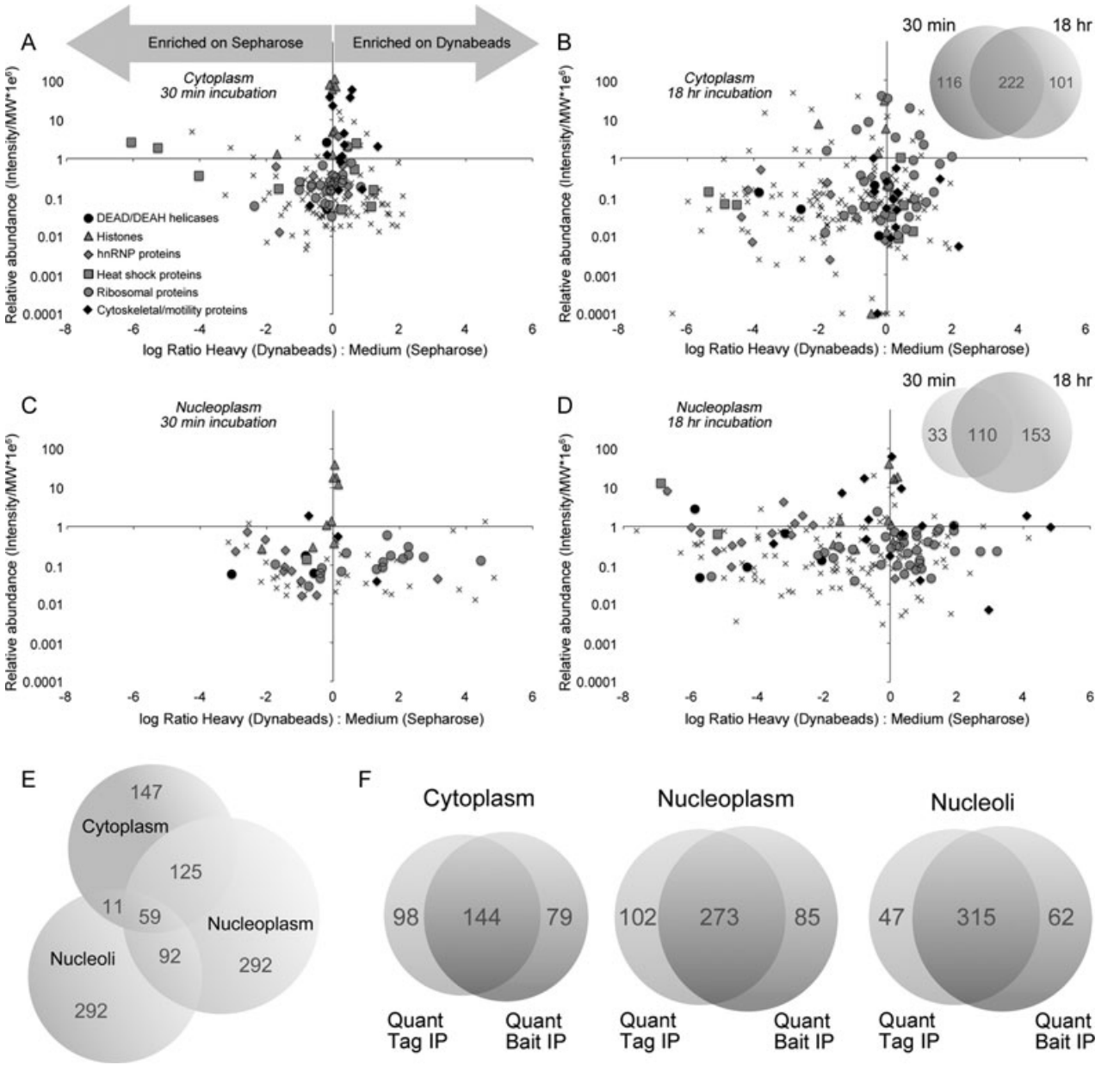


Figure 4. Investigating the source of background contaminants in AP-MS experiments. Panels (A)–(D) plot the distribution of proteins found to bind nonspecifically to Protein G-Sepharose (heavy-encoded) and/or Protein G-Dynabeads (medium-encoded) in cytoplasmic and nucleoplasmic HeLa cell extracts with either a brief preclearing incubation (30 min) or a longer incubation (18 h). Previously published families of common contaminants are highlighted in each data set. All extracts were clarified prior to incubation with the beads by centrifuging at 2800 × *g* (3500 rpm, Beckman GS-6, GH3.8 rotor) for 10 min at 4°C. A surprisingly similar number of proteins were identified/quantified in the 30 min cytoplasmic extract incubation (338) compared to the 18 h cytoplasmic extract incubation (323), whereas significantly fewer proteins were identified with a short nucleoplasmic extract incubation (143) compared to a longer nucleoplasmic extract incubation (263). Venn diagrams illustrate the significant overlap found between the short and long incubations (B and D, insets). (E) The overlap of Protein G-Sepharose bead contaminants identified in separate quantitative control experiments using cytoplasmic, nucleoplasmic and nucleolar HeLa cell extracts likely reflects the abundant and sticky proteins that are found in more than one cellular compartment, while a surprising number of unique contaminants were identified for each cell fraction. (F) Demonstration of the overlap between separate quantitative cell fraction-specific pulldowns of tagged bait protein versus tag alone. As shown here, a significant fraction of the proteins identified in the bait IP are also identified in the tag IP in cytoplasmic (~60%), nucleoplasmic (~70%), and nucleolar (~80%) experiments. Annotating these proteins as likely contaminants significantly enhances the ability to mine subthreshold data.

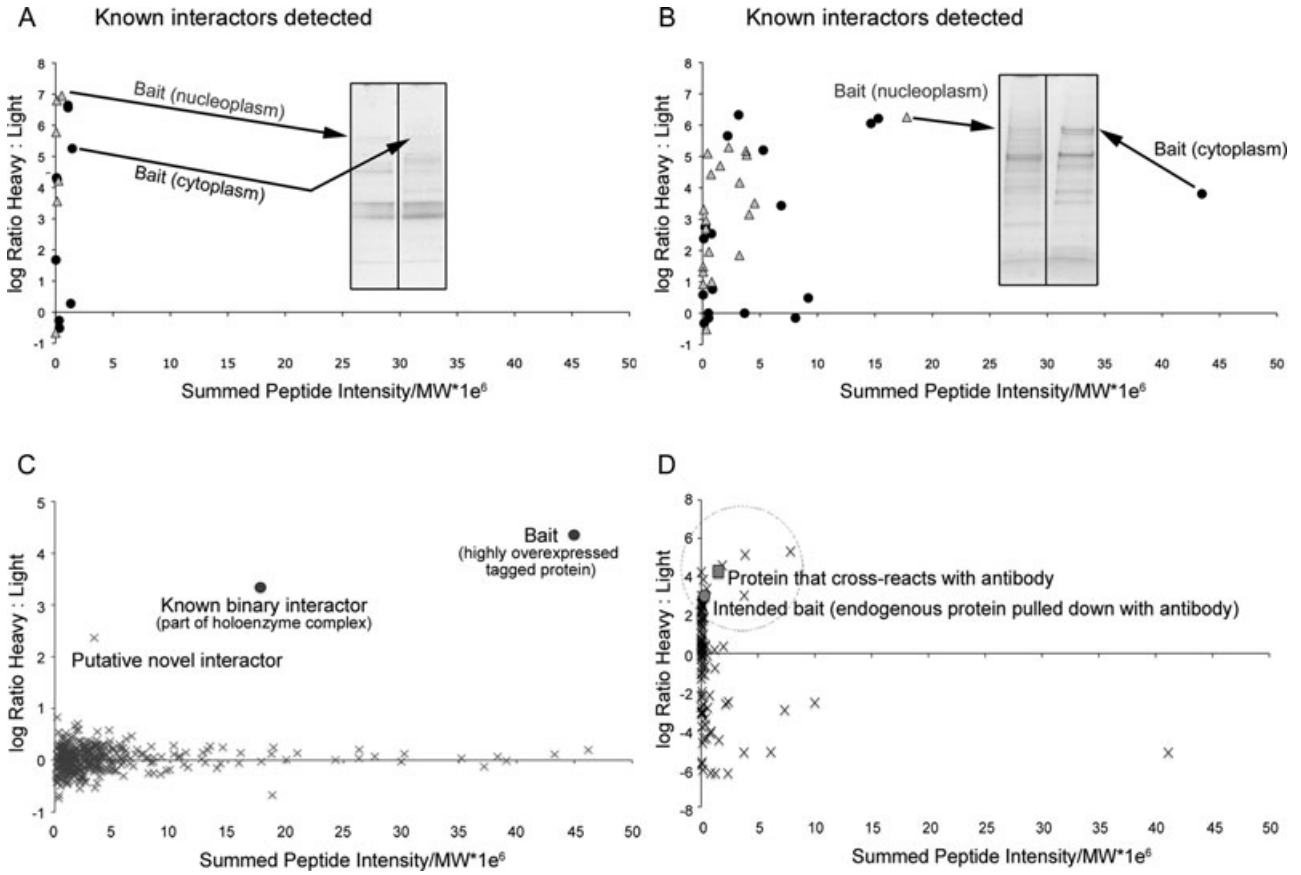


Figure 5. Factors that can affect the success of an AP-MS experiment. (A) Enrichment and abundance of known interactors detected in an IP of stably expressed GFP-SMN depleted from cytoplasmic (black circles) and nucleoplasmic (gray triangles) extracts using a monoclonal GFP antibody. Arrows indicate the position of the bait protein on the Coomassie-stained SDS-PAGE gel prior to tryptic digestion and LC-MS/MS analysis. A relatively small amount of bait protein was recovered (<50% depleted from the extracts), and subsequently only a handful of known interactors were detected by MS analysis. (B) When the experiment was repeated using a higher affinity reagent to recover significantly more bait protein from the same amount of starting material (nearly 100% depletion from both extracts), the bait was more readily detected on the Coomassie-stained gel prior to digestion (arrows). Consequently, all of its known interactors, both stable and dynamic, were identified. (C) In this example, an epitope-tagged bait protein was transiently over-expressed in cells at a high level prior to affinity purification. Although a significant amount of the bait was recovered, overexpression hampered its incorporation into endogenous signaling complexes. One known interactor (a binary partner in a holoenzyme complex) was highly enriched, as was one novel putative interactor. None of the protein's other known binding partners were detected. (D) In this example, an endogenous bait protein (filled circle) was affinity purified using a commercial peptide antibody. Although not predicted by a BLAST search, validation experiments revealed cross-reactivity and subsequent pulldown of an unrelated protein (filled square). Data analysis is complicated by the mix of protein complexes that co-purified with the two separate antigens (hashed circle).

It is also reasonable to assume that there will be differences in contaminant profiles between extracts from different cell types, or between different types of extracts from the same cell line, due to variations in relative protein abundance and other factors that affect nonspecific binding. Figure 4E shows the overlap between proteins identified in three separate pulldowns of the same tag control (free GFP) from cytoplasmic, nucleoplasmic, and nucleolar HeLa cell extracts prepared in RIPA buffer [85]. The presence of a significant number of extract-specific contaminants highlights the importance of separate contaminant profiles. Figure 4F demonstrates the overlap between these control GFP experiments and pulldowns of a GFP-tagged bait protein, which is ~60–80%. These experiments were designed as an additional step to filter the subthreshold data with heavy:light ratios close to the mean. Elution conditions will also influence the level and type of background contaminants. Affinity matrices such as FLAG antibody and Streptactin offer the potential for specific elution of bait-associated proteins under mild conditions using the FLAG peptide or biotin, with the caveat that these reagents should be removed prior to MS analysis.

Overall, it can be concluded that the affinity matrices constitute a major source of nonspecific protein binding for all protein interaction studies, and that no single type of bead is ideally suited to all applications. Rather, improved results

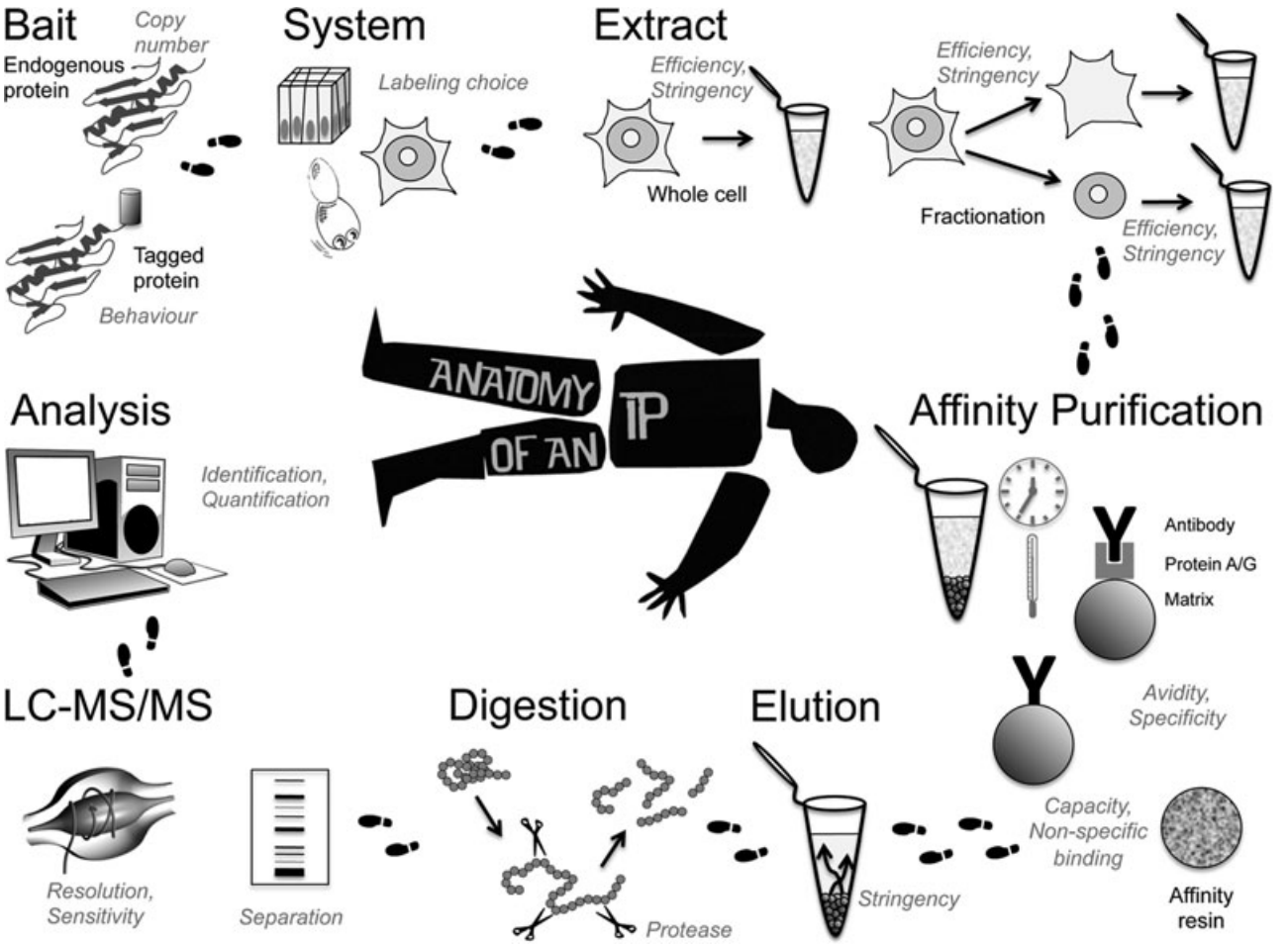


Figure 6. Anatomy of an IP. Diagram of the steps involved in a typical AP-MS experiment (label-free or labeling-based), with key issues highlighted in italics for each step. Taking all of these issues into account during the careful design of an AP-MS experiment will greatly increase the chances of success.

can be obtained by choosing the most suitable combination of reagents based on the specific details of the experiment to be performed, and by independently assessing cell type- and cell fraction-dependent contaminants.

2.4 When things go wrong

When properly designed and implemented, label-based quantitative AP-MS experiments have a high success rate. An important issue for maximizing the identification of protein interaction partners is ensuring both efficient isolation of the target protein under study and achieving a high signal-to-noise ratio. This is demonstrated in Fig. 5, in which known interactors detected in cytoplasmic and nuclear pulldowns of a GFP-tagged fusion of the well-characterized Survival of Motor Neuron (SMN) protein are plotted on a ratio versus abundance graph. The fusion protein was stably expressed in HeLa cells at slightly lower than endogenous levels, and recovered from the same amount of extract using either a monoclonal GFP antibody covalently conjugated to Protein G Sepharose (Fig. 5A) or the high affinity GFP-Trap_A® reagent (Fig. 5B; Chromotek #gta-20). As indicated on the inset Coomassie-stained gels and on the graphs, the latter enriched a substantially larger amount of the bait protein, with a concomitant increase in the number of interactors that were identified, along with the total number of peptides detected for each. Thus, although the original experiment did work (SMN and interactors were identified), increasing the signal:noise ratio by recovering more protein permitted mapping of the entire known SMN interactome and identification of a novel interactor [13].

It should be noted that transiently overexpressing high amounts of a bait protein is unlikely to increase interactome coverage, unless there is a concomitant upregulation of all of its binding partners. In addition, with overexpression it is possible that nonphysiological protein interactions will occur and be detected during MS analysis. An example is shown in Fig. 5C. Although a substantial amount of the bait protein (transiently over-expressed at high levels) was recovered, only

two putative interaction partners were detected, one of which is a high affinity binding partner with which the bait forms a holoenzyme complex (and which was likely recruited away from its other intracellular binding partners). The fact that the numerous other known binding partners for this particular protein were not detected indicates that it was not incorporated efficiently into its endogenous multiprotein complexes.

Labeling approaches can also be successfully applied to the analysis of endogenous proteins directly immunoprecipitated with antibodies. However, the overall quality of the resulting data will inevitably be affected by the specificity and efficiency of the available antibodies. This is illustrated in Fig. 5D, which shows a data set that initially looked quite promising. The relatively low abundance of the bait protein was not a particular concern, as it was known to be present at a low copy number. Validation of the putative interactors, however, revealed that one of them cross-reacted with the antibody used to IP the bait protein. Thus, the interactors identified in this screen represent the mix of complexes that co-purified with either this protein or with the intended bait. This substantially complicated the data analysis and wasted time, money, and reagents.

3 Concluding remarks and future perspectives

Strategies combining label-based AP-MS with bead proteome filtering and enhanced data analysis procedures can reliably be used to characterize specific protein interaction partners while using less stringent isolation procedures that minimize the loss of lower affinity and lower abundance proteins. Fractionation and additional controls can extend this to resolution of interaction events confined to specific cellular compartments. Figure 6 summarizes key factors that need to be taken into account and decisions that need to made at each stage of a label-based AP-MS experiment. Even when conditions must be varied, general principles still apply, including the importance of maintaining short incubation times while optimizing overall efficiency during affinity purification and recognizing that the efficiency and stringency of protein extraction will impact the preservation and detection of specific multiprotein complexes. Inevitable differences in the biochemical properties of different proteins mean that no unique isolation protocol may be ideal in every case, however high-throughput studies have already been carried out using similar isolation protocols to successfully analyze hundreds of interactomes in parallel. And with the number of protein interactions in the human cell estimated at 375 000 [86], we still have a long way to go!

The author thanks Lawrence Puente at the Ottawa Hospital Research Institute Proteomics Core Facility and Douglas Lamond and Kenneth Beattie at the University of Dundee Fingerprints Proteomics Facility for technical support. This work was supported in part by a Canadian Cancer Research/Terry Fox Foundation New Investigator Award (Ref: 20148) and NSERC Discovery Grant (Ref: 372370). LTM holds a CIHR New Investigator Salary Support Award.

The authors have declared no conflict of interest.

4 References

[1] Zhong, J., Chaerkady, R., Kandasamy, K., Gucek, M. et al., The interactome of a PTB domain-containing adapter protein, Odin, revealed by SILAC. *J. Proteomics* 2011, *74*, 294–303.

[2] Chamousset, D., De Wever, V., Moorhead, G. B., Chen, Y. et al., RRP1B targets PP1 to mammalian cell nucleoli and is associated with Pre-60S ribosomal subunits. *Mol. Biol. Cell* 2010, *21*, 4212–4226.

[3] Bezstarosti, K., Ghamari, A., Grosveld, F. G., Demmers, J. A. A., Differential proteomics based on ^{18}O Labeling to determine the Cyclin Dependent Kinase 9 Interactome. *J. Proteome Res.* 2010, *9*, 4464–4475.

[4] Malovannaya, A., Lanz, R. B., Jung, S. Y., Bulynko, Y. et al., Analysis of the human endogenous coregulator complexome. *Cell* 2011, *145*, 787–799.

[5] Hubner, N. C., Bird, A. W., Cox, J., Splettstoesser, B. et al., Quantitative proteomics combined with BAC TransgeneOmics reveals *in vivo* protein interactions. *J. Cell Biol.* 2010, *189*, 739–754.

[6] Ho, Y., Gruhler, A., Heilbut, A., Bader, G. et al., Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 2002, *415*, 180–183.

[7] Gavin, A.-C., Bösche, M., Krause, R., Grandi, P. et al., Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002, *415*, 141–147.

[8] Bicho, C., de Lima Alves, F., Chen, Z., Rappsilber, J., Sawin, K., A genetic engineering solution to the "Arginine Conversion Problem" in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *Mol. Cell. Proteomics* 2010, *9*, 1567–1577.

[9] Graumann, J., Applicability of tandem affinity purification MudPIT to pathway proteomics in Yeast. *Mol. Cell. Proteomics* 2004, *3*, 226–237.

[10] Gingras, A.-C., Gstaiger, M., Raught, B., Aebersold, R., Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell. Biol.* 2007, *8*, 645–654.

[11] Paul, F. E., Hosp, F., Selbach, M., Analyzing protein–protein interactions by quantitative mass spectrometry. *Methods* 2011, *54*, 387–395.

[12] Kaake, R., Wang, X., Huang, L., Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. *Mol. Cell. Proteomics* 2010, *9*, 1650–1665.

[13] Trinkle-Mulcahy, L., Boulon, S., Lam, Y. W., Urcia, R. et al., Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* 2008, *183*, 223–239.

[14] Domon, B., Mass spectrometry and protein analysis. *Science* 2006, *312*, 212–217.

[15] Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O. et al., Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* 2011, *10*, M111.011015

[16] Selbach, M., Mann, M., Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK). *Nat. Meth* 2006, *3*, 981–983.

[17] Tackett, A. J., DeGrasse, J. A., Sekedat, M. D., Oeffinger, M. et al., I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. *J. Proteome Res.* 2005, *4*, 1752–1756.

[18] Vermeulen, M., Hubner, N. C., Mann, M., High confidence determination of specific protein–protein interactions using quantitative mass spectrometry. *Curr. Opin. Biotechnol.* 2008, *19*, 331–337.

[19] Cox, J., Mann, M., Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* 2011, *80*, 273–299.

[20] Collins, M. O., Choudhary, J. S., Mapping multiprotein complexes by affinity purification and mass spectrometry. *Curr. Opin. Biotechnol.* 2008, *19*, 324–330.

[21] Brizzard, B., Epitope tagging. *Biotech.* 2008, *44*(Suppl.), 693–695.

[22] Tsai, A., Carstens, R. P., An optimized protocol for protein purification in cultured mammalian cells using a tandem affinity purification approach. *Nat. Protoc.* 2007, *1*, 2820–2827.

[23] Malovannaya, A., Li, Y., Bulynko, Y., Jung, S. Y., et al., Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc. Nat. Acad. Sci.* 2010, *107*, 2431–2436.

[24] Cristea, I. M., Fluorescent proteins as proteomic probes. *Mol. Cell Proteomics* 2005, *4*, 1933–1941.

[25] Cheeseman, I. M., A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Sci. STKE* 2005, *266*, 1–15.

[26] Trinkle-Mulcahy, L., Andersen, J., Lam, Y. W., Moorhead, G. et al., Repo-Man recruits PP1 gamma to chromatin and is essential for cell viability. *J. Cell Biol.* 2006, *172*, 679–692.

[27] Poser, I., Sarov, M., Hutchins, J. R. A., Hériché, J.-K. et al., BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Meth.* 2008, *5*, 409–415.

[28] Hubner, N. C., Mann, M., Extracting gene function from protein–protein interactions using Quantitative BAC InteraCtomics (QUBIC). *Methods* 2011, *53*, 453–459.

[29] Schulze, W. X., A novel proteomic screen for peptide–protein interactions. *J. Biol. Chem.* 2003, *279*, 10756–10764.

[30] Mittler, G., Butter, F., Mann, M., A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res.* 2008, *19*, 284–293.

[31] Butter, F., Sheibe, M., Mörl, M., Mann, M., Unbiased RNA–protein interaction screen by quantitative proteomics. *PNAS* 2009, *106*, 10626–10631.

[32] Tsai, B. P., Wang, X., Huang, L., Waterman, M. L., Quantitative profiling of *In Vivo*-assembled RNA-Protein complexes using a novel integrated proteomic approach. *Mol. Cell Proteomics* 2011, *10*, M110.007385

[33] Yang, W., Steen, H., Freeman, M. R., Proteomic approaches to the analysis of multiprotein signaling complexes. *Proteomics* 2008, *8*, 832–851.

[34] Domon, B., Aebersold, R., Options and considerations when selecting a quantitative proteomics strategy. *Nat. Publ. Group* 2010, *28*, 710–721.

[35] Mueller, L. N., Brusniak, M.-Y., Mani, D. R., Aebersold, R., An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J. Proteome Res.* 2008, *7*, 51–61.

[36] Huang, X., Liu, M., Nold, M. J., Tian, C. et al., Software for quantitative proteomic analysis using stable isotope labeling and data independent acquisition. *Anal. Chem.* 2011, *83*, 6971–6979.

[37] Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T. et al., A guided tour of the Trans-Proteomic Pipeline. *Proteomics* 2010, *10*, 1150–1159.

[38] Cox, J., Matic, I., Hilger, M., Nagaraj, N. et al., A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protoc.* 2009, *4*, 698–705.

[39] Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A. et al., Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 2011, *10*, 1794–1805.

[40] Gygi, S., Rist, B., Gerber, S., Turecek, F. et al., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999, *17*, 994–999.

[41] Ranish, J. A., Yi, E. C., Leslie, D. M., Purvine, S. O. et al., The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* 2003, *33*, 349–355.

[42] Brand, M., Ranish, J. A., Kummer, N. T., Hamilton, J. et al., Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat. Struct. Mol. Biol.* 2003, *11*, 73–80.

[43] Himeda, C. L., Ranish, J. A., Hauschka, S. D., Quantitative proteomic identification of MAZ as a transcriptional regulator of muscle-specific genes in skeletal and cardiac myocytes. *Mol. Cell. Biol.* 2008, *28*, 6521–6535.

[44] Aggarwal, K., Shotgun proteomics using the iTRAQ isobaric tags. *Brief. Func. Genomics Proteomics* 2006, *5*, 112–120.

[45] Pflieger, D., Jünger, M., Müller, M., Rinner, O. et al., Quantitative proteomic analysis of protein complexes: CONCURRENT identification of interactors and their state of phosphorylation. *Mol. Cell. Proteomics* 2008, *7*, 326–346.

[46] Butler, G., Dean, R., Morrison, C., Overall, C., *Identification of Cellular MMP Substrates Using Quantitative Proteomics: Isotope-Coded Affinity Tags (ICAT) and Isobaric Tags for Relative and Absolute Quantification (iTRAQ)*, vol. *622*, Humana Press, Totowa, NJ 2010.

[47] Bai, Y., Markham, K., Chen, F., Weerasekera, R. et al., The *in vivo* Brain Interactome of the Amyloid Precursor Protein. *Mol. Cell. Proteomics* 2008, *7*, 15–34.

[48] Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S. et al., Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* 2007, *25*, 1035–1044.

[49] Molina, H., Yang, Y., Ruch, T., Kim, J.-W. et al., Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy. *J. Proteome Res.* 2009, *8*, 48–58.

[50] Ishihama, Y., Sato, T., Tabata, T., Miyamoto, N. et al., Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nat. Biotechnol.* 2005, *23*, 617–621.

[51] Geiger, T., Wisniewski, J. R., Cox, J., Zanivan, S. et al., Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat. Protoc.* 2011, *6*, 147–157.

[52] Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Meth.* 2010, *7*, 383–385.

[53] Blagoev, B., Mann, M., Quantitative proteomics to study mitogen-activated protein kinases. *Methods* 2006, *40*, 243–250.

[54] Lossner, C., Warnken, U., Pscherer, A., Schnölzer, M., Preventing arginine-to-proline conversion in a cell-line-independent manner during cell cultivation under stable isotope labeling by amino acids in cell culture (SILAC) conditions. *Anal. Biochem.* 2011, *412*, 123–125.

[55] Bendall, S. C., Hughes, C., Stewart, M. H., Doble, B. et al., Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. *Mol. Cell Proteomics* 2008, *7*, 1587–1597.

[56] Van Hoof, D., Pinkse, M., Ward-Van Oostwaard, D., Mummery, C. et al., An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat. Meth.* 2007, *4*, 677–678.

[57] Park, S., Liao, L., Kim, J., Yates, J., III, A computational approach to correct arginine-to-proline conversion in quantitative proteomics. *Nat. Meth.* 2009, *6*, 185–186.

[58] van Breukelen, B., van den Toorn, H. W. P., Drugan, M. M., Heck, A. J. R., StatQuant: a post-quantification analysis toolbox for improving quantitative mass spectrometry. *Bioinformatics* 2009, *25*, 1472–1473.

[59] Mortensen, P., Gouw, J. W., Olsen, J. V., Ong, S.-E. et al., MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J. Proteome Res.* 2010, *9*, 393–403.

[60] Larance, M., Bailly, A., Pourkarimi, E., Hay, R. et al., Stable isotope labeling with amino acids in Nematodes. *Nat. Methods* 2011, *8*, 849–851.

[61] Nikolov, M., Stützer, A., Mosch, K., Krasauskas, A. et al., Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns. *Mol. Cell. Proteomics* 2011, *10*, M110.005371

[62] Gunaratne, J., Goh, M. X., Swa, H. L. F., Lee, F. Y. et al., Protein interactions of phosphatase and tensin homologue (PTEN) and its cancer-associated G20E mutant compared by using stable isotope labeling by amino acids in cell culture-based parallel affinity purification. *J. Biol. Chem.* 2011, *286*, 18093–18103.

[63] Andersen, J., Matic, I., Vertegaal, A., *Identification of SUMO Target Proteins by Quantitative Proteomics*, Vol. *497*, Humana Press, Totowa, NJ 2009.

[64] Guerrero, C., An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified *in vivo* Cross-linked Protein Complexes (qtax) to Decipher the 26 s Proteasome-interacting network. *Mol. Cell Proteomics* 2005, *5*, 366–378.

[65] Wang, X., Huang, L., Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. *Mol. Cell. Proteomics* 2008, *7*, 46–57.

[66] Mousson, F., Kolkman, A., Pijnappel, W. W. M. P., Timmers, H. T. M., Heck, A. J. R., Quantitative proteomics reveals regulation of dynamic components within TATA-binding Protein (TBP) transcription complexes. *Mol. Cell Proteomics* 2008, *7*, 845–852.

[67] Fang, L., Wang, X., Yamoah, K., Chen, P.-L. et al., Characterization of the Human COP9 signalosome complex using affinity purification and mass spectrometry. *J. Proteome Res.* 2008, *7*, 4914–4925.

[68] Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. et al., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 2002, *1*, 376–386.

[69] Sury, M., Chen, J.-X., Selbach, M., The SILAC fly allows for accurate protein quantification *in Vivo*. *Mol. Cell. Proteomics* 2010, *9*, 2173–2183.

[70] Gouw, J. W., Krijgsveld, J., Heck, A. J. R., Quantitative proteomics by metabolic labeling of model organisms. *Mol. Cell Proteomics* 2010, *9*, 11–24.

[71] Engen, J., Bradbury, E., Chen, X., Using stable-isotope-labeled proteins for hydrogen exchange studies in complex mixtures. *Anal. Chem.* 2002, *74*, 1680–1686.

[72] Dreisbach, A., Otto, A., Becher, D., Hammer, E. et al., Monitoring of changes in the membrane proteome during stationary phase adaptation of Bacillus subtilis using *in vivo* labeling techniques. *Proteomics* 2008, *8*, 2062–2076.

[73] de Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L. et al., Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 2008, *455*, 1251–1254.

[74] Georgianna, D. R., Hawkridge, A. M., Muddiman, D. C., Payne, G. A., Temperature-dependent regulation of proteins in *Aspergillus flavus*: whole organism stable isotope labeling by amino acids. *J. Proteome Res.* 2008, *7*, 2973–2979.

[75] Prieto, J. H., Koncarevic, S., Park, S. K., Yates, J., Becker, K., Large-Scale Differential Proteome Analysis in Plasmodium falciparum under Drug Treatment. *PLoS ONE* 2008, *3*, e4098.

[76] Gruhler, A., Schulze, W., Matthiesen, R., Mann, M., Jensen, O., Stable isotope labeling of *Arabidopsis thaliana* cells and quantitative proteomics by mass spectrometry. *Mol. Cell. Proteomics* 2005, *4*, 1697–1709.

[77] Schutz, W., Hausmann, N., Krug, K., Hampp, R., Macek, B., Extending SILAC to proteomics of plant cell lines. *The Plant Cell Online* 2011, *23*, 1701–1705.

[78] Kruger, M., Moser, M., Ussar, S., Thievessen, I. et al., SILAC mouse for quantitative proteomics uncovers Kindlin-3 as an essential factor for red blood cell function. *Cell* 2008, *134*, 353–364.

[79] Doherty, M. K., Whitehead, C., McCormack, H., Gaskell, S. J., Beynon, R. J., Proteome dynamics in complex organisms:

using stable isotopes to monitor individual protein turnover rates. *Proteomics* 2005, *5*, 522–533.

[80] Looso, M., Borchardt, T., Kruger, M., Braun, T., Advanced identification of proteins in uncharacterized proteomes by pulsed *in Vivo* stable isotope labeling-based mass spectrometry. *Mol. Cell Proteomics* 2010, *9*, 1157–1166.

[81] Boulon, S., Ahmad, Y., Trinkle-Mulcahy, L., Verheggen, C. et al., Establishment of a protein frequency library and its application in the reliable identification of specific protein interaction partners. *Mol. Cell Proteomics* 2010, *9*, 861–879.

[82] Lavallée-Adam, M., Cloutier, P., Coulombe, B., Blanchette, M., Modeling contaminants in AP-MS/MS experiments. *J. Proteome Res.* 2011, *10*, 886–895.

[83] Choi, H., Larsen, B., Lin, Z.-Y., Breitkreutz, A. et al., SAINT: probabilistic scoring of affinity purification–mass spectrometry data. *Nat. Meth.* 2010, *8*, 70–73.

[84] Kim, B., Nesvizhskii, A., Rani, P., Hahn, S. et al., The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc. Nat. Acad. Sci.* 2007, *104*, 16063–16073.

[85] Chamousset, D., Mamane, S., Boisvert, F.-M., Trinkle-Mulcahy, L., Efficient extraction of nucleolar proteins for interactome analyses. *Proteomics* 2010, *10*, 3045–3050.

[86] Ramani, A., Bunescu, R., Mooney, R., Marcotte, E., Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol.* 2005, *6*, R40.

[87] Abu-Farha, M., Elisma, F., Figeys, D., *Advances in Biochemical Engineering/Biotechnology*, Springer Berlin Heidelberg, Berlin, Heidelberg 2008, Vol. *110*.

[88] Einarson, M. B., Pugacheva, E. N., Orlinick, J. R., Identification of protein-protein interactions with Glutathione-S-Transferase (GST) Fusion Proteins. *Cold Spring Harbor Protocols* 2007, doi:10.1101/pdb.top11.

[89] Carey, M. F., Peterson, C. L., Smale, S. T., Purification of epitope-tagged transcription factor IID. *Cold Spring Harbor Protocols* 2010, doi:10.1101/pdb.prot5450.

[90] Kirkpatrick, D. S., Weldon, S. F., Tsaprailis, G., Liebler, D. C., Gandolfi, A. J., Proteomic identification of ubiquitinated proteins from human cells expressing His-tagged ubiquitin. *Proteomics* 2005, *5*, 2104–2111.

[91] Schmidt, T. G., Skerra, A., The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* 2007, *2*, 1528–1535.

