

# Co-immunoprecipitations revisited: an update on experimental concepts and their implementation for sensitive interactome investigations of endogenous proteins

Kelly Markham · Yu Bai · Gerold Schmitt-Ulms

Received: 8 March 2007 / Revised: 14 May 2007 / Accepted: 22 May 2007 / Published online: 22 June 2007  
© Springer-Verlag 2007

**Abstract** The study of protein–protein interactions involving endogenous proteins frequently relies on the immunoaffinity capture of a protein of interest followed by mass spectrometry-based identification of co-purifying interactors. A notorious problem with this approach is the difficulty of distinguishing physiological interactors from unspecific binders. Additional challenges pose the need to employ a strategy that is compatible with downstream mass spectrometry and minimizes sample losses during handling steps. Finally, the complexity of data sets demands solutions for data filtering. Here we present an update on co-immunoprecipitation procedures for sensitive interactome mapping applications. We define the relevant terminology, review methodological advances that reduce sample losses, and discuss experimental strategies that facilitate recognition of candidate interactors through a combination of informative controls and data filtering. Finally, we provide starting points for initial validation experiments and propose conventions for manuscripts which report on co-immunoprecipitation work.

**Keywords** Co-immunoprecipitations · Mass spectrometry · Isotopic labeling · Antibodies · In vivo crosslinking

## Abbreviations

BSA	Bovine serum albumin
co-IP	Co-immunoprecipitation
FRET	Fluorescence resonance energy transfer
GO	Gene ontology
GST	Glutathione- <i>S</i> -transferase
HA	Hemagglutinin
HPLC	High-performance liquid chromatography
MS–MS	Tandem mass spectrometry
PAGE	Polyacrylamide gel electrophoresis
SDS	Sodium dodecylsulfate
TFA	Trifluoroacetic acid
THS	Two-hybrid system

## Introduction

Proteins do not act in isolation but engage in complex and dynamic interactions with other proteins to fulfill their diverse cellular roles [1–3]. In recognition of this, researchers in pursuit of novel therapeutic targets and diagnostic markers turn to strategies that provide insights into the molecular environments of established disease targets.

Traditionally, protein interactions have been investigated by exploiting the unique physicochemical properties of individual protein complexes. In such studies, a protein complex is subjected to a biochemical purification scheme that often relied on cell fractionation, salt precipitation, and/or conventional chromatography steps [4, 5]. If performed under “mild” conditions, this approach may lead to the parallel recovery of all proteins that constitute a protein

**Electronic supplementary material** The online version of this article (doi:10.1007/s00216-007-1385-x) contains supplementary material, which is available to authorized users.

K. Markham · Y. Bai · G. Schmitt-Ulms (✉)  
Centre for Research in Neurodegenerative Diseases,  
University of Toronto,  
Tanz Neuroscience Building,  
6 Queen’s Park Crescent West,  
Toronto, ON M5S 3H2, Canada  
e-mail: g.schmittulms@utoronto.ca

G. Schmitt-Ulms  
Department of Laboratory Medicine and Pathobiology,  
University of Toronto,  
Toronto, ON M5S 3H2, Canada

complex. Because this strategy capitalizes on the unique properties of a given protein complex, however, no two purification schemes are identical, and the constant challenge remains the empirical adjustment of procedures to new protein complexes. Affinity purification steps that rely on the immobilization of a protein of interest, hereafter referred to as the bait protein, using bait-specific antibodies of high-affinity biological ligands provide an attractive solution to this problem [6–9]. In early studies of this kind, the main objective was to replace the last few steps of a complex purification scheme with a high-affinity purification step. Bait protein interactors were then either resolved by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to Western blotting membranes or were digested and their peptides separated by high-performance liquid chromatography (HPLC). Subsequent protein identification relied on Edman sequencing, a strategy that routinely required the presence of approximately 1 pmol of highly purified starting material and frequently would give no results because of blocked N-termini [10]. The development of highly sensitive mass spectrometry instrumentation has reduced the quantity of starting material required for successful protein identifications more than one-hundredfold [11, 12]. This development, combined with the relative ease with which antibodies can be obtained that specifically recognize virtually any bait protein, has largely removed the need for complex purification schemes for interactome work [13]. The above trend was paralleled by improvements in protein-tagging strategies that facilitate high-throughput interactome mapping investigations. While some of the most popular tags (FLAG [14], c-Myc, hemagglutinin (HA), and V5) continue to rely on epitope-specific antibodies for interactome work, other tagging strategies, for example polyhistidine tags, glutathione-S-transferase (GST) tagging [15], calmodulin, or streptavidin-binding peptides are based on antibody-independent affinity capture steps [16–18].

Clearly, despite the recent dominance of reports based on tagged proteins in the interactome mapping literature, there is a continued need for conventional co-immunoprecipitations (co-IPs) [19]. Co-IPs remain a powerful interactome mapping tool in:

1. studies directed against endogenous proteins;
2. investigations of the molecular environment of a protein in a complex tissue; and
3. interactome mapping efforts for proteins that do not tolerate addition of tags or for which biology is altered after addition of tags.

As such, co-IPs are frequently the first experiments in a tedious and protracted experimental process with the ultimate purpose of identifying functional interactors of a protein of interest. Given the large investment in time and

resources that protein–protein interaction validation work demands, rigorous planning is essential to ensure highest quality of candidate interactors. We hope this article will contribute to this field by:

1. helping researchers avoid common pitfalls of co-IPs for interactome mapping studies;
2. reporting improvements to conventional immunoprecipitation procedures with the objective of increasing the sensitivity of downstream protein identification by tandem mass spectrometry (MS–MS); and
3. assisting to establish peer-review acceptance criteria for publications that employ co-IPs in their experimental strategy.

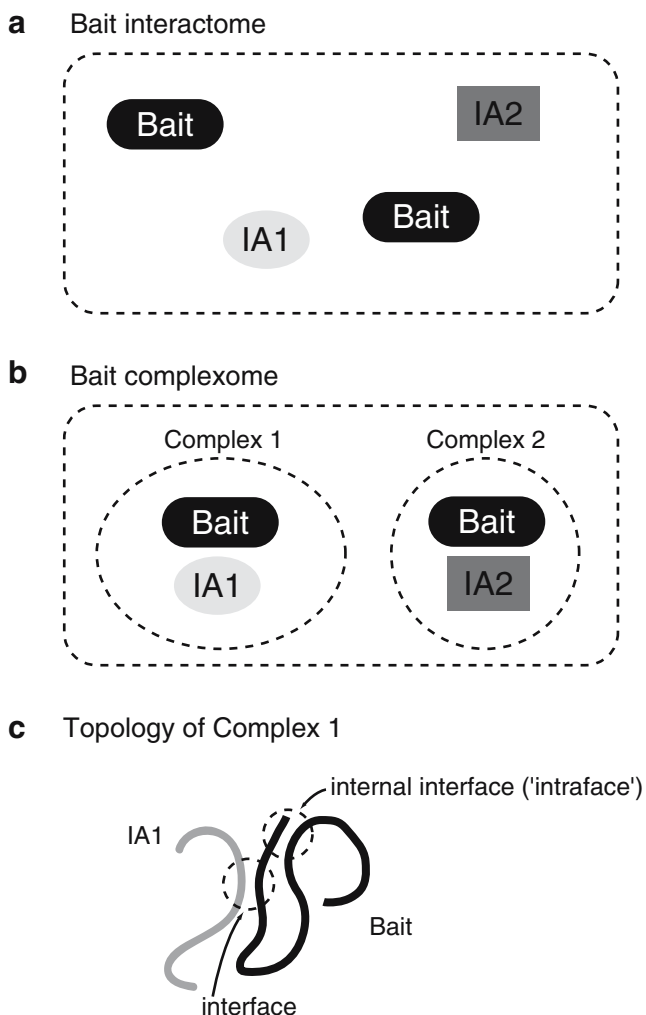
## Expected outcomes and general strategy

### Terminology of protein complexes

It is important to note that co-IP-based interactome mapping strategies alone will not establish direct interactions among proteins. Co-IP experiments further are poorly suited to defining boundaries of protein complexes or determining the stoichiometry of protein constituents within a given protein complex. Instead, the outcome of a co-IP experiment is a list of proteins that is populated by bait-specific interactors (bait interactome) and various other proteins which co-purify as a result of unspecific interactions (see below). Assignment of the bait protein within such a list to distinct protein complexes (bait complexome) typically requires the application of complementary methods. In systematic interactome studies, however, the composition of individual protein complexes can sometimes be deduced from bioinformatics analyses of data obtained from a large number of protein baits which give rise to overlapping interactome lists. Extensive downstream work is usually required to determine the topology of a protein complex, i.e. identify residues that reside in close spatial proximity within a protein (“intrafaces”) or define contact sites between two proteins that engage in direct interactions (interfaces) within a quaternary protein assembly (Fig. 1).

### The immunoprecipitation procedure

Co-IP experiments consist of a modular arrangement of steps. Individual steps are connected by well-defined interfaces, may be rearranged in order to some extent, and can be replaced with alternative submethods (Fig. 2). First, a suitable extract derived from cultured cells or a tissue extract is generated and cleared of all particulate and aggregated material. Next, an immunoaffinity matrix, generated by immobilizing a target-specific antibody to a matrix—for easy recovery of the antibody—is added to the



**Fig. 1** Terminology used to define the architecture of protein complexes. **(a)** Hypothetical interactome consisting of “bait” protein and two interactors, IA1 and IA2. **(b)** Hypothetical complexome depicting bait protein as a constituent of distinct protein complexes containing IA1 (“Complex 1”) or IA2 (“Complex 2”) (a more detailed classification of protein complexes has recently been proposed [82]. Please note that the direct interaction of proteins can often be inferred after extensive biochemical characterization of highly purified protein complexes or systematic application of interactome protocols to a large number of bait proteins. **(c)** Co-IP work does not usually provide insights into regions within a protein that contribute either to internal contact sites (“intraface”) or to the binding to another protein (interface) and thereby would be indicative of the topology of a protein complex

cell extract, and protein complexes containing bait proteins are immunocaptured. After extensive washes, proteins are eluted from the immunoaffinity matrix and are reduced, alkylated, and trypsinized with or without prior gel-separation. The resulting peptide mixture is subjected to tandem mass spectrometry. Finally, processing of MS–MS peptide fragmentation spectra by protein identification algorithms reveals the identity of proteins that putatively engage in interactions with the target protein.

### Concepts for the elimination of unspecific interactors

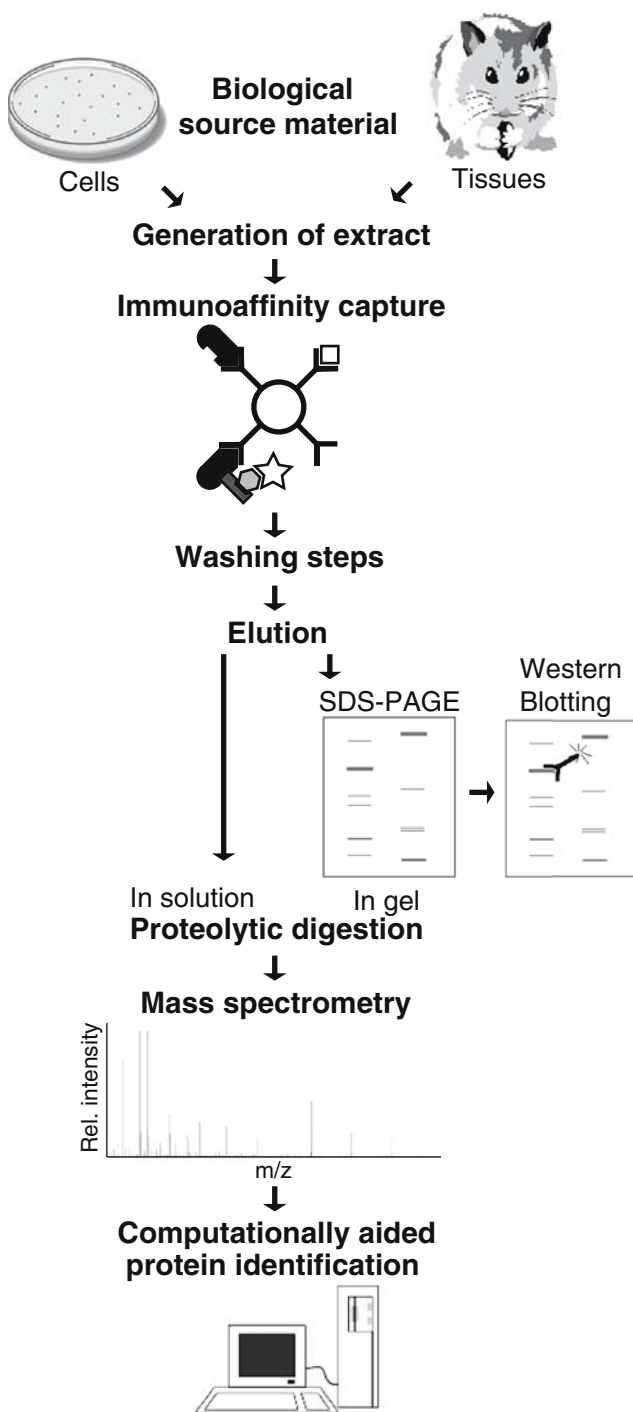
A key objective with any co-IP experiment is to minimize the risk of identifying unspecific interactors. Common explanations of the presence of these unspecific interactors in co-IP eluates are the presence of:

1. aggregated proteins in the sample that co-sediment with the immunoaffinity matrix;
2. proteins that bind directly to the immunoaffinity matrix;
3. proteins which under physiological conditions are found in a different cellular compartment than the protein target but have an intrinsic propensity to bind to the protein target when present in an extract;
4. abundant cellular proteins that populate eluate fractions when samples are subject to less than stringent washing conditions;
5. proteins that originate from the immunoaffinity matrix themselves, e.g. if crude antibody preparations are coupled to chemically activated matrix beads; and, finally,
6. proteins such as trypsin, human skin and hair proteins, etc. introduced into the sample as a result of sample-handling procedures.

With this many possible sources of unspecific proteins, rather than aiming to eliminate all unspecific contaminants the objective is to minimize their occurrence and, more importantly, to know their identities.

### Bait exclusion strategy

Two alternative design concepts that we refer to as the “bait exclusion strategy” and the “tool exclusion strategy” are particularly important and are suited to identification of most non-specific interactors (Fig. 3). The bait exclusion strategy is conceptually the most rigorous approach. It uses side-by-side co-IPs from starting materials which differ only in the presence or absence of the bait protein. As all steps of the process employ the same reagents and procedures, differences in the final list of candidate protein interactors are attributed to the expression versus “knock-out” of the protein of interest. In instances where no knock-out samples are available two derivative strategies can be used. The first of these employs a side-by-side co-IP of identical wild-type extracts but relies on presaturation of the bait-specific antibody with the peptides used to raise that antibody in the control sample. In the second strategy the knock-out of a bait protein is replaced with a mere knock-down using RNA interference technology. These derivative strategies may not achieve complete suppression of bait protein capture in the control sample, however. As a result it can be difficult to distinguish between specific candidate protein interactors and non-specific interactors, because the



candidate protein interactors may also be present in control samples, albeit at lower quantities. It is therefore advisable to consider combining these derivative strategies with isotope-labeling procedures that enable quantitative comparison of samples [20–22]. Although the simplicity of the bait exclusion strategy is compelling, caution is nevertheless warranted, because a protein may end up in the immunoprecipitated sample as a result of its ability to non-physiologically, and thus unspecifically, bind directly

◀ **Fig. 2** Schematic representation of a generic co-IP experiment. Complex biological source materials—most frequently an extract obtained from cells or tissues—are cleared of all insoluble protein material. After addition of bait-specific immunoaffinity matrices, protein complexes containing the bait protein are captured and matrices are subject to washing steps that reduce the amount of unspecific binders. Elution from the matrix is followed by reduction, alkylation, and proteolytic cleavage either in solution or within excised gel slices. Peptide mixtures are analyzed by tandem MS or peptide mass fingerprinting. Subsequent queries of genomic databases with mass spectrometric data enable identification of candidate interactors. In small-scale co-IP experiments used for method development and validation of candidate interactors the gel-separation step is usually followed by transfer to Western blots which then are probed with bait-specific antibodies or antibodies directed against candidate interactors

to the bait protein or the immunoprecipitated complex on disruption of sample tissue. In addition, a protein for which expression is down-regulated as an indirect result of bait protein knock-out and thus conspicuously absent from the immunoprecipitated material may be mistaken for a physical interactor.

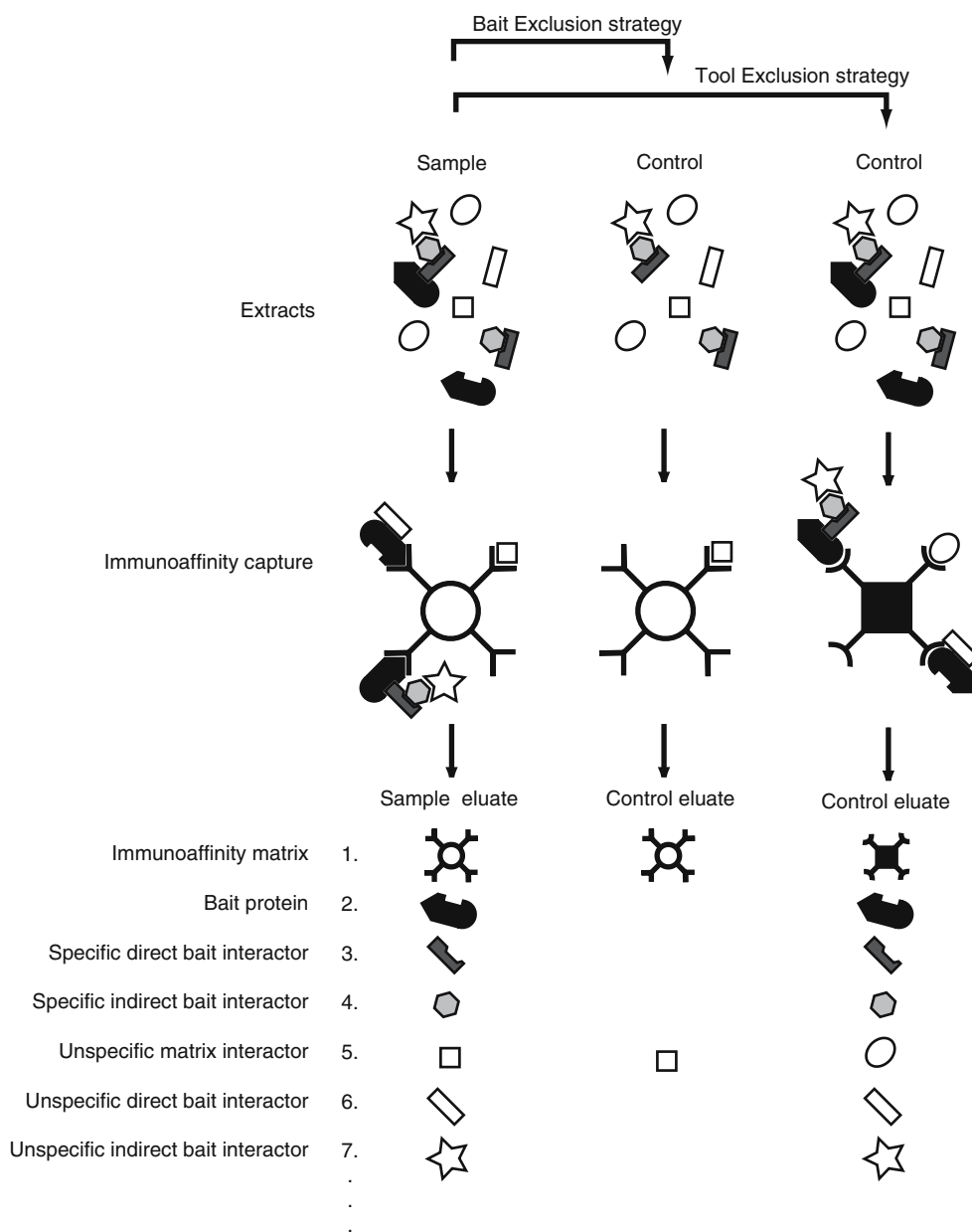
#### Tool exclusion strategy

This approach is essentially built around concepts that are diametrically opposed to concepts underlying the bait exclusion strategy. With the latter, identical tools are applied to sample and control immunoprecipitations, and bait interactors populate sublists of non-overlapping protein identifications (IDs). The tool exclusion strategy, however, relies on common protein IDs in immunoprecipitation data sets obtained with a minimum overlap of samples and tools. The most faithful implementation of this concept would employ samples from diverse origins, would have no overlap in any of the tools used for immunoprecipitation, i.e. the matrix, coupling reagents, antibodies etc. Samples handled in parallel would have one shared feature, however—they would employ antibodies that were raised against the same bait protein. It is immediately apparent that a perfect implementation of this strategy is not realistic. Consequently, in applying this approach one needs to be alert to the fact that some of the proteins common to parallel-processed samples will represent highly abundant proteins that unspecifically and promiscuously bind to a large number of protein partners.

#### The conventional approach

The vast majority of co-IP based interactome data reported to date used neither bait exclusion nor tool exclusion strategies. Although in recent times application of these concepts has found more widespread use, a significant share of contemporary studies still omit safeguards facili-

**Fig. 3** Schematic diagram depicting bait exclusion and tool exclusion strategies for identification of non-specific interactors in co-IP experiments. The bait exclusion strategy employs parallel co-IPs from starting materials which differ in the presence or absence of the bait protein. As a result bait-specific interactors represent a subpopulation of proteins exclusively found in eluate fractions derived from the bait-containing biological source material. The tool exclusion strategy is, for both sample and control samples, based on biological source materials that contains the bait protein but relies on a minimum overlap of tools (biological source material, antibodies, immunoaffinity matrix, etc.). Specific bait interactors will, as a result, constitute a subset of proteins common to sample and control co-IP data sets and unspecific interactors are expected to differ substantially between sample and control eluate fractions



tating the identification of unspecific proteins. The list of candidate interactors is frequently obtained after co-IPs with a single antibody performed in parallel with either a “matrix only” IP or a “mock” IP in which the specific antibody is replaced with a generic control antibody. Protein lists are then manually screened for the presence of a protein which, from the known biology of the bait protein (location, function, domain structure, etc.), makes “sense” as a potential physiological interactor. For publication in a peer-reviewed journal, the minimum requirement is, typically, to show that at least one of the candidate interactors co-localizes with the bait protein and/or co-enriches with the bait protein in reciprocal co-IPs. Naturally, the inherent bias underlying the selection of proteins for validation studies favors the study of better-known proteins.

Also, not surprisingly, a large share of candidate interactors that populate such interactome data lists can be expected to represent unspecific binders. The reason this approach enjoys great popularity despite its shortcomings is the relative ease with which data can be generated and the fact that it requires neither bait-specific knockout samples nor a second bait-specific antibody. In our own experience, this strategy has merit only when discriminatory power is added through side-by-side comparison of interactome data from a range of different bait proteins. For this to work, bait-specific pull-downs directed against different bait proteins must be performed from the same biological source material using antibodies coupled in the same fashion to identical matrices. Thus most unspecific binders in the eluate sample will be found in all interactome datasets and

thereby greatly facilitate the mining of candidate interactor lists. A caveat with this strategy remains the uncertainty that an unspecific protein may appear in the candidate list of specific interactors as a result of a cross-reacting antibody. Before significant investments in downstream validations it is therefore advisable to verify that a protein does not merely represent a target for a cross-reacting antibody (Fig 4).

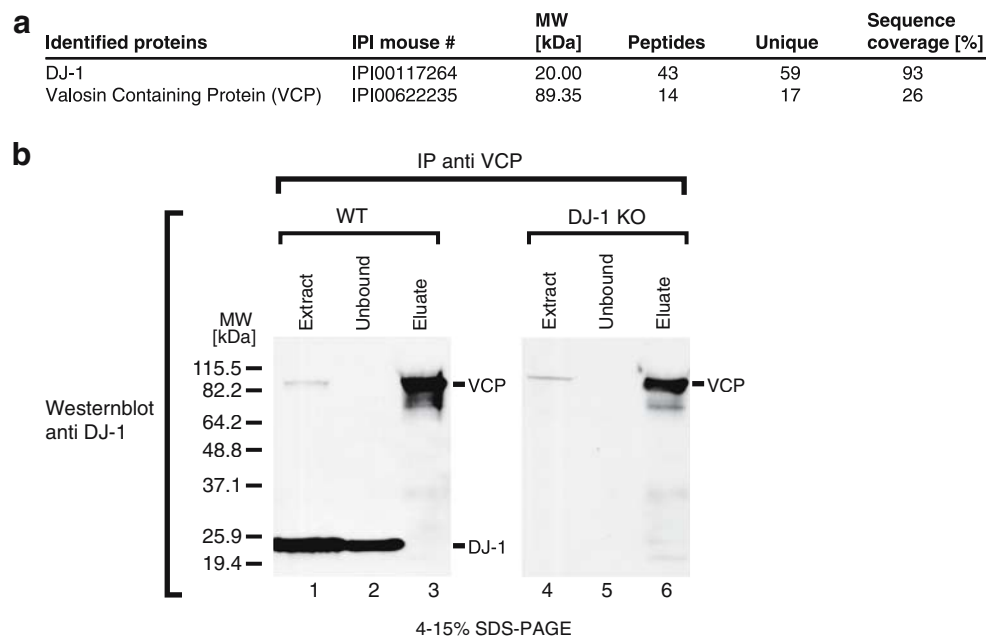
#### Quantitative mass spectrometry

The challenge in designing a co-IP experiment is the need to capture relevant physiological interactors under conditions that promote as few non-specific protein interactions as possible. In recent years two divergent strategies have emerged to address this problem. The first strategy employs isotopic labeling reagents to label samples and negative controls differently [20, 21, 23]. Rather than trying to limit unspecific binding, this philosophy capitalizes on the quantitative power provided by isotopic labeling reagents and aims to distinguish specific from unspecific interactors by their relative abundances in side-by-side analyzed samples [24]. The strength of this approach is its ability to tolerate the presence of a large number of contaminant proteins. When combined with either the overexpression [25] or the siRNA-mediated down-regulation of the bait

protein in a negative control sample [22] this approach can readily reveal candidate interactors of interest. This approach comes at a price, however—the requirement for significant front-end separation and a concomitant reduction in sensitivity. It is nevertheless the only practical strategy in interactome studies directed toward abundant cellular proteins where, unless quantitative data are available, it may be impossible to distinguish bona fide interactors from unspecific contaminants. This approach further eliminates the risk of a mass spectrometry inherent sampling bias, a common phenomenon whereby run-to-run variances in the analyses of complex samples may be misinterpreted to reflect sample-to-sample differences [26].

#### In vivo crosslinking

The objective of this second strategy is to minimize the presence of unspecific interactors by stringent salt and detergent washing of immunoaffinity-captured target protein complexes. For protein complexes to survive this treatment their constituents must be covalently coupled to each other through an in-vivo crosslinking procedure, in which cells [27–30] or tissues [31] are subjected to short treatments with chemical or photoactivatable crosslinkers [32]. This approach greatly reduces the risk of being misled



**Fig. 4** Example of initial validation experiment which firmly establishes a candidate interactor as an unspecifically cross-reacting protein. **(a)** Valosin-containing protein (VCP) was identified in a DJ-1 specific large-scale co-IP data set as a strong candidate interactor on the basis of more than ten unique CID spectra. **(b)** VCP-directed IP from wild-type and DJ-1 knockout mice followed by Western blotting with DJ-1-directed antibodies unequivocally establishes unspecific crossreactivity of DJ-1-directed antibody with VCP. Brains of WT mice and DJ-

1-knockout mice were subject to small-scale co-IPs with a VCP-directed antibody. Extracts before co-IP (*lanes 1 and 4*), unbound co-IP material (*lanes 2 and 5*), and co-IP eluate fractions (*lanes 3 and 6*) were analyzed by Western blotting with DJ-1-directed antibody (the same antibody that had been used for original large-scale DJ-1 co-IP experiment). Please note the increase in the DJ-1 antibody signal for a band of ~95 kDa (VCP has apparent MW of 97 kDa) after VCP-directed co-IP from DJ-1 knockout mouse extracts (*lanes 1 and 3*)

in situations where interactors bind directly to the protein of interest or the immunoprecipitated protein complex, when present in an extract, but do not physiologically interact with the target protein as long as cellular integrity is maintained. This strategy also has a clear advantage over alternative approaches for the study of labile protein complexes and protein complexes that require a particular milieu for integrity [33], for example membrane protein complexes, which notoriously and unpredictably may disassociate when subjected to common solubilization buffers. The sensitivity of this approach will be affected by the spatial distribution of chemical groups within bait protein complexes that can engage in productive crosslinks [34].

## Modules of a Co-IP experiment

### Biological source material

For interactome studies of multicellular organisms a choice must be made whether a bait protein is purified via co-IP from either cultured cells or tissue material. Naturally, for any multicellular organism tissue represents the most authentic environment of a protein—the exceptions here are blood cells and other non-sessile cell types. Tissue also has the ability to capture the authentic molecular environment of membrane proteins that may involve interactions in trans with proteins expressed on the surface of a different cell type [31]. Even for intracellular proteins, however, interactomes may differ, depending on whether a given protein is expressed in a homogenous cell population (cell clone) or in a heterogeneous and complex cell environment, because extracellular contacts and stimuli have been shown to initiate intracellular molecular rearrangements [35]. This added feature of tissue-based co-IP work, however, may complicate efforts to delineate which cell type contributed to a given protein interaction later during data analysis. While metabolic labeling with stable isotopes can be used even for complex organisms [36], cell culture work is clearly the economic choice if co-IP work is combined with metabolic labeling strategies (see below) for mass spectrometry-based quantitation.

### Generation of extract

Naturally, initial cell and tissue handling steps will depend on the biochemical properties of the bait protein and its subcellular location. For large-scale experiments it is advisable to first test a range of buffer conditions for their ability to solubilize the bait protein and monitor relative yields by Western blotting. In choosing salt and detergent conditions, a key concern is to maintain the integrity of the bait complex while minimizing unspecific interactions.

Most co-IP procedures reported in the literature use near-physiological salt concentrations and ~1% (v/v) non-ionic detergents, for example Triton X-100 or NP-40, for the capture of bait proteins. For co-IP work it is advisable to use concentrated extracts (5–20 mg mL<sup>-1</sup> protein), because a high concentration of all extract constituents increases the chance of detecting low-affinity protein–protein interactions with high binding constants.

### Source of antibody

By far the most frequently encountered antibodies in the co-IP literature are rabbit polyclonal and mouse monoclonal antibodies. Detailed descriptions of procedures for generation and characterization of these antibodies can be found elsewhere [37]. Here, comments will be restricted to aspects of antibody use in co-IPs not commonly covered in the pertinent literature. Both classes of antibody are well suited to immunoaffinity capture, assuming a given antibody recognizes its protein target in solution with sufficient affinity and specificity. The affinity of an antibody will determine the minimum concentration at which a bait protein must be present for its capture. The antibody specificity is indicative of the selectivity with which an antibody recognizes its antigen as opposed to crossreactive binding to other proteins. Whereas the low affinity of an antibody can frequently be dealt with by pre-fractionation steps that increase the concentration of target proteins, there is no good remedy to counteract low specificity other than rigid implementation of strategies that facilitate identification of unspecifically enriched proteins. A small-scale study that involves immunoprecipitation of bait proteins followed by Western blotting analysis is recommended to test the utility of an antibody for co-IP work. If polyclonal rabbit antibodies are used, antibodies raised against protein fragments are preferred over those raised against recombinant full length protein, because the probability the latter may be contaminated with a cross-reacting subpopulation of immunoglobulins increases with the size of the antigen. Antibody preparations are frequently either contaminated with serum proteins or contain bovine serum albumin (BSA) that has been added to stabilize the antibody during transport and storage; this again may cause the appearance of common serum contaminants of commercial BSA preparations in data sets. If not removed, these proteins may undermine the coverage and sensitivity with which protein constituents of bait protein complexes can be detected in subsequent mass spectrometric analysis.

### Choice of matrices

A large selection of matrices is available for immobilization of antibodies. Most of these matrices use a beaded-agarose

core. For accelerated workflows, magnetically derivatized beads (Dynabeads; Invitrogen Dymal) have been in use for some time. In the simplest implementation, beads are merely surface derivatized with a reactive group chemistry (often provided at the distal end of an extended linker arm) that enables rapid covalent capture of purified antibodies [38]. Alternatively, similar beads may be purchased pre-derivatized either with an antibody capture reagent, for example protein A [39] or protein G [40], or with a protein such as avidin/streptavidin or genetically engineered neutravidin, which captures biotinylated antibodies [41–43]. Chemically activated matrices work well if highly purified antibody preparations are used, because they provide ease of handling and reduce the level of cross-contaminating proteins and unspecific binders that may bind to protein A/G or avidin derivatives themselves. In particular, the capture of biotinylated antibodies with avidin/streptavidin-derivatized matrices tends to be problematic, because it invariably leads to concomitant enrichment of physiologically biotinylated proteins and their interactors (Electronic Supplementary Material; Table S2) [44]. The use of chemically activated matrices is also appropriate when co-IPs have to rely on antibodies such as chicken immunoglobulins that bind neither to protein A nor protein G. A limitation of chemically activated beads is, however, the need to use about twice as much antibody as in protein A/G based co-IPs, probably because of a significant share of antibody that unproductively couples in an orientation which sterically interferes with subsequent capture of target proteins. In instances when the purity of an antibody preparation is questionable, or when only low levels of antibody are available, use of protein A/G pre-derivatized beads is, therefore, recommended. The protein A/G bacterial preparations from common manufacturers seem to be of high purity, because prominent bacterial contaminants are not usually encountered in protein A/G agarose-derived data sets.

Unless the coupling of capture antibodies to agarose matrices relies on the high-affinity avidin–biotin bond, it is critical to covalently link the antibody to the matrix. Whereas such a crosslinking step may not be needed if co-IPs are followed by conventional Western blotting analyses, it is critical for applications involving subsequent mass spectrometry, because excessive leakage of the antibody into the eluate may mask low-abundance peptides derived from low-stoichiometry interacting proteins [45]. Most frequently, reactive chemistries used for such a crosslinking step target accessible amino groups on the antibody surface [46]. Despite this crosslinking step, residual leakage of antibody into the eluate does occur, and peptides derived from the antibody often feature prominently in mass spectrometry data sets. If the antibody retains affinity for its target under elution conditions, it is therefore advisable to subject the affinity matrix to a brief wash in elution buffer to minimize subsequent leakage of

antibodies before the bait protein-capture step. It is important to quench unreacted groups with a quenching reagent after antibody coupling. Here, the commonly employed chemical ethanolamine can be problematic, because it tends to cause selective enrichment of a subset of cellular dehydrogenases (unpublished results; Electronic Supplementary Material; Table S3), an observation which has previously been exploited in a different context for deliberate enrichment of bacterial dehydrogenases [47].

Immunoprecipitation procedures frequently include overnight incubations at reduced temperature for capture of bait proteins. In co-IP work such a relatively long incubation period is only advisable for covalently crosslinked samples. It has been reported that shorter incubations with high-affinity capture antibodies may increase the yield of weakly associated specific interactors not crosslinked to the bait [48].

#### Handling of immunoaffinity matrices

Efficient washing of immunoaffinity matrices after the immunocapture step is essential for overall success of the experiment when co-IP eluates are subjected to mass spectrometric analysis. As far as we are aware, systematic analysis of the effect of washing stringency on the presence of unspecific protein IDs in mass spectrometry data sets has not been reported. The following general comments and recommendations for removal of unspecific interactors are therefore based solely on observations and have not been subjected to more rigorous scrutiny. Once captured, harsh conditions of up to  $1 \text{ mol L}^{-1}$  salt and addition of small amounts of denaturing detergents (e.g. 0.2% SDS) are frequently tolerated by protein complexes and may reduce the number of unspecifically purified proteins. As in most applications, however, the stability of a protein complex in the presence of harsh environments is not predictable and it may initially be safer to use extensive washing with lower stringency buffers rather than to risk losing relevant but weakly-binding interactors as a result of harsh washing conditions. A stringent washing procedure may subject bead material to three to five consecutive washes with a five-hundredfold excess of washing buffer. To avoid trapping unspecific proteins between squeezed agarose beads it is advisable to use minimum centrifugal force for collection of the beads. In fact, because of the relatively high density of most agarose bead materials, it is usually sufficient to collect beads between washing steps by gravity sedimentation. To avoid contamination of eluates with agarose bead material and to maximize recovery of protein targets, co-IP tubes containing the matrix slurry can be pierced at their bottom with a fine needle before the elution step. Alternatively, commercially available mini-spin columns [49] can be used for large-scale co-IPs. Both approaches enable collection of eluates by unidirectional



flow and provide good control of elution flow-rates when a syringe is mounted on the top-opening of the tube and gentle air pressure is applied to the elution buffer. Irrespective of the matrix sedimentation approach used, the final handling steps of the washed bead material and all subsequent steps up to the proteolytic digestions (e.g. trypsin) should be performed in a dust-free environment (e.g. a horizontal laminar flow hood) to minimize the risk of external contamination of samples with keratins.

#### Elution from matrices

A key consideration in the design of the elution step is choosing buffer compositions that are compatible with subsequent mass spectrometric analysis. A variety of strategies have been used for dissociation of antibodies from their antigens; these include adjustment of pH, addition of chaotropic salts, and competitive elution with excess peptide antigen [50]. Of these, by far the most popular strategy utilizes a rapid decrease in pH (with or without the addition of additional organic solvents for the disruption of hydrophobic interactions). For conventional immunoprecipitations this step usually employs an acidified glycine solution to generate the low pH. To achieve maximum recovery of bait protein complexes for subsequent mass spectrometry it is advisable to replace the low-pH glycine buffer with a solution of trifluoroacetic acid (for example 0.2% TFA) in water (with or without acetonitrile). This change to a chemistry based on volatile reagents enables elution from the immunoaffinity capture matrix with a larger volume of elution buffer without introducing unwanted salts. Similarly, instead of using concentrated Tris buffers for subsequent pH neutralization, it is advantageous to use a mixture of ammonium bicarbonate and aqueous ammonia (for example 0.05%  $\text{NH}_4\text{OH}$  in 25 mM  $\text{NH}_4\text{HCO}_3$ ) during this step. These reagents enable the rapid concentration of immunoaffinity eluates by centrifugal vacuum concentration. If the immunocapture step is based on a monoclonal antibody that recognizes a well-characterized linear peptide epitope, competitive elution in the presence of an excess of synthetic peptide antigen has the advantage of selective elution at neutral pH and thereby reduces contamination of the eluates with unspecific matrix interactors. To be successful with this method, the antibody must have a dissociation constant ( $K_{\text{off}}$ ) that is low enough to avoid extensive leakage from the matrix during washing steps, but high enough to mediate efficient displacement in the presence of reasonable concentrations (typically added at 50–500  $\mu\text{g mL}^{-1}$ ) of competitive peptide antigen [51, 52].

#### In-solution versus in-gel processing

The pros and cons of in-solution or in-gel processing are contentious. So far the co-IP literature is dominated by

studies that use denaturing SDS-PAGE for initial interactome analysis [53]. After gel staining, bands interpreted as being those of candidate interactors are typically excised from the gel, subjected to in-gel digestion, and subsequently analyzed by peptide mass fingerprinting or tandem mass spectrometry [54]. This strategy enables convenient assessment of sample complexity and relative stoichiometry of interactome components. Shortcomings of this approach are the sample losses associated with multiple sample handling, an inherent bias toward proteins that can be resolved within the molecular weight range limitations of the SDS-PAGE gel, and its reliance on time-consuming manipulations. It has increasingly been recognized that apart from the obvious increase in speed, a move to in-solution digests may also be paralleled by improvements in sequence coverage and peptide yield [55].

#### From raw interactome data sets to lists of candidate bait interactors

It is helpful to sort proteins in an order that reflects sequence coverage because this often correlates well with the abundance of a protein in a sample [56]. Any raw interactome data set typically includes protein IDs (for example IgG, trypsin, keratins, etc.) that are readily recognizable as being introduced during sample handling steps and can, therefore, be eliminated from the list. A critical requirement for meaningful subsequent validation studies is that the bait protein is not only represented in a given data set but gives rise to the most unequivocal protein identification in terms of both signal intensities of parent ions and percent protein coverage. Rare scenarios in which a bait protein engages in very tight associations with other proteins that give rise to peptides with more favorable ionization characteristics, or that are present within a core complex in numbers that exceed the quantity of the bait, are exceptions to this requirement. Conversely, it is expected that physiologically relevant but intrinsically dynamic or weak interactions would give rise to low spectral counts. Thus, relatively weak representation within an interactome data set does not exclude a protein from being a physiologically important interactor. Incorporation of negative controls in the experimental strategy, in particular when combined with isotopic labeling for quantitative mass spectrometric analysis, enables efficient removal of unspecific interactors by comparative analysis of sample and control protein lists. For quantitative analysis a relative abundance threshold must be set before this filtering step. Reasonable strategies here are to base this abundance threshold on the empirical abundance distribution of peptides added as internal standards before the isotopic labeling step or peptides expected to be present at equal levels in sample and negative control lists (e.g. IgGs, trypsin,

and some common endogenous contaminants, for example tubulin and actin). Additional elimination of proteins from bait interactome lists can be based on thresholds for low confidence identifications. Finally, the list can be simplified by removal of proteins whose identifications were exclusively based on peptides they share with other proteins on the bait interactome list and thus are not supported by unique and independent peptides—a heuristic approach which favors proteins identified with the largest number of peptides to address a complex problem in large data sets commonly referred to as the protein inference problem [57, 58].

With regard to the predicted nature and size of data sets, it is reasonable to assume that any target protein may engage in interactions with proteins that facilitate its formation, transport, post-translational modification, cellular function, and degradation. Clearly, an interactome study that reveals no interaction or an excessive number of interactions (e.g. >100 specific candidate interactors) should raise suspicions about the quality of the data. In such circumstances the presence of a significant subset of proteins known to physically or genetically interact with each other may indicate an excessive presence of indirect interactors.

For protein baits with previously validated interactors the presence of these known interactors in the data set can serve as an internal validation and facilitate assessment of data quality. A range of databases which assist in this task are available [90]. It is frequently necessary to screen lists of candidate interactors for the most likely functional interactors. Caution must be exercised during these steps to avoid bias, and the following comments may merely serve as starting points in this difficult task.

#### Gene ontology annotations (GO)

These describe the molecular function, biological process, and cellular component with which a target protein has been associated and may help to group candidate interactors [59]. A reasonable strategy here is to limit initial validation efforts to the candidate interactor within a classification group for which the highest percentage of sequence coverage was obtained. Similarly, a detailed sequence analysis occasionally reveals the presence of shared subdomains in a subset of candidate interactors, and suggests selective affinity of the target protein for proteins that harbor this domain [60]. As above, in such a scenario it may be advisable to initially restrict validation efforts to the candidate interactor containing this shared sequence domain which gave rise to the most confident identification.

#### “Expert eyes” and PubMed searches

This approach is regarded as the most controversial, because it is potentially fraught with bias. It is, however, unlikely that

anyone would resist the temptation to screen literature databases for corroborating evidence supporting the notion of a possible interaction between a bait protein and its candidate interactors. There is also little doubt that valuable resources can be saved by recruiting the opinion of experts who understand the biology of the bait protein and others who have previously seen dozens of similar interactome data sets (Fig. 5).

#### Initial validation of candidate interactors

When promising interactors of a bait protein have been selected for further investigation, an array of biochemical and genetic methods can be used to characterize these proteins and probe for their involvement in physiological activity that governs the biology of the bait protein. If more than just a few candidate interactors must be validated in parallel, it may be most economical to base an initial screening on rtPCR. Underlying this recommendation is the observation that many functional interactors are subject to transcriptional co-regulation [61]. This approach requires parallel harvesting of mRNA from cells or tissues that express different levels of either the bait or candidate interactor. Alternatively, the effect of an RNAi-based knockdown of candidate interactors on the expression and post-translational modification of the bait protein may be investigated [62]. Both approaches generate data that are orthogonal to the physical interaction data sets collected in co-IP studies. Additional biochemical validation tools to be considered are:

- overexpression analyses of selected targets;
- reciprocal co-IPs;
- glycerol velocity gradient centrifugation [63];
- size-exclusion chromatography [64–66];
- iodixinol gradient centrifugation [67];
- immunocytochemistry;
- in-situ hybridization of tissue sections; and
- functional cell based assays.

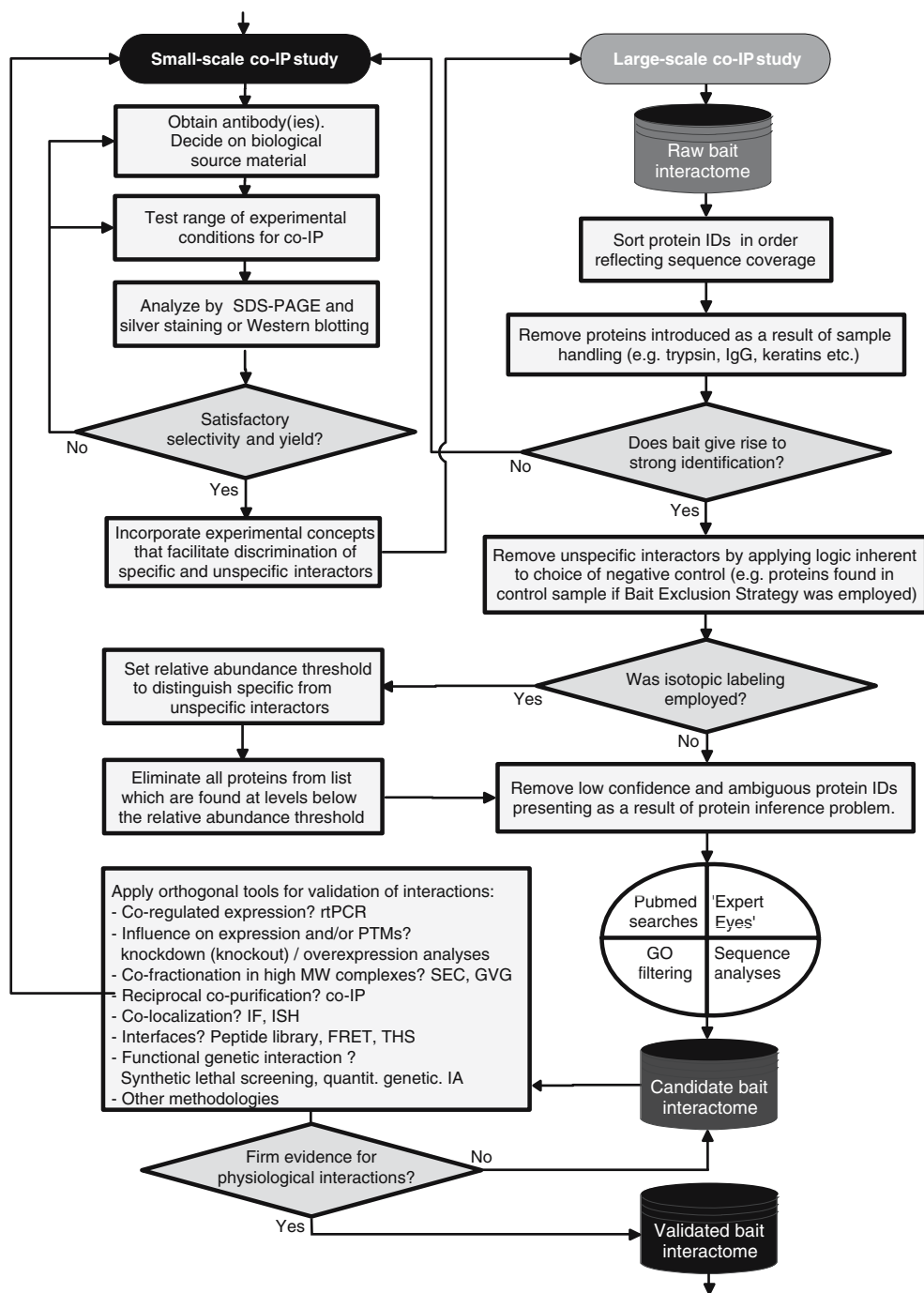
Orthogonal genetic tools for validation of protein interactions are:

- the two-hybrid system (THS) [68, 69];
- fluorescence resonance energy transfer analysis (FRET) [70, 71] and derivative strategies [72–78];
- synthetic lethal screening [79, 80]; and
- the recently reported quantitative genetic interaction mapping method [81].

In addition to establishing whether a given candidate interactor is a physiological interactor of the bait protein, the objective of the above studies should be to:

1. delineate whether the bait protein engages in interactions with multiple candidate interactors as part of a

**Fig. 5** Diagram summarizing work flows of co-IP experiments. The diagram puts emphasis on the incorporation of pilot studies, informative controls, and strategies for data mining. A subset of complementary strategies for initial candidate interactor validation is provided. Additional explanations can be found in the text



- single protein complex or binds to a subset of its interactors within multiple distinct complexes [82];
- 2. provide insights into the stoichiometry of protein constituents within a complex [83]; and
- 3. determine the contact sites (interfaces) between proteins [32, 84–87].

Naturally, the choice of method will depend on the nature of the candidate interactor, the availability of specific immunoreagents, cell or animal models, specific inhibitors/

agonists, and whether a knock-down/overexpression of the bait protein and/or candidate interactor can be achieved.

### A common standard for co-IP datasets

There is a trend toward development of standards and guidelines for publication of proteomics data [88, 89]. A similar discussion on useful conventions for co-IP datasets is

much needed. As a starting point we recommend adherence to the bait exclusion or tool exclusion concepts and their derivative strategies (summarized above). It should thus be considered insufficient if negative control data on unspecific binders are based on either “matrix only” precipitations or “mock IPs” employing generic IgGs. We further strongly recommend incorporation of isotope-labeling strategies into the experimental strategy to minimize the effects of sampling bias for complex samples and to afford relative quantitation of candidate interactors in bait-specific and control datasets. Finally, adherence to these more rigorous design concepts should extend to small-scale reciprocal co-IPs that frequently constitute the first validation experiment in publications which document novel protein–protein interactions.

### Concluding remarks

In this review we have reported improved co-IP procedures for applications that use subsequent mass spectrometry to generate interactome maps of endogenous bait proteins. We have presented alternative experimental design strategies that facilitate recognition of unspecific interactors in interactome data sets. The need to incorporate concepts that facilitate recognition of false positive interactors has also been emphasized and recommendations for successful setup of co-IP experiments have been presented. We have also discussed strategies for the filtering of raw interactome data and provided guidelines for assessment of data quality and the design of pilot experiments for subsequent validation of candidate interactors.

There are, clearly, many ways to design and implement a successful co-IP-based interactome mapping experiment. The experiment involves a complex series of individual steps and, frequently, there are multiple directions one can follow without compromising the overall outcome of the experiment. Thus, in writing this article we attempted to find a compromise between offering specific advice and providing broad guidelines. Whenever a design element or experimental step seemed superior to alternative choices this was pointed out in the text. We do, however, accept that any specific implementation of a co-IP experiment will depend on the target protein under investigation, the sample material from which it is to be purified, and the immunoaffinity reagents available for its capture. Similarly, the unspecific interactors encountered in individual data sets are expected to depend on the nature of the sample and the particular co-IP strategy used.

The intention of this article was to fill a void in the proteomics literature with regard to co-IP work. We hope this article will assist researchers who embark on co-IP work to avoid common pitfalls and will stimulate further improvements to the concepts and current practices presented here.

**Acknowledgements** We thank Rasanjala Weerasekera, Tujin Shi (Centre for Research in Neurodegenerative Diseases, University of Toronto, Canada) and Rick Bagshaw (Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada) for their help during the preparation of this manuscript. Work on this article was funded through grant support from the Canadian Institute for Health Research (MOP-74734), the Canadian Foundation for Innovation (CIHR #10087), and the National Institute of Health (#GM61898). YB was funded through the CIHR Strategic Program in Protein Folding. GSU received generous support from the W. Garfield Weston Foundation.

### References

1. Aloy P, Russel RB (2002) *Trends Biochem Sci* 12:633–638
2. Pandey A, Mann M (2000) *Nature* 405:837–846
3. Sobott F, Robinson CV (2002) *Curr Opin Struct Biol* 12:729–734
4. Falkenburg PE, Haass C, Kloetzel PM, Niesel B, Kopp F, Kuehn L, Dahlmann B (1988) *Nature* 331:190–192
5. Keshgegian AA, Austin GE, Meltzer SM, Furth JJ (1975) *Cancer Res* 35:310–315
6. Cuatrecasas P (1970) *J Biol Chem* 245:3059–3065
7. Fritze CE, Anderson TR (2000) *Methods Enzymol* 327:89–92
8. Jarvik JW, Telmer CA (1998) *Annu Rev Genet* 32:601–618
9. Wilchek M, Miron T, Kohn J (1984) *Methods Enzymol* 104:3–55
10. Matsudaira P (1993) *A practical guide to protein and peptide purification for microsequencing*. Academic Press, San Diego, CA, USA
11. Aebersold R, Mann M (2003) *Nature* 422:198–207
12. Gevaert K, Vandekerckhove J (2000) *Electrophoresis* 21:1145–1154
13. Otto J, Lee S (1993) *Methods Cell Biol* 37:119–127
14. Einhauser A, Jungbauer A (2001) *J Biochem Biophys Methods* 49:455–465
15. Smith DB, Johnson KS (1988) *Gene* 67:31–40
16. Bauer A, Kuster B (2003) *Eur J Biochem* 270:570–578
17. Chang IF (2006) *Proteomics* 6:6158–6166
18. Monti M, Orru S, Pagnozzi D, Pucci P (2005) *Biosci Res* 25:45–56
19. Ransone L (1995) *Methods Enzymol* 254:491–497
20. Goshe MB, Smith RD (2003) *Curr Opin Biotechnol* 14:101–109
21. Sechi S, Oda Y (2003) *Curr Opin Chem Biol* 7:70–77
22. Selbach M, Mann M (2006) *Nat Methods* 3:981–983
23. Blagoev B, Mann M (2006) *Methods* 40:243–250
24. Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, Aebersold R (2003) *Nat Genet* 33:349–355
25. Tackett AJ, DeGrasse JA, Sekedat MD, Oeffinger M, Rout MP, Chait BT (2005) *J Prot Res* 4:1752–1756
26. Liu H, Sadygov RG, Yates JR (2004) *Anal Chem* 76:4193–4201
27. Farrell IS (2005) *Nat Methods* 2:377–384
28. Hino N, Okazaki Y, Kobayashi T, Hayashi A, Sakamoto K, Yokoyama S (2005) *Nat Methods* 2:201–206
29. Schmitt-Ulms G, Legname G, Baldwin MA, Ball HL, Bradon N, Bosque PJ, Crossin KL, Edelman GM, DeArmond SJ, Cohen FE, Prusiner SB (2001) *J Mol Biol* 314:1209–1225
30. Vasilescu J, Guo XJ, Kast J (2004) *Proteomics* 12:3845–3854
31. Schmitt-Ulms G, Hansen K, Liu J, Cowdrey C, Yang J, DeArmond S, Cohen FE, Prusiner SB, Baldwin MA (2004) *Nat Biotechnol* 22:724–731
32. Melcher K (2004) *Curr Protein Pept Sci* 4:287–296
33. Herrmann JM, Westermann B, Neupert W (2001) *Methods Cell Biol* 65:217–230
34. Wong SS, Wong LJ (1992) *Enzyme Microb Technol* 14:866–874
35. Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Crougton K, Cruciat C, Eberhard D, Gagneur J, Ghidelli S, Hopf C, Hushe B, Mangano R, Michon AM, Schirle M, Schlegl J,

- Schwab M, Stein MA, Bauer A, Casari G, Drewes G, Gavin AC, Jackson DB, Joberty G, Neubauer G, Rick JM, Kuster B, Superti-Furga G (2004) *Nat Cell Biol* 6:97–105
36. Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR (2004) *Anal Chem* 76:4951–4959
37. Harlow E, Lane D (1988) *Antibodies: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
38. Penzol G, Armisen P, Fernandez-Lafuente R, Rodes L, Guisan JM (1998) *Biotechnol Bioeng* 60:518–523
39. Ey PL, Prowse SJ, Jenkin CR (1978) *Immunochemistry* 15:429–436
40. Akerstrom B, Brodin T, Reis K, Bjorck L (1985) *J Immunol* 135:2589–2592
41. Gretch DR, Suter M, Stinski MF (1987) *Anal Biochem* 163:270–277
42. Hiller Y, Gershoni JM, Bayer EA, Wilchek M (1987) *Biochem J* 248:167–171
43. Updyke TV, Nicolson GL (1986) *Methods Enzymol* 121:717–725
44. Chapman-Smith A, Cronan JE (1999) *Biomol Eng* 16:119–125
45. Qoronfle MW, Ren L, Emery D, Perr M, Kaboord B (2003) *J Biomed Biotechnol* 291–298
46. Sisson TH, Castro CW (1990) *J Immunol Methods* 127:215–220
47. Visser J, Van Dongen W, Strating M (1978) *FEBS Lett* 85:81–85
48. Cristea IM, Williams R, Chait BT, Rout MP (2005) *Mol Cell Prot* 4:1933–1941
49. Brymora A, Cousin MA, Roufogalis BD, Robinson PJ (2001) *Anal Biochem* 295:119–122
50. Absolom DR, van Oss CJ (1986) *CRC Crit Rev Immunol* 6:1–46
51. Diedrich G, Bangia N, Pan M, Cresswell P (2001) *J Immunol* 166:1703–1709
52. Plug MJ, Dijk J, Maassen A, Moller W (1992) *Eur J Biochem* 206:123–130
53. Figeys D, McBroom LD, Moran MF (2001) *Methods* 24:230–239
54. Cramer R, Saxton M, Barnouin K (2004) *Methods Mol Biol* 261:499–510
55. Havlis J, Shevchenko A (2004) *Anal Chem* 76:3029–3036
56. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevensky JR, Resing KA, Ahn NG (2005) *Mol Cell Prot* 4:1487–1502
57. Adachi J, Kumar C, Zhang Y, Mann M (2007) *Mol Cell Proteomics* 6: Epub ahead of print
58. Adamski M, Blackwell T, Menon R, Martens L, Hermjakob H, Taylor C, Omenn GS, States DJ (2005) *Proteomics* 5:3246–3261
59. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) *Nat Genet* 25:25–29
60. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D (1999) *Science* 285:751–753
61. Washburn MP, Koller A, Oshiro G, Ulaszek RR, Plouffe D, Deciu C, Winzeler E, Yater JR (2003) *Proc Natl Acad Sci USA* 100:3107–3112
62. Colland F, Jacq X, Trouplin V, Mougin C, Groizelaeau C, Hamburger A, Meil A, Wojcik J, Legrain P, Gauthier JM (2004) *Genome Res* 14:1324–1332
63. Howlett GJ, Minton AP, Rivas G (2006) *Curr Opin Chem Biol* 10:430–436
64. Mogridge J (2004) *Methods Mol Biol* 261:113–118
65. Wen J, Arakawa T, Philo JS (1996) *Anal Biochem* 240:155–166
66. Winzor DJ (2003) *J Biochem Biophys Methods* 56:15–52
67. Yeaman C (2003) *Methods* 30:198–206
68. Fields S, Song O (1989) *Nature* 340:245–246
69. Ito T, Ota K, Kubota H, Yamaguchi Y, Chiba T, Sakuraba K, Yoshida M (2002) *Mol Cell Prot* 1:561–566
70. Fernandez SM, Berlin RD (1976) *Nature* 264:411–415
71. Wallrabe H, Periasamy A (2005) *Curr Opin Biotechnol* 16:19–27
72. Aronheim A (2001) *Methods* 24:29–34
73. Deyev SM, Waibel R, Lebedenko E, Schubiger AP, Plueckthun A (2003) *Nat Biotechnol* 21:1486–1492
74. Hu J (2001) *Trends Microbiol* 9:219–222
75. Lippincott-Schwartz J, Patterson GH (2003) *Science* 300:87–91
76. Stagljar I, Fields S (2002) *Trends Biochem Sci* 27:559–563
77. Tsien RY (1998) *Annu Rev Biochem* 67:509–544
78. Zal T, Gascoigne NR (2004) *Curr Opin Immunol* 16:674–683
79. Ooi SL, Shoemaker DD, Boeke JD (2003) *Nat Genet* 35:277–286
80. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C (2001) *Science* 294:2364–2368
81. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, Recht J, Shales M, Ding H, Xu H, Han J, Ingvarsdottir K, Cheng B, Andrews B, Boone C, Berger SL, Hieter P, Zhang Z, Brown GW, Ingles CJ, Emili A, Allis CD, Toczycki DP, Weissman JS, Greenblatt JF, Krogan N (2007) *Nature* 446:806–810
82. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dimpfelfeld B, Edelmann A, Heurtier MA, Hoffmann V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G (2006) *Nature* 440:631–636
83. Kito K, Ota K, Fujita T, Ito T (2007) *J Proteome Res* 6:792–800
84. Dihazi GH, Sinz A (2003) *Rapid Commun Mass Spec* 17:2005–2014
85. Eyles SJ, Kaltashov IA (2004) *Methods* 34:88–99
86. Fancy DA (2000) *Curr Opin Cell Biol* 4:28–33
87. Weerasekera R, Schmitt-Ulms G (2006) *Biot Gen Eng Rev* 23:41–62
88. Carr S, Aebersold R, Baldwin M, Burlingame AL, Clauser K, Nesvizhskii A (2004) *Mol Cell Prot* 3:531–533
89. Bradshaw RA, Burlingame AL, Carr S, Aebersold R (2006) *Mol Cell Prot* 5:787–788
90. Mathivanan S, Periaswamy B, Gandhi TKB, Kandasamy K, Suresh S, Mohmood R, Ramachandra YL, Pandey A (2006) *BMC Bioinformatics* 7:S5–S19