

# Analysis of protein complexes using mass spectrometry

Anne-Claude Gingras\*, Matthias Gstaiger†, Brian Raught§ and Ruedi Aebersold†¶

**Abstract** | The versatile combination of affinity purification and mass spectrometry (AP–MS) has recently been applied to the detailed characterization of many protein complexes and large protein–interaction networks. The combination of AP–MS with other techniques, such as biochemical fractionation, intact mass measurement and chemical crosslinking, can help to decipher the supramolecular organization of protein complexes. AP–MS can also be combined with quantitative proteomics approaches to better understand the dynamics of protein–complex assembly.

The past decade has seen the complete sequencing of several eukaryotic genomes, providing a comprehensive inventory of predicted proteins for many different species. However, such protein lists are not sufficient to describe biological processes. Vital cellular functions such as DNA replication, transcription and mRNA translation require the coordinated action of a large number of proteins that are assembled into an array of multiprotein complexes of distinct composition and structure. Similarly, biological processes are orchestrated and regulated by dynamic signalling networks of interacting proteins that link chemical or physical stimuli to specific effector molecules. The analysis of protein complexes and protein–protein interaction networks — and the dynamic behaviour of these networks as a function of time and cell state — are therefore of central importance in biological research. Various large-scale efforts have thus attempted to define protein interactomes in several organisms, including *Saccharomyces cerevisiae*<sup>1–6</sup>, *Drosophila melanogaster*<sup>7–9</sup>, *Caenorhabditis elegans*<sup>10</sup> and *Homo sapiens*<sup>11–13</sup>.

Different approaches have been used to characterize protein complexes and protein–protein interaction networks. The first interactome maps were obtained using the yeast two-hybrid system (the yeast two-hybrid approach and its strengths and weaknesses are described elsewhere<sup>14–16</sup>). More recently, a combination of affinity purification and mass spectrometry (AP–MS) has been used to greatly advance our understanding of protein–complex composition. With the AP–MS method, multiprotein complexes are isolated directly from cell lysates through one or more AP steps. Complex components are then identified by MS (FIG. 1). In contrast to yeast two-hybrid and related methods, AP–MS can be performed under near physiological conditions, in the relevant

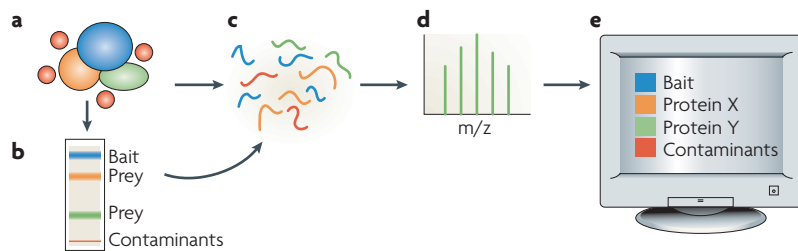
organism and cell type. AP–MS does not typically perturb relevant post-translational modifications, which are often crucial for the organization and/or activity of complexes. These post-translational modifications can also be identified by MS. Another advantage of AP–MS is that it can be used to probe dynamic changes in the composition of protein complexes, especially when used in combination with quantitative proteomics techniques<sup>17–22</sup>.

Two recent reports<sup>2,4</sup> describe high-confidence AP–MS data that connect an estimated 60% of the yeast proteome, and these studies demonstrate that large-scale protein–interaction mapping by MS is feasible. So, it is timely to reflect on AP–MS approaches and to outline some emerging strategies that are designed to better exploit this powerful technique. In this review, we briefly summarize experimental design for AP–MS, and we discuss how AP–MS data can be used to characterize the components of protein complexes. We then describe how AP–MS can be combined with other strategies (such as biochemical fractionation, co-elution profiles and chemical crosslinking) to reveal protein–complex stoichiometry and structural organization. Last, we discuss the use of quantitative proteomics to study the dynamics of protein–complex composition.

## AP–MS: the technique

Several recent advances have enabled researchers to successfully apply AP–MS techniques to high-throughput acquisition of interaction data, especially in *S. cerevisiae*, in which epitope-tagged proteins (or baits) can be routinely generated for many open reading frames (ORFs). These advances also include the isolation of protein complexes using generic protocols and powerful MS methods for the identification and quantification of isolated proteins.

\*Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Ave, Room 992A, Toronto, Ontario, M5G 1X5, Canada; and Department of Medical & Molecular Genetics and Graduate Department of Molecular & Medical Genetics, University of Toronto, Toronto, Canada. †Institute of Molecular Systems Biology, Swiss Federal Institute of Technology, ETH Hönggerberg HPT E 78, CH-8093 Zürich, Switzerland; and Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland. ‡University Health Network, Ontario Cancer Institute and McLaughlin Centre for Molecular Medicine, 101 College St, MaRS TMDT 9-805, Toronto, Ontario, M5G 1L7, Canada; and Department of Medical Biophysics, University of Toronto, Toronto, Canada. §Institute for Systems Biology, 1441 N. 34th St, Seattle, Washington 98103, USA; and Faculty of Science, University of Zurich, Switzerland. Correspondence to A.-C.G. e-mail: gingras@mshri.on.ca doi:10.1038/nrm2208 Published online 27 June 2007



**Figure 1 | General overview of an affinity purification and mass spectrometry experiment.** **a** | The protein of interest (often epitope tagged; blue) is purified from a cell lysate together with its binding partners (orange and green). Contaminants (red) can also be present. **b** | In an optional step, proteins in the complex can be separated by SDS-PAGE (followed by silver or coomassie staining) or by some type of liquid chromatography. Although analysis of gel-purified proteins has been used most often so far, gel-free approaches allow for a more rapid and generic analysis and are increasingly used. **c** | Proteins are subjected to proteolysis (usually with trypsin). **d** | Mass spectrometry (MS) analysis of peptides. In most cases, this involves peptide separation by reversed-phase liquid chromatography followed by two MS events: in the first scan, the mass/charge ratio ( $m/z$ ) of the intact peptide is measured. The most abundant peptides are then specifically selected and subjected to fragmentation, yielding a tandem MS (MS/MS) spectrum (a simplified MS/MS scan is shown for one of the peptides). **e** | Database searching and statistical software are used to interpret the MS data to yield a list of proteins that were present in the initial sample, including the tagged protein, its interacting partners and contaminants.

**The first step — affinity purification.** Generic approaches that use affinity-tagged recombinant proteins have allowed for parallel sample preparation without the need to optimize the purification protocol for each protein complex. Proteins of interest are simply expressed in-frame with an epitope tag (at either the N or C terminus), which is then used as an affinity handle to purify the tagged protein (the bait) along with its interacting partners (the prey). Although several different tags or tag combinations have been successfully used in many low-throughput studies (see REF. 23), high-throughput studies have primarily used either the flag tag<sup>3,23</sup> or the tandem affinity purification (TAP) tag<sup>2,4,5</sup> system (BOX 1).

In the flag-tag approach, as used by Ho *et al.*, C-terminally flag-tagged proteins were expressed under the control of a GAL-inducible promoter and isolated in a single step using an anti-flag antibody resin<sup>3</sup>. Tagging 10% of the yeast ORFs, the authors were able to connect 25% of the yeast proteome. In the TAP-tag approach, as used by Gavin *et al.*<sup>4,5</sup> and Krogan *et al.*<sup>2</sup>, yeast genes for the proteins of interest were fused to a C-terminal dual-epitope tag via homologous recombination, such that the proteins were expressed under their own promoters. Protein purification was carried out in two steps, first via the protein A moiety in the TAP tag (which binds immunoglobulin G (IgG)–sepharose), and then via the calmodulin-binding peptide (which exhibits high affinity to calmodulin–sepharose; BOX 1). Further discussion on false positives and false negatives in AP-MS experiments can be found in BOX 2.

**The second step — mass spectrometry.** Two main strategies to ionize peptide ions, electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), and their implementation on several types of tandem mass

spectrometers<sup>24</sup>, have allowed for efficient sequencing of peptides derived from proteolytic digests of protein complexes (an in-depth discussion of MS is beyond the scope of this review; we refer the reader to excellent recent publications<sup>25–29</sup>).

MS is currently the method of choice for peptide sequencing because it is sensitive; it routinely allows for the identification of peptides that are present at femtomole levels. MS is also rapid; sequencing of individual peptides can be achieved within hundreds of milliseconds, and thousands of peptides can therefore be identified in a single MS run. Last, MS is compatible with high-throughput strategies and is easily automated. It also allows for the characterization of peptide modifications (including naturally occurring post-translational modifications, such as phosphorylation, and exogenously added modifications, such as chemical crosslinkers). MS can also be adapted to quantitatively measure peptide abundance and does not require pre-existing knowledge of the proteins to be analysed. Advances in sample processing and instrumentation have gone hand-in-hand with the development of software tools that automatically retrieve sequence information from acquired mass spectra and provide statistical validation of the accuracy of the determined sequences<sup>30,31</sup>.

Although individual research groups have used different methods for the analysis of protein complexes by MS, the basic principles are essentially the same (FIG. 1).

### From interactors to complexes

The AP-MS technique only generates a list of proteins detected in a given sample, and does not necessarily reveal the composition of individual protein complexes. The data from a single AP-MS experiment represents an average of binding partners and protein complexes. If the bait protein is a component of multiple alternative complexes, a single AP-MS analysis cannot be used to decipher this multiplicity of associations. This is an important limitation because proteins can have dramatically different roles as components of different types of complexes.

**Identifying PP2A-interacting partners.** The acquisition of high-density interaction data sets in which each component (or multiple components) of a particular complex is affinity tagged and purified can greatly assist in deciphering the association of a given protein with multiple alternative complexes. For example, TAP-tag AP-MS was used to identify the interacting partners for the catalytic (C) subunit of the human serine/threonine phosphatase PP2A (FIG. 2; A.-C.G., B.R. and R.A., unpublished observations). PP2A is an important serine/threonine phosphatase in eukaryotes and generally functions as a trimer in which the C subunit is associated with a regulatory protein (B, B' or B''); 14 genes in humans) through an adaptor (A) molecule<sup>32,33</sup>. Mutually exclusive interaction of the C subunit with a protein known as  $\alpha 4$  is also observed<sup>34–36</sup>. A single TAP-tag AP-MS of the PP2A C subunit revealed a large list of interactors without shedding light on individual protein–protein complexes. However, reciprocal TAP analyses of each of the proteins that were identified in

#### Interactomes

Global networks of interactions (protein–protein interactions in this article).

#### Yeast two-hybrid

A genetic technique that detects binary interactions between protein pairs.

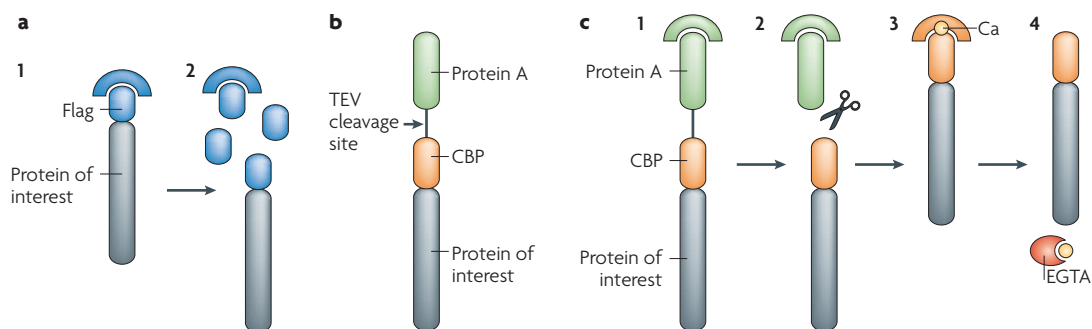
#### Flag tag

A short amino-acid sequence (usually 8–11 amino acids) that is recognized by monoclonal antibodies and is used to tag proteins for detection and purification.

#### Tandem affinity purification (TAP) tag

A recombinant fusion tag formed of two moieties, used sequentially for protein purification. The original version consists of the immunoglobulin G (IgG)-binding portion of *Staphylococcus aureus* protein A (distal tag) and the calmodulin-binding peptide (proximal tag); these tags are separated by a tobacco etch virus protease cleavage site.

## Box 1 | Affinity-purification strategies used in high-throughput studies



Although a number of affinity tags and tag combinations have been used in low-throughput studies, recent high-throughput studies in yeast have primarily used two tagging systems: a single flag epitope (8–11 amino acids) or the tandem affinity purification (TAP) system, which adds a >20-kDa tag. Both purification approaches have advantages and disadvantages, and the appropriate technique should be selected depending on the goals of the experiment. For example, the single-step flag purification can preserve weaker or more transient protein–protein interactions. By contrast, despite the lengthier purification procedure, the two purification steps of TAP dramatically decrease background levels (that is, proteins that bind non-specifically to the solid phase support or epitope tag), and thereby increases confidence in the interactions.

During flag purification, the protein of interest is expressed with a flag epitope and is purified on anti-flag sepharose (panel a). Elution is effected by the addition of a large molar excess of the competing flag peptide. In the original TAP tag, the protein of interest is expressed in-frame with a dual affinity moiety (panel b). The proximal tag (calmodulin-binding peptide (CBP)) is separated from the distal tag (protein A) by a protease cleavage site (specifically recognized by the tobacco etch virus protease (TEV)). During TAP purification (panel c), protein A is first immobilized on immunoglobulin G (IgG)–sepharose (c1). The tag is then cleaved with TEV (c2) and the protein of interest, now fused only to the CBP, is next immobilized on calmodulin–sepharose in the presence of calcium (c3). Calcium chelation (with EGTA) releases the recombinant protein, along with its interacting partners (c4).

**Electrospray ionization (ESI).** One of two common techniques (see MALDI) used in mass spectrometry to ionize large, non-volatile molecules (such as peptides). ESI charges analytes directly from the liquid phase and is easily coupled with reversed-phase high-pressure liquid chromatography.

**Matrix-assisted laser desorption/ionization (MALDI).** One of two common techniques (see ESI) used in mass spectrometry to ionize large, non-volatile molecules (such as peptides). MALDI ionizes molecules through laser excitation of a matrix in which the molecules are embedded.

**Graph-clustering algorithm** Protein–protein interactions are often represented in a graph format, in which individual proteins are depicted as circles (nodes) and the observed associations are depicted as lines (edges) that link the nodes.

**Socio-affinity index** Measure of the propensity of proteins to form partnerships. This index represents the log odds of the number of times particular proteins are observed together in a tandem affinity purification, as compared to what would be expected for their individual frequency of detection in the data set.

**Tobacco etch virus protease** Specific endopeptidase that recognizes the linear epitope EXXYXQG/S, in which X is any amino acid (although some preferences are observed).

the initial experiment recapitulated the known PP2A supramolecular architecture. Consistent with all previous biochemical data, TAP- $\alpha 4$  pulldowns contained only the C subunit of PP2A, without adaptor and regulatory proteins. On the other hand, the association of one regulatory subunit (for example, the B subunit) was mutually exclusive to that of another (B') (A.-C.G. and R.A., unpublished observations).

Gavin *et al.* also generated a high-density map of the interactions of the yeast orthologues of PP2A and recapitulated the mutually exclusive association of the B (Cdc55) and B' (Rts1) PP2A subunits with the PP2A C (Pph21 and Pph22) and A (Tpd3) proteins<sup>5</sup>. We also recently applied this strategy to the organization of the PP2A-related phosphatase PP4 and demonstrated that the PP4 C subunit assembles into several different alternative complexes<sup>37</sup>. High-density data can thus be used to identify mutually exclusive and/or cooperative protein assemblies.

**High-density data in high-throughput studies.** The use of high-density data to characterize alternative protein complexes can also be extended to large interaction networks. Two recent *S. cerevisiae* AP–MS studies<sup>2,4</sup> have made serious and promising attempts to dissect protein complexes from moderately dense interaction data. Although the two reports differ in the methodologies that were used to extract data and identify complexes, both make use of internally generated data sets that comprise roughly 2,000 successful TAP-tag purifications.

Krogan *et al.*<sup>2</sup> describe the identification of protein complexes on the basis of a graph-clustering algorithm that identifies highly connected modules in protein–protein interaction networks. The method was applied to the interaction data after the removal of common contaminants, and values for the clustering parameters were selected to optimize overlap with hand-curated protein complexes (available from Munich Information Center for Protein Sciences (MIPS)). This method identified 547 distinct heteromeric protein complexes, about half of which were novel<sup>3</sup>. However, although this method excels at identifying interacting partners, it does not necessarily separate complexes that contain shared subunits into different entities, but instead often groups them into a single larger complex. As such, it is difficult to directly compare the ‘complexes’ identified in this approach to *bona fide* biochemically stable protein assemblies. For example, although four separate yeast proteins (Tif4631, Tif4632, Caf20 and Eap1) share a common binding site for the translation-initiation factor eIF4E (Cdc33) (REF. 38) — such that only one of these proteins can interact with eIF4E at any given time — these four proteins were reported as part of a single ‘complex’.

Gavin *et al.* analysed the propensity of proteins to associate directly from raw data (without removal of contaminants), computing a composite index (the socio-affinity index) that represents the frequency with which two proteins were observed together as compared to their respective frequency of identification in the data set. In an attempt to capture the association of a given protein into multiple alternative assemblies, they

## Box 2 | False positives and false negatives in affinity purification and mass spectrometry data

**False positives**

**Dealing with background.** Different types of contaminants are present in typical affinity purified protein samples. Common contaminants include abundant proteins (for example, actin, tubulin and ribosomal proteins), proteins that bind to unfolded polypeptides (heat-shock proteins) and proteins that interact with affinity matrices (for example, STK38, arginine *N*-methyltransferase-5 and 52-kDa Ro protein in flag immunoprecipitates from human cells<sup>90</sup>). Careful experimental design and inclusion of proper controls allows users to discriminate between contaminants and *bona fide* interactors. It is important to express proteins at near physiological levels, as overexpression (for example, transient transfections using strong promoters) will probably result in the association of the bait protein with chaperones and can lead to improper intracellular localization. A dual-purification strategy (for example, tandem affinity purification (TAP)) helps to avoid isolation of proteins with affinity for the distal tag or the matrix. Including more controls — for example, control proteins expressed at similar levels with the same tag — can also help to identify true interactors. In this respect, high-throughput studies have an advantage over low-throughput studies: a contaminant protein is often isolated with many unrelated bait proteins, a result that is easily recognized through analysis of the high-throughput data. Combining high-throughput data sets as well as the use of noise filtering techniques promise to yield higher confidence interactors.

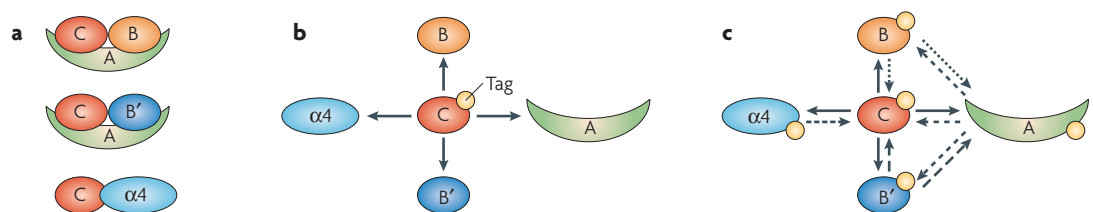
**False negatives**

**Where did the interactors go?** It is clear that high-throughput data sets (for example, those from the two recent TAP studies in yeast<sup>2,4</sup>) lack many previously documented protein–protein interactions. There are many reasons why a known interactor might not be identified in a given affinity purification and mass spectrometry (AP–MS) experiment. The proteins might not have interacted under the tested conditions (high-throughput data are typically generated under a single growth condition). The nature and location of the tag might have disrupted interactions, or the conditions of the AP might have been too harsh to preserve the interaction. However, the absence of detection is often due to a problem with the relative abundances of the proteins in the AP sample. This typically occurs if the bait is in large molar excess over its interacting partners (a problem that is exacerbated when the bait is overexpressed) or if the prey is in much lower abundance (that is, low stoichiometry) than other components of the sample. AP–MS thus tends to capture more stable, and fairly abundant, interactors in contrast to other approaches such as immunoprecipitation–western blot or yeast two-hybrid assays.

performed matrix clustering to generate an initial list of complexes. After subtracting a penalty from the original values, clustering was repeated with the assumption that tight associations would not be drastically affected by the penalty, whereas weaker associations would be gradually lost and could be replaced by interactions not initially detected. After varying clustering parameters (including the number of iterations and the penalty value) and comparing the resulting hits to a manually curated group of known complexes, the best parameters were chosen, yielding 491 protein complexes, of which about half were previously unknown<sup>4</sup>.

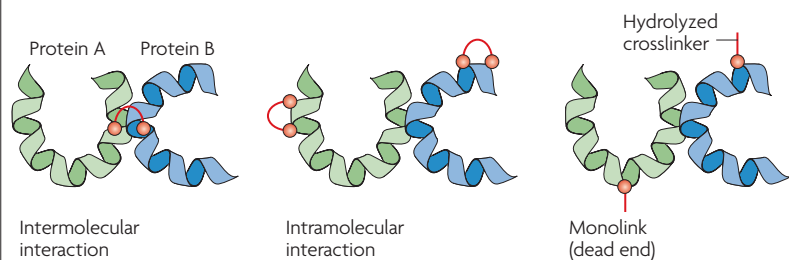
An important difference from the Krogan *et al.* approach is that Gavin *et al.* identified ‘complex isoforms’ by comparing these 491 protein complexes to similar complexes detected (albeit with slightly poorer

accuracy or coverage) from clustering with different parameters. Inclusion of ‘complex isoforms’ increased the coverage of components of known complexes, but also allowed the authors to partition proteins in complexes into two types: core components, which are present in most complexes, and attachments, which are present in only some. In addition, the authors also defined protein ‘modules’, in which two or more proteins are always found together, and these modules can be found with various other partners in multiple complexes. By doing so, Gavin *et al.* could in theory resolve the issue of a protein belonging to two complexes, as well as distinguish between essential and accessory components of a protein assembly. Although this approach was successful for the PP2A module, for which the density of information was sufficient, it was



**Figure 2 | High-density data acquisition to elucidate complex composition. a** | Three different biochemically defined complexes that contain the catalytic (C) subunit of the serine/threonine phosphatase PP2A (note that this is not an exhaustive list). **b** | Results from an affinity purification and mass spectrometry (AP–MS) analysis of the tandem AP (TAP)-tagged PP2A C subunit. These data represent an average of all PP2A-containing complexes; the composition of each of the different PP2A-containing complexes cannot be resolved. **c** | After tagging each component and repeating the purification procedure, the AP–MS data reveal several distinct PP2A-containing complexes, as depicted in **a** (indicated by the arrows).

## Box 3 | Crosslinkers in proteomics



Chemical crosslinkers possess at least two reactive moieties, directed towards one of several different types of functional groups (typically amines, sulphhydrydes or carboxyls). When bifunctional crosslinkers are incubated with a protein complex, several types of reactions can occur (see figure): the two reactive groups of the crosslinker can each interact with a different component of the complex (intermolecular interaction); the reactive groups can effect an intramolecular interaction with only one of the proteins; or only one of the reactive groups can interact with a protein, generating a monolink or a dead-end (the other end of the crosslinker is hydrolyzed). Identifying the type of link is crucial for obtaining structural information. Strategies for characterizing the architecture of protein complexes are generally aimed at identifying intermolecular interactions, although intramolecular interactions can provide information regarding how individual proteins fold in the complex, and monolinks can identify solvent-exposed amino acids.

**Gel filtration**

A chromatographic method (also called size-exclusion chromatography) that separates proteins or protein complexes on the basis of their size (or hydrodynamic volume).

**Density gradients**

Centrifugation-based separation technique that often uses sucrose or glycerol density gradients to separate components on the basis of their density. It is applicable to protein complexes, organelles and intact cells.

**Selective precipitation**

Common enrichment procedure (also known as salting out) that most often uses increasing concentrations of ammonium sulphate to alter the solubility of proteins and cause their precipitation. It is usually used in the early steps of a biochemical fractionation protocol.

**Ion-exchange chromatography**

A frequently used separation technique for peptides, proteins and protein complexes that exploits the charge properties of the molecules to be fractionated.

not successful for all complexes; for instance, both of the eIF4G isoforms (Tif4631 and Tif4632) were assigned to the same core complex with eIF4E, even though these associations are mutually exclusive.

It is to be expected that as these large data sets are combined and re-analysed (and as new data become available), the density of the data will increase and the sensitivity and accuracy of complex determination will improve. Collins *et al.* recently combined the Gavin and Krogan data sets and applied a new probabilistic method to filter out false positives in each of the data sets<sup>39</sup>. They found 9,074 high confidence interactions among 1,622 individual proteins, with an apparent error rate similar to the low-throughput data from MIPS. The Collins *et al.* scoring function has three components: direct evidence for an interaction following recovery of protein Y when protein X is a bait; indirect evidence of co-association, as was used by Gavin *et al.*; and evidence of non-association, when protein Y does not come down when protein X is used as bait. Combining the high-throughput data with this improved scoring function should help resolve protein complexes; however, this is easier said than done, and some mutually exclusive interactions (for example, those of Tif4631 and Tif4632 with eIF4E) still remain annotated as part of the same complex in the data set. In spite of these limitations, these studies have significantly improved our understanding of protein–protein interactions.

**AP–MS and biochemical fractionation**

As discussed above, the structure of multiprotein complexes can only be revealed indirectly through high-density AP–MS approaches. However, as described below, analysing the composition of an intact protein complex with defined biochemical properties can be used to directly reveal the composition of a given complex.

**Biochemical fractionation in protein-complex analysis.**

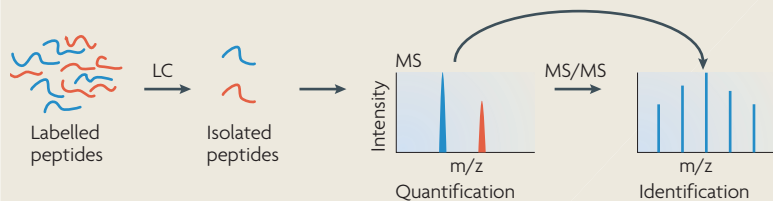
Size fractionation (via gel filtration or density gradient), selective precipitation and ion-exchange chromatography have been used widely for the separation and enrichment of protein complexes<sup>40</sup>. AP of at least one of the sample components, using for example an inhibitor or a ligand, has also frequently been included in biochemical purification schemes to significantly increase enrichment factors. Depending on the nature of the particular protein complex, a combination of these separation methods can yield pure preparations. The composition of the semi-purified or enriched complex can then be determined by MS<sup>41–44</sup>. For example, Moyer *et al.*<sup>43</sup> purified a high molecular mass complex that contains the *D. melanogaster* DNA-replication protein CDC45 by performing serial chromatography steps, including gel filtration (Superdex), ion exchange (DEAE sepharose, Mono S and Mono Q) and affinity chromatography (heparin and anti-CDC45). The presence of CDC45 in each fraction was monitored by immunoblotting. Silver staining revealed 10 CDC45 interactors, which were identified by MS and immunoblotting to be components of the minichromosome maintenance-2 (MCM2)–7 and GINS complexes.

Although fractionation approaches have been used successfully for the characterization of the composition of numerous biologically relevant protein assemblies, they are not generic and must be tailored to a particular complex of interest. This limitation prevents their application to genome-wide studies. However, combining one or more of such biochemical fractionation techniques with a generic AP protocol (such as an epitope tag) can provide a surrogate for a complete biochemical isolation of a protein complex. For example, a combination of TAP or flag purification with standard gel filtration or ion exchange has allowed for a better characterization of several nuclear complexes<sup>45–47</sup>; this type of approach should be scalable to high-throughput projects. Emerging techniques such as free-flow electrophoresis (FFE)<sup>48</sup> and blue native gel electrophoresis<sup>49–51</sup> also offer great promise for the purification of intact multiprotein assemblies. In particular, blue native gels have been increasingly used to isolate protein complexes (including complexes from membranes<sup>51–53</sup>), and could easily be incorporated into an AP–MS strategy.

**Guilt by association.** Another strategy for the analysis of large multiprotein assemblies (or organelles) is to monitor co-fractionation profiles using quantitative MS and then to compare the acquired profiles with those of known components of the protein complex or organelle of interest. This can be accomplished by monitoring the number and intensity of the peptide signals for each detected protein across adjacent fractions (for example, throughout a sucrose or glycerol gradient). Using such an approach, Mann and colleagues identified novel components of the centrosome<sup>54</sup> and have recently extended this approach to the description of components of multiple organelles from mouse liver<sup>55</sup>.

A related strategy for the identification of organellar components combines classical biochemical fractionation and affinity enrichment with stable-isotope-based

## Box 4 | Stable isotopes and quantitative proteomics



'Heavy' stable isotopes (such as  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^{18}\text{O}$ ) are frequently used in quantitative proteomics approaches. Because these molecules are chemically equivalent to the more abundant, naturally occurring, 'light' isotopes, they behave identically throughout biochemical fractionation procedures, yet differ in mass — a property that can be measured in the mass spectrometer. Stable isotopes can thus be used to determine relative amounts of proteins in different samples, either to monitor dynamic changes in protein levels or to filter out contaminants. (For methods allowing the incorporation of stable isotopes into peptides, see FIGS 3, 4.)

Isotope-labelled peptides are separated by reversed-phase liquid chromatography. Chemically equivalent peptides labelled with different isotopes co-fractionate. Differentially labelled, but chemically equivalent, peptides thus enter the mass spectrometer simultaneously and are subjected to an intact mass measurement: at this step, the instrument measures the mass/charge ( $m/z$ ) ratio and intensity of both peptide species. Multiple mass and intensity measurements are taken over time to generate two ion chromatograms, which represent the intensity of each peptide over time. The difference in the areas under the curves in these chromatograms is used to determine a relative abundance ratio between the different molecular species (see figure, the light peptide in blue is more abundant than the heavy peptide in red). Peptides are also subsequently fragmented in a tandem mass spectrometry (MS/MS) event to identify the quantified species (only the selection/fragmentation of one of the peptides is shown). The use of isotope-labelled peptide standards of known concentration allows for absolute quantification. LC, liquid chromatography.

quantitative proteomics (described in more detail below<sup>56</sup>). The use of stable isotopes allows for the measurement of enrichment ratios, which define the likelihood of a protein being in a particular complex or organelle. In addition to identifying new components of previously known complexes or organelles, groups of co-eluting proteins can identify novel organelles or protein complexes when no known component is available (for example, when the generated profiles are subjected to unsupervised clustering).

### Crosslinking of protein complexes

**Stabilization of interactions with crosslinkers.** A problem that is encountered during the isolation of intact native protein complexes from cells or tissue is that only protein-protein interactions that are resistant to the lysis and purification conditions will survive to be detected by MS. Several different strategies have thus been devised to freeze transient or labile protein interactions by using chemical crosslinking reagents (BOX 3). Crosslinkers possess at least two reactive groups that form covalent bonds with target molecules. These reactive groups are separated by a spacer arm of a defined length (usually in the range of 5–15 Å) that determines the maximal distance between two molecules. This confers some degree of specificity to the crosslinking process: molecules in close proximity are more likely to be crosslinked than distant species. However, protein-protein crosslinking techniques present multiple experimental and analytical

challenges. The choice of crosslinker is crucial, as crosslinkers vary in cell-permeability, reactivity and spacer arm length. Crosslinking reaction conditions must also be closely monitored, such that *bona fide* protein-protein interactions are stabilized and undesired crosslinks (to contaminating proteins) are minimized.

Although many chemical crosslinkers can be used to stabilize complexes in theory, only a few have been successfully used for *in vivo* crosslinking followed by MS analysis. Formaldehyde and di-thiobis-succinimidyl-propionate (DSP; an amine-reactive, homobifunctional, thiol-cleavable and membrane-permeable crosslinker) have been used most often to identify novel interacting partners<sup>57–60</sup>. Crosslinkers are also particularly attractive for revealing interactions that involve membrane proteins (or microsomes), as the detergent concentrations used to solubilize the membranes and extract the proteins typically also disrupt protein-protein interactions<sup>61</sup>. Mild formaldehyde crosslinking has also been performed in animals and has allowed for the stringent immunopurification of an intact  $\gamma$ -secretase complex as well as the identification of protein interactors for the cellular prion protein PrP<sup>C</sup> (REF. 62). A combination of techniques integrating a modified TAP strategy, *in vivo* crosslinking with formaldehyde and an isotope-labelling strategy (BOX 4; FIG. 3) was used to identify novel proteasome interactors<sup>63,64</sup>. Improved variations of combined crosslinking, quantitative proteomics and AP-MS approaches will probably continue to be developed.

**Crosslinkers and structural studies.** Several new strategies have also been developed to measure distance constraints through intra- and intermolecular crosslinking data (reviewed in REF. 65). As a low-resolution approach, crosslinking-MS is rapid, requires little material and allows structural analysis of complexes that are isolated under native conditions. A main challenge to the application of crosslinkers to structural determination is the identification of those pairs of peptides that are specifically crosslinked. This issue has recently been addressed by ingenious experimental designs and novel data analysis solutions. For example, several groups have devised approaches that use isotope-labelled crosslinkers<sup>66–68</sup>. The mass of a pair of peptides crosslinked with a heavy-isotope-labelled crosslinking reagent is larger than that of the same peptides crosslinked with the corresponding light-isotope-labelled compound. Peptides modified by light- and heavy-isotope-coded crosslinkers appear as doublets in mass spectra, whereas unmodified peptides appear as singlets; this difference facilitates the selection of the crosslinked peptides for further analysis.

Seebacher *et al.* introduced an additional step to distinguish monolinks from crosslinks<sup>68</sup>. The protein sample is prepared in a mixture of heavy and light water. The non-reacted end of a monolink (BOX 3) is hydrolyzed and incorporates  $^{18}\text{O}/^{16}\text{O}$  from the solvent, resulting in a splitting of the peaks<sup>68</sup>, and differentiating them from intra- or intermolecularly crosslinked peptides. This type of approach can be combined with data-analysis strategies to identify the crosslinked pairs and can establish distance constraints for structural modelling.

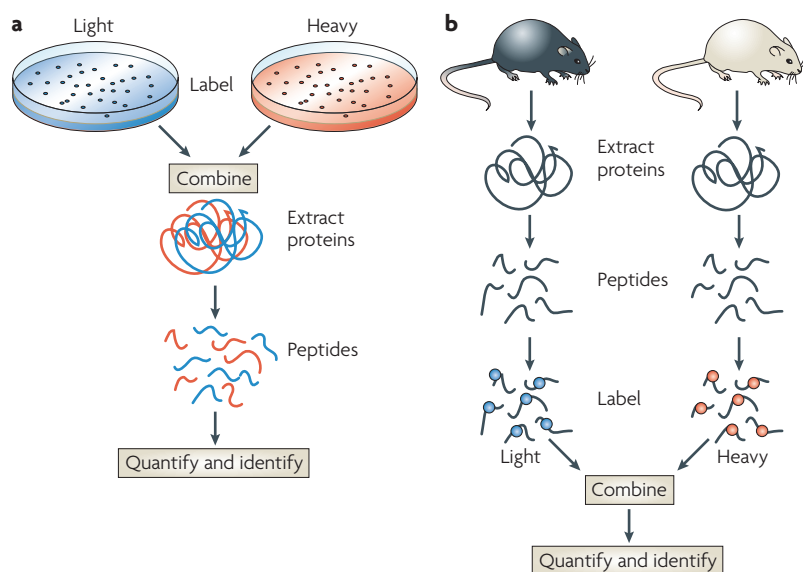
**Free flow electrophoresis (FFE).** Separation procedure that fractionates the sample (ranging in size from small molecules, including peptides, to objects of cellular dimension) on the basis of the isoelectric point (or electrophoretic mobility). FFE operates continuously in the absence of a solid support.

### Blue native gel electrophoresis

A native gel-based approach that efficiently separates protein complexes. It can be used alone or in combination with a denaturing second dimension.

### Spacer arm length

Spacer arm length refers to the distance between two reactive groups in a bifunctional crosslinker. The spacer arm length defines the maximal distance between the two crosslinked moieties.



**Figure 3 | Incorporation of stable isotopes into proteins. a** | Metabolic labelling. In the stable-isotope labelling with amino acids in cell culture (SILAC) technique, one population of cells is maintained for several generations in growth media in which one or more naturally occurring 'light' amino acids have been replaced by their isotopically 'heavy' (usually containing  $^{13}\text{C}$  or  $^{15}\text{N}$ ) counterparts. Heavy- and light-labelled samples are combined, and the proteins of interest are isolated using standard biochemical procedures. Protein mixtures are proteolyzed and subjected to mass spectrometry (MS). **b** | Chemical labelling. Chemical labelling is used after isolation, and often following proteolysis, of proteins of interest. Chemical labelling can therefore be used when SILAC-type labelling is difficult or impossible (for example, in human tissue biopsies). Chemically labelled peptides can be combined and then analysed by MS (often following an optional purification step).

A new generation of crosslinkers, referred to as protein-interaction reporters (PIR), contain two labile bonds in the spacer domain. These bonds are broken by low-energy MS<sup>2</sup> to yield a reporter ion, which facilitates the identification of spectra from crosslinked peptides; spectra that contain non-crosslinked peptides do not contain the reporter ion. Another advantage of the PIR approach is that the two crosslinked peptides are released following the cleavage of the labile crosslinker and can be easily identified using standard database-search algorithms<sup>69</sup>. Future use of such reagents holds great promise for obtaining structural information of proteins in a complex.

### Complex stoichiometry

Another issue which has not yet been addressed in high-throughput studies concerns the stoichiometry of complex components — this is important information for understanding the structural organization of a protein assembly. In recent years, mass spectrometers have been modified for the analysis of protein assemblies in the MDa range, and conditions have been developed to maintain intact protein complexes in the gas phase<sup>70</sup>. Measurement of the intact mass of a protein complex, as well as the exact masses of each of its components, can thus be used to determine the stoichiometry of each component. This strategy was used to decipher the heptameric nature of the stalk complex dissociated from

bacterial ribosomes<sup>71</sup> as well as interactions in complicated multiprotein structures, such as the proteasome 19S lid<sup>72</sup> and the yeast exosome<sup>73</sup>. Importantly, such an approach is compatible with AP by TAP tagging<sup>73</sup>. Because this approach is compatible with many purification strategies and because the parameters of operation of the mass spectrometer can be manipulated to partially disrupt the protein complex, this technique also has the potential to shed light on subcomplexes or weakly associated binding partners (reviewed in REF. 74).

Another promising strategy to determine protein stoichiometry in a complex is to combine complex isolation with isotope-based absolute quantitative proteomics (BOX 4). If all of the components of a complex are known, synthetic tryptic peptides can be generated to monitor the abundance of each of the proteins in the complex. These peptides can be synthesized with heavy isotopes and then mixed with an unlabelled sample (as in the AQUA approach<sup>75</sup> or the QCAT strategy<sup>76</sup>) or labelled in parallel to the samples (for example by reaction with isobaric tags for relative and absolute quantification (iTRAQ)<sup>77</sup> or other amino-reactive reagents). This approach was successfully used to confirm the 1:1:1 stoichiometry of the human spliceosomal U1 small nuclear ribonucleoprotein<sup>78</sup>. It should be noted that this method is dependent on the isolation of a single homogeneous complex, which can be difficult to achieve using common isolation methods.

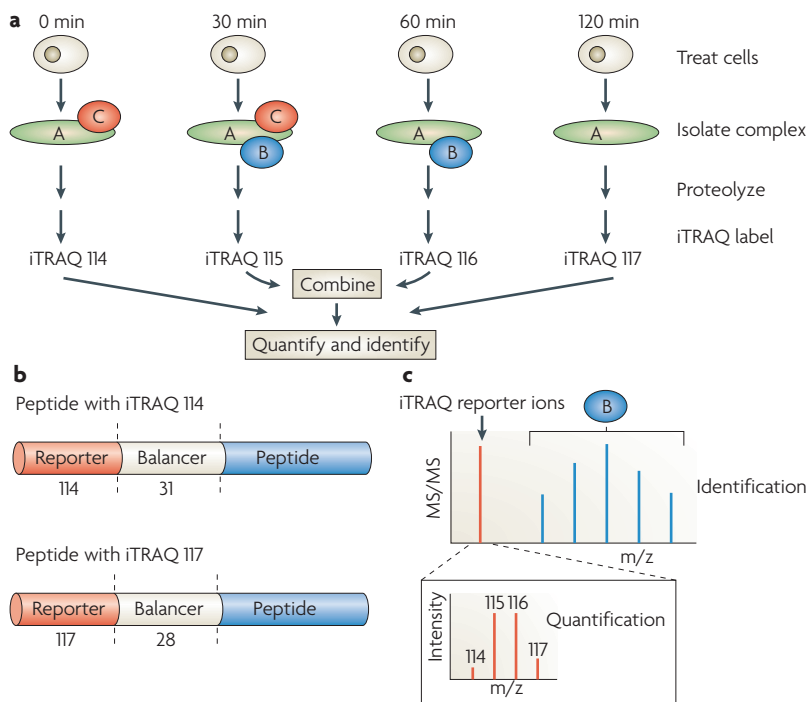
### Complex dynamics

Most of the high-throughput AP-MS data generated so far represent a static view of protein complexes. Owing to the dynamically changing composition of many complexes — particularly those involved in cell signalling — it has been crucial to develop strategies to capture regulated interactions. Quantitative proteomics approaches are ideally suited for this task. One effective way to generate quantitative information is to incorporate stable isotopes into peptides. This is typically accomplished through metabolic labelling, in which isotopically 'heavy' compounds replace the natural 'light' isotopes in growth medium or through the addition of an isotopic label after lysis (and often after purification) using chemical or enzymatic reactions<sup>79,80</sup> (FIG. 3).

**Metabolic labelling with SILAC.** In the stable-isotope labelling with amino acids in cell culture (SILAC) procedure<sup>81</sup>, cells are grown in culture for several generations in the presence of an isotopically heavy amino acid, thereby replacing essentially all of the naturally occurring light amino acid. Peptides derived from labelled cells behave identically to the corresponding peptide from unlabelled cells throughout any biochemical enrichment or purification steps. However, the difference in mass between the light and heavy peptides can be measured in the survey scan of the mass spectrometer. The relative intensities of the MS peaks are proportional to the abundance of the peptides in the two samples. SILAC has been successfully used in several recent studies (see REFS 17, 18, 81–83), including for the profiling of the dynamic association of proteins isolated via an

**Isobaric tags for relative and absolute quantification**  
A type of chemical labelling for quantitative proteomics in which isobaric tags are added to the N terminus of every peptide (and to the lysine  $\epsilon$ -amine).

**Stable-isotope labelling with amino acids in cell culture**  
A type of metabolic labelling for quantitative proteomics in which heavy isotopic versions of one or more amino acids are used in the culture medium to replace their normal (light) isotopic counterparts.



**Figure 4 | Isobaric tags to elucidate complex formation dynamics.** **a** | Desired treatment of cells is followed by isolation of protein complexes and proteolysis. Isobaric tags (iTRAQ) are chemically added to the N terminus of every peptide (as well as to lysine  $\epsilon$ -amine groups). Samples from multiple treatment time points are combined and subjected to analysis. **b** | A peptide labelled with the iTRAQ 114 and iTRAQ 117 reagents. iTRAQ is isobaric, such that addition of the 114-Da or 117-Da mass tags alter the mass of a given peptide by the same amount. To maintain a constant mass, the reporter moiety (for example, of mass 114) is separated from the peptide by a balancer group. The reporter and balancer groups fragment in the collision cell of the mass spectrometer during the tandem mass spectrometry (MS/MS) event, and the intensity of the reporter ions is monitored. **c** | Analysis of an iTRAQ experiment. MS/MS analysis of a labelled peptide generates a fragmentation spectrum that yields the sequence of the peptide. The iTRAQ reagent is fragmented in the same step and reporter ions are quantified by magnifying the low mass range (114–117) area. In the example shown, protein B associates with protein A (the bait) after 30 and 60 minutes of stimulation, but not after 120 minutes of treatment. *m/z*, mass/charge ratio.

**Isotope-coded affinity tags (ICAT).** A chemical labelling reagent used for quantitative proteomics. The current version of ICAT consists of an acid-cleavable biotin-labelled tag that reacts with cysteine groups on proteins. Isotopically heavy ( $^{13}\text{C}_6$ ) and light ( $^{12}\text{C}_6$ ) variants are commercially available.

**DNA-repair syndrome trichothiodystrophy group A** Trichothiodystrophy is a human syndrome characterized by brittle hair and nails, ichthyotic skin, and developmental and mental retardation. Photosensitivity is observed in roughly half of the patients and has been linked to defects in nucleotide-excision repair.

anti-phosphotyrosine resin following stimulation with growth factors, such as epidermal growth factor (EGF) or platelet-derived growth factor (PDGF), in mesenchymal stem cells<sup>17</sup>.

**Chemical labelling with ICAT reagents.** Quantitative approaches based on isotopic labelling of proteins (or peptides) after lysis using chemical reactions can also be used for detecting changes in complex composition. Chemical labelling approaches are complementary to SILAC and are particularly useful for measuring dynamic changes in complexes isolated from tissues or organisms that cannot be metabolically labelled. Isotope-coded affinity tags<sup>84</sup> (ICAT) are typical of this type of approach and are comprised of an affinity handle (often biotin), an isotopic linker (often cleavable) and a protein reactive group (for example, a thiol- or amine-reactive group). ICAT reagents have been used in many quantitative proteomics studies (reviewed in REFS 79,85). For example, the ICAT approach uncovered a novel tenth subunit

of the TFIIH transcription complex; mutation of this subunit is responsible for the DNA-repair syndrome trichothiodystrophy group A<sup>20,86</sup>. ICAT has also been applied to the study of dynamically regulated associations, for example to determine changes in transcription-factor complexes during erythroid-cell differentiation<sup>22</sup>. In addition to the original ICAT reagents, various different types of chemical isotope-labelling reagents, targeting various different reactive groups, have now been developed.

**Multiplex isobaric tags.** A newer generation of labelling reagent, iTRAQ<sup>77</sup>, is particularly well suited for resolving complex dynamics because it allows multiplexed quantitative analysis of four different samples (FIG. 4). iTRAQ reagents are amine reactive and are therefore covalently linked to the N terminus of every peptide (as well as to the  $\epsilon$ -amine of lysine residues). In contrast to SILAC and ICAT labelling, in which the heavy and light labels alter the parent peptide masses, iTRAQ reagents are isobaric and are therefore not quantified in the survey scan. Instead, iTRAQ reagents fragment during the tandem MS (MS/MS) step, yielding a single low-mass reporter ion (at 114, 115, 116 or 117 Da for the four-plexed reagent set). Intensities of each of the reporter ions are used to calculate the relative abundance of each of the peptides in the samples. The remaining peaks in the MS/MS spectrum are used for sequence identification. So, a single MS/MS spectrum can yield both quantitative information and can identify a peptide. Temporal profiles can be reconstructed from the purification of complexes at different time points following a specific treatment and the labelling of the derived peptides with a single iTRAQ reagent. iTRAQ technology has been used to analyse dynamically regulated tyrosine phosphorylation<sup>87</sup> and to characterize the carbon-source dependency of the Snf1 kinase complex in *Candida albicans*<sup>88</sup>.

**Which method to choose?** We have highlighted only a subset of the quantitative proteomics techniques currently available; each of the techniques has strengths and limitations, and the choice of the proper technique depends on the question being asked. The first choice to be made concerns the use of metabolic versus chemical labelling. SILAC can be used only with cells grown in culture for several generations, but has the advantage of introducing the isotopic label before purification, thereby reducing errors due to sample handling. Chemical labelling (ICAT and iTRAQ) is not as limited in applicability, but the efficiency of the labelling reaction can be problematic in some cases.

A second decision point regards the use of standard isotopic versus isobaric labelling reagents: combining SILAC or ICAT samples prior to MS analysis results in an increase in sample complexity (proportional to the number of isotopic variants used), a problem that does not occur in iTRAQ experiments. However, whereas iTRAQ quantification is extremely powerful for direct comparison of up to four samples in a single experiment, comparison across different experiments requires that the peptide of interest be fragmented in each data set. The same is not the case in SILAC- or ICAT-based

experiments, in which quantification at the MS level can be obtained even if the peptide was not selected for fragmentation.

Although isotopic labelling approaches have clearly been at the forefront of the quantitative proteomics field, label-free approaches are also being developed, and these might change the way researchers perform quantitative experiments. These represent a promising alternative in situations in which metabolic labelling is not feasible, and changes in protein-complex components must be measured with high sensitivity. For example, Rinner *et al.*<sup>89</sup> recently applied a label-free strategy on the basis of the computational alignment of MS spectra across different samples (MasterMap) to capture insulin-mediated changes in the formation of human FOXO3A complexes.

### Conclusions

We have presented the general principles of the AP-MS approach and have highlighted some recent successes in deciphering the composition of protein complexes. Although AP-MS was traditionally considered a low-throughput method, ground-breaking studies in *S. cerevisiae* have demonstrated the power of this approach in high-throughput interactome studies. One crucial issue regarding the interpretation of AP-MS data is that a single protein can be present in multiple

distinct complexes. Purification of a single protein can thus result in a mixture of different unique protein-protein assemblies. We have discussed methods to characterize individual protein complexes by AP-MS either indirectly, through high-density data acquisition, or directly, through the coupling of AP-MS with biochemical fractionation.

AP-MS can also be combined with a number of different approaches to dramatically increase information content. For example, the use of crosslinkers in AP-MS design was shown to have two advantages: stabilizing weak or transient interactions and providing important clues regarding the architecture of the protein complex. Structural studies of complexes will also be facilitated by new strategies to determine the stoichiometry of each of the components of a complex, either through intact mass measurements (of the complex and its constituents) or through absolute quantitative proteomics. Last, the modern generation of quantitative proteomics techniques truly opens the door for studies of protein-complex dynamics. This is an exciting area of research — all high-throughput maps that have been generated so far have been static; however, by using quantitative proteomics methods, dynamic protein-protein interaction maps will be generated in the future, allowing us to discover how proteins are reorganized following intracellular signalling events.

- Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
- Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).  
**One of the two most comprehensive AP-MS studies in yeast to date (see REF. 4) using TAP to generate a high-confidence interaction network and define protein complexes.**
- Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
- Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).  
**One of the two most comprehensive AP-MS studies in yeast (see REF. 2). It introduced the notion of the socio-affinity index to define the propensity of proteins to associate into specific complexes.**
- Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Formstecher, E. *et al.* Protein interaction mapping: a *Drosophila* case study. *Genome Res.* **15**, 376–384 (2005).
- Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
- Stanyon, C. A. *et al.* A *Drosophila* protein-interaction map centered on cell-cycle regulators. *Genome Biol.* **5**, R96 (2004).
- Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
- Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
- Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
- Gandhi, T. K. *et al.* Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nature Genet.* **38**, 285–293 (2006).
- Fields, S. & Sternglanz, R. The two-hybrid system: an assay for protein-protein interactions. *Trends Genet.* **10**, 286–292 (1994).
- von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399–403 (2002).
- Parrish, J. R., Gulyas, K. D. & Finley, R. L. Jr. Yeast two-hybrid contributions to interactome mapping. *Curr. Opin. Biotechnol.* **17**, 387–393 (2006).
- Kratzmarova, I., Blagoev, B., Haack-Sorensen, M., Kassem, M. & Mann, M. Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* **308**, 1472–1477 (2005).
- Blagoev, B. *et al.* A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nature Biotechnol.* **21**, 315–318 (2003).
- Himeda, C. L. *et al.* Quantitative proteomic identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer. *Mol. Cell Biol.* **24**, 2132–2143 (2004).
- Ranish, J. A. *et al.* Identification of TFB5, a new component of general transcription and DNA repair factor IIIH. *Nature Genet.* **36**, 707–713 (2004).
- Ranish, J. A. *et al.* The study of macromolecular complexes by quantitative proteomics. *Nature Genet.* **33**, 349–355 (2003).
- Brand, M. *et al.* Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nature Struct. Mol. Biol.* **11**, 73–80 (2004).
- Terpe, K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60**, 523–533 (2003).
- Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212–217 (2006).
- Nesvizhskii, A. I. Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol. Biol.* **367**, 87–120 (2006).
- Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
- Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nature Rev. Mol. Cell Biol.* **5**, 699–711 (2004).
- Domon, B. & Aebersold, R. Challenges and opportunities in proteomic data analysis. *Mol. Cell Proteomics* **5**, 1921–1926 (2006).
- Mikesh, L. M. *et al.* The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta* **1764**, 1811–1822 (2006).
- Nesvizhskii, A. I. *et al.* Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides. *Mol. Cell Proteomics* **5**, 652–670 (2006).
- Pedrioli, P. G. *et al.* Automated identification of SUMOylation sites using mass spectrometry and SUMOn pattern recognition software. *Nature Methods* **3**, 533–539 (2006).
- Janssens, V., Goris, J. & Van Hoof, C. PP2A: the expected tumor suppressor. *Curr. Opin. Genet. Dev.* **15**, 34–41 (2005).
- Goldberg, Y. Protein phosphatase 2A: who shall regulate the regulator? *Biochem. Pharmacol.* **57**, 321–328 (1999).
- Kong, M. *et al.* The PP2A-associated protein  $\alpha 4$  is an essential inhibitor of apoptosis. *Science* **306**, 695–698 (2004).
- Chen, J., Peterson, R. T. & Schreiber, S. L.  $\alpha 4$  associates with protein phosphatases 2A, 4, and 6. *Biochem. Biophys. Res. Commun.* **247**, 827–832 (1998).
- Inui, S. *et al.* Ig receptor binding protein 1 ( $\alpha 4$ ) is associated with a rapamycin-sensitive signal transduction in lymphocytes through direct binding to the catalytic subunit of protein phosphatase 2A. *Blood* **92**, 539–546 (1998).
- Gingras, A. C. *et al.* A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol. Cell Proteomics* **4**, 1725–1740 (2005).
- Cosentino, G. P. *et al.* Eap1p, a novel eukaryotic translation initiation factor 4E-associated protein in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **20**, 4604–4613 (2000).
- Collins, S. R. *et al.* Towards a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell Proteomics* **6**, 439–450 (2007).
- Abelson, J. N. & Simon, M. I. (eds) *Guide to protein purification* (Academic Press, 1990).
- Weiner, O. D. *et al.* Hem-1 complexes are essential for Rac activation, actin polymerization, and myosin regulation during neutrophil chemotaxis. *PLoS Biol.* **4**, e38 (2006).

42. Panigrahi, A. K. *et al.* Association of two novel proteins, TbMP52 and TbMP48, with the *Trypanosoma brucei* RNA editing complex. *Mol. Cell Biol.* **21**, 380–389 (2001).
43. Moyer, S. E., Lewis, P. W. & Botchan, M. R. Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl Acad. Sci. USA* **103**, 10236–10241 (2006).
- A nice example of classical biochemical fractionation combined with AP–MS. CDC45 interactors were identified as components of the Mcm2–7 and GINS complexes.**
44. Poot, R. A. *et al.* HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. *EMBO J.* **19**, 3377–3387 (2000).
45. Mueller, C. L. & Jaehning, J. A. Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol. Cell Biol.* **22**, 1971–1980 (2002).
46. Lindstrom, D. L. *et al.* Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell Biol.* **23**, 1368–1378 (2003).
47. Ducut Sigala, J. L. *et al.* Activation of transcription factor NF- $\kappa$ B requires ELKS, an I $\kappa$ B kinase regulatory subunit. *Science* **304**, 1965–1967 (2004).
48. Weber, G. & Boeck, P. Recent developments in preparative free flow isoelectric focusing. *Electrophoresis* **19**, 1649–1653 (1998).
49. Lasserre, J. P. *et al.* A complexomic study of *Escherichia coli* using two-dimensional blue native/SDS polyacrylamide gel electrophoresis. *Electrophoresis* **27**, 3306–3321 (2006).
50. Schagger, H. & von Jagow, G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem.* **199**, 223–231 (1991).
51. Camacho-Carvajal, M. M., Wollscheid, B., Aebersold, R., Steimle, V. & Schamel, W. W. Two-dimensional blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates: a proteomics approach. *Mol. Cell. Proteomics* **3**, 176–182 (2004).
- Example of the use of blue native electrophoresis combined with MS for the study of multiprotein complexes.**
52. Nijtmans, L. G., Henderson, N. S. & Holt, I. J. Blue native electrophoresis to study mitochondrial and other protein complexes. *Methods* **26**, 327–334 (2002).
53. Fandino, A. S. *et al.* LC-nanospray–MS/MS analysis of hydrophobic proteins from membrane protein complexes isolated by blue-native electrophoresis. *J. Mass Spectrom.* **40**, 1223–1231 (2005).
54. Andersen, J. S. *et al.* Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570–574 (2003).
- Description of the protein correlation profiling strategy and its use in the identification of novel components of the centrosome. This approach can also be applied to other organelles.**
55. Foster, L. J. *et al.* A mammalian organelle map by protein correlation profiling. *Cell* **125**, 187–199 (2006).
56. Marelli, M. *et al.* Quantitative mass spectrometry reveals a role for the GTPase Rho1p in actin organization on the peroxisome membrane. *J. Cell Biol.* **167**, 1099–1112 (2004).
57. Kim, D. H. *et al.* mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell* **110**, 163–175 (2002).
58. Vasilescu, J., Guo, X. & Kast, J. Identification of protein–protein interactions using *in vivo* cross-linking and mass spectrometry. *Proteomics* **4**, 3845–3854 (2004).
- Example of the use of crosslinking (with formaldehyde) coupled to AP–MS to identify protein interactors for an activated form of myc-tagged Ras.**
59. Meunier, L., Usherwood, Y. K., Chung, K. T. & Henderson, L. M. A subset of chaperones and folding enzymes form multiprotein complexes in endoplasmic reticulum to bind nascent proteins. *Mol. Biol. Cell* **13**, 4456–4469 (2002).
60. Kim, S., Wong, P. & Coulombe, P. A. A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. *Nature* **441**, 362–365 (2006).
61. Rashid, K. A., Hevi, S., Chen, Y., Le Caherec, F. & Chuck, S. L. A proteomic approach identifies proteins in hepatocytes that bind nascent apolipoprotein B. *J. Biol. Chem.* **277**, 22010–22017 (2002).
62. Schmitt-Ulms, G. *et al.* Time-controlled transcardiac perfusion cross-linking for the study of protein interactions in complex tissues. *Nature Biotechnol.* **22**, 724–731 (2004).
63. Tagwerker, C. *et al.* A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with *in vivo* cross-linking. *Mol. Cell. Proteomics* **5**, 737–748 (2006).
64. Guerrero, C., Tagwerker, C., Kaiser, P. & Huang, L. An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified *in vivo* cross-linked protein complexes (QTAX) to decipher the 26S proteasome-interacting network. *Mol. Cell. Proteomics* **5**, 366–378 (2006).
- A nice example of a combination of approaches, including quantitative proteomics, AP and crosslinking, to define protein–protein interactions.**
65. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev.* **25**, 663–682 (2006).
66. Schmidt, A., Kalkhof, S., Ihling, C., Cooper, D. M. & Sinz, A. Mapping protein interfaces by chemical cross-linking and Fourier transform ion cyclotron resonance mass spectrometry: application to a calmodulin/adenylyl cyclase 8 peptide complex. *Eur. J. Mass Spectrom.* **11**, 525–534 (2005).
67. Muller, D. R. *et al.* Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal Chem.* **73**, 1927–1934 (2001).
68. Seebacher, J. *et al.* Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. *J. Proteome Res.* **5**, 2270–2282 (2006).
69. Tang, X., Munske, G. R., Siems, W. F. & Bruce, J. E. Mass spectrometry identifiable cross-linking strategy for studying protein–protein interactions. *Anal Chem.* **77**, 311–318 (2005).
70. Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem.* **74**, 1402–1407 (2002).
71. Ilag, L. L. *et al.* Heptameric (L12)6/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria. *Proc. Natl Acad. Sci. USA* **102**, 8192–8197 (2005).
72. Sharon, M., Taverner, T., Ambroggio, X. I., Deshaies, R. J. & Robinson, C. V. Structural organization of the 19S proteasome lid: insights from MS of intact complexes. *PLoS Biol.* **4**, e267 (2006).
73. Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B. & Robinson, C. V. Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep.* **7**, 605–610 (2006).
- A nice example of the use of intact-complex MS in the characterization of component stoichiometry, subunit interactions and overall assembly of protein complexes. TAP-purified material was used to elucidate the structure of the exosome.**
74. Benesch, J. L. & Robinson, C. V. Mass spectrometry of macromolecular assemblies: preservation and dissociation. *Curr. Opin. Struct. Biol.* **16**, 245–251 (2006).
75. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl Acad. Sci. USA* **100**, 6940–6945 (2003).
76. Beynon, R. J., Doherty, M. K., Pratt, J. M. & Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature Methods* **2**, 587–589 (2005).
77. Ross, P. L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004).
78. Hochleitner, E. O. *et al.* Protein stoichiometry of a multiprotein complex, the human spliceosomal U1 small nuclear ribonucleoprotein: absolute quantification using isotope-coded tags and mass spectrometry. *J. Biol. Chem.* **280**, 2536–2542 (2005).
79. Gingras, A. C., Aebersold, R. & Raught, B. Advances in protein complex analysis using mass spectrometry. *J. Physiol.* **563**, 11–21 (2005).
80. Ong, S. E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature Chem. Biol.* **1**, 252–262 (2005).
81. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
- First description of SILAC, a metabolic labelling approach that is now widely used for quantitative proteomics.**
82. Foster, L. J. *et al.* Insulin-dependent interactions of proteins with GLUT4 revealed through stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* **5**, 64–75 (2006).
83. Bose, R. *et al.* Phosphoproteomic analysis of Her2/neu signaling and inhibition. *Proc. Natl Acad. Sci. USA* **103**, 9773–9778 (2006).
84. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999 (1999).
- First description of ICAT, a chemical labelling approach for quantitative proteomics with isotopes. Different labelling reagents that target different protein reactive groups have subsequently been developed.**
85. Flory, M. R., Griffin, T. J., Martin, D. & Aebersold, R. Advances in quantitative proteomics using stable isotope tags. *Trends Biotechnol.* **20**, S23–S29 (2002).
86. Giglia-Mari, G. *et al.* A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nature Genet.* **36**, 714–719 (2004).
87. Zhang, Y. *et al.* Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* **4**, 1240–1250 (2005).
88. Corvey, C. *et al.* Carbon source-dependent assembly of the Snf1p kinase complex in *Candida albicans*. *J. Biol. Chem.* **280**, 25323–25330 (2005).
89. Rinner, O. *et al.* An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nature Biotechnol.* **25**, 345–352 (2007).
90. Chen, G. I. & Gingras, A. C. Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. *Methods* **42**, 298–305 (2007).

**Acknowledgements**

Funding was provided in part by the Terry Fox Foundation to A.-C.G., the Canada Institutes of Health Research to B.R. and the National Heart, Lung, and Blood Institute, National Institute of Health contract N01-HV-28179 to R.A. A.-C.G. and B.R. are recipients of Canada Research Chairs, Canadian Foundation for Innovation and Ontario Research Fund grants.

**Competing interests statement**

The authors declare no competing financial interests.

**FURTHER INFORMATION**

Anne-Claude Gingras's homepage: <http://www.mshri.on.ca/?page=Dr%20Anne%20Claude%20Gingras>  
 Brian Raught's homepage: <http://www.uhnresearch.ca/researchers/profile.php?lookup=11069>  
 Matthias Gstaiger's homepage: <http://www.imsb.ethz.ch/researchgroup/gstaiger>  
 Ruedi Aebersold's homepage: [http://www.systemsbio.org/Scientists\\_and\\_Research/Faculty\\_Groups/Aebersold\\_Group](http://www.systemsbio.org/Scientists_and_Research/Faculty_Groups/Aebersold_Group)  
<http://www.imsb.ethz.ch/researchgroup/rudolf>  
**Access to this links box is available online.**