DOI 10.1002/pmic.200600225 *Proteomics* 2006, *6,* 6158–6166

REVIEW

Mass spectrometry-based proteomic analysis of the epitope-tag affinity purified protein complexes in eukaryotes

Ing-Feng Chang

Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV, USA

Received: March 30, 2006
Revised: July 14, 2006
Accepted: August 12, 2006

In recent years, MS has been widely used to study protein complex in eukaryotes. The identification of interacting proteins of a particular target protein may help defining protein–protein interaction and proteins of unknown functions. To isolate protein complexes, high-speed ultracentrifugation, sucrose density-gradient centrifugation, and coimmunoprecipitation have been widely used. However, the probability of getting nonspecific binding is comparatively high. Alternatively, by use of one- or two-step (tandem affinity purification) epitope-tag affinity purification, protein complexes can be isolated by affinity or immunoaffinity columns. These epitope-tags include protein A, hexahistidine (His), c-Myc, hemaglutinin (HA), calmodulin-binding protein, FLAG, maltose-binding protein, Strep, *etc.* The isolated protein complex can then be subjected to protease (*i.e.*, trypsin) digestion followed by an MS analysis for protein identification. An example, the epitope-tag purification of the *Arabidopsis* cytosolic ribosomes, is addressed in this article to show the success of the application. Several representative protein complexes in eukaryotes been isolated and characterized by use of this approach are listed. In this review, the comparison among different tag systems, validation of interacting relationship, and choices of MS analysis method are addressed. The successful rate, advantages, limitations, and challenges of the epitope-tag purification are also discussed.

Keywords:
Epitope-tag affinity purification / Interactome / Mass spectrometry / Protein complex / Tandem affinity purification

1 Introduction

It is known that cellular processes, *i.e.*, DNA replication, RNA transcription, protein synthesis, protein degradation, and molecular signaling are accomplished by large protein complexes in the cells. In general, these molecular machineries are dynamic, and their individual components associate and disassociate constantly in response to intra- or intercellular signals. Some regulatory proteins can have multiple functions and associate with different protein complexes in a dynamic manner. Although there are hundreds of protein complexes in the cells, the detailed composition of these protein complexes is little known. Mapping the protein–protein interaction (interactome) may better define the organization of individual molecular machines and their dynamic nature in the cells. More importantly, mapping the interacting proteins may define the target proteins with unknown functions [1].

Correspondence: Dr. Ing-Feng Chang, Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV 89557, USA
E-mail: ifchang@unr.edu
Fax: +1-775-784-1650

Abbreviations: CBP, calmodulin-binding protein; **MBP,** maltose-binding protein; **RACK1,** receptor of activated kinase C; **TAP,** tandem affinity purification

2 High-speed sucrose density-gradient centrifugation and the coimmunoprecipitation of protein complex – conventional approaches

Since the growth of MS technology, MS-based proteomic analysis has been widely introduced into the study of system biology [2]. Several review articles have discussed the applications of this approach to study the interactome [1, 3–5]. Due to the completely sequenced genome of some model organisms (*i.e.*, human, *Arabidopsis*, and rice), protein identification is much easier than before. These studies vary from analysis of protein complexes to the organellar proteome [2, 6–13]. MS has been rapidly applied to study the organellar proteins, protein complexes, or protein–protein interactions in eukaryotes [6, 8, 11, 14, 15]. These organellar proteome or protein complexes include the 26S proteasome [16, 17], the cytosolic 80S ribosome [18–22], the mitochondria [23, 24], the splicesome [25], the chloroplast [26], the thylakoid membranes [27], the cell wall [28], the peroxisome [29], the Golgi [30], the nucleus [31], the plasma membrane [32–34], the mitochondrial ribosome [35–37], the vacuole [38], and the plastid ribosome [39, 40].

The MS-based proteomic approach is generally summarized into two major steps, (1) the isolation, fractionation and purification of proteins, and (2) the MS analyses of proteins [41]. A schematic flow chart of the MS-based proteomic analysis is shown in Fig. 1. Protein complexes can be isolated by use of the ultracentrifugation, the sucrose density-gradient centrifugation, affinity chromatography, coimmunoprecipitation, or epitope-tag affinity purification. After the protein complex is isolated, it can be subjected to either gel-based or gel-free separation followed by an MS analysis. For gel-free protein fractionation, the protein complex is subjected to in-solution protease (*i.e.*, trypsin) digestion followed by a 2-D-LC separation. Digested peptides are then subjected to MS/MS analysis using multidimensional protein identification technology (MudPIT) followed by a database search for protein identification [42]. As shown in Fig. 1, protein complex isolation serves the first step of the MS-based proteomic analysis. The strategy to isolate the protein complex is very critical to yield high quality protein purification and to identify *bona fide* interacting proteins. Two conventional approaches are discussed in this chapter. The first approach to isolate a protein complex is by use of the high-speed ultracentrifugation. With the ultracentrifuge, protein complexes are fractionated, on the basis of the differential sedimentation rate which is predominately determined by the molecular mass of the complex. The higher the mass is, the higher the sedimentation rate will be. For example, two independent groups utilized the ultracentrifuge to isolate the cytosolic ribosomes in *Arabidopsis* [21, 22]. With a density-gradient generated by different concentrations of sucrose, the protein complex was further separated and purified [21, 22]. The advantage of this approach is that it is an easy step. However, of nonspecific contaminants may be copurified in the same fraction with a similar sedimentation coefficient. This will be addressed in a later section of this paper.

The second approach to isolate protein complex is through a coimmunoprecipitation. By use of a specific antibody raised against the target proteins, the interacting proteins can be copurified due to the coimmunoprecipitation.

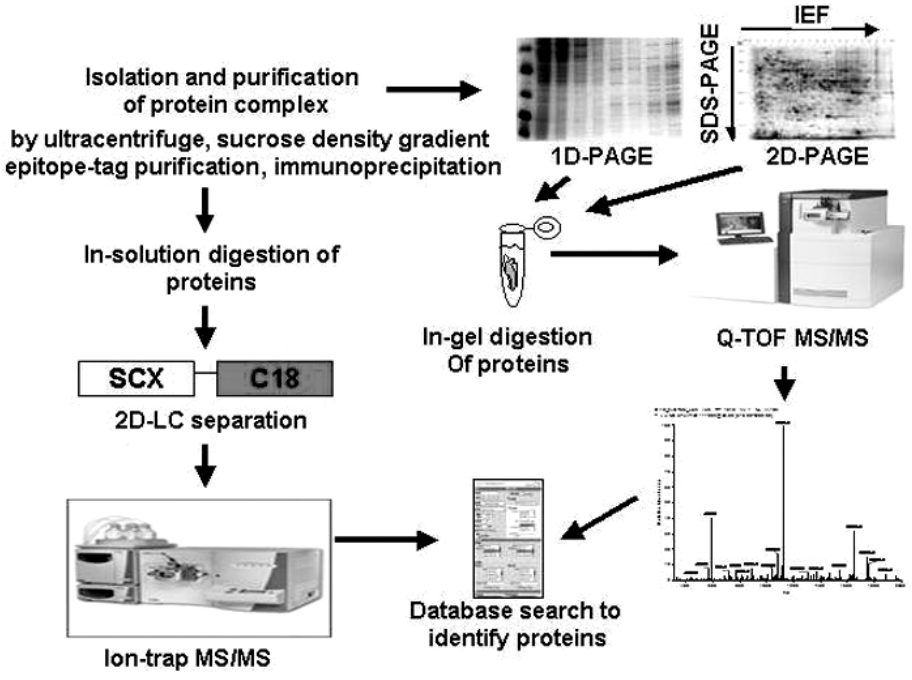

Figure 1. MS-based proteomic analysis of a protein complex. Protein complex is purified by ultracentrifugation, sucrose density-gradient ultracentrifugation, coimmunoprecipitation, or epitope-tag affinity purification (including TAP). The purified protein complex is separated by a 1-D or 2-D gel system depending on the complexity of the complex. The fractionated proteins are then digested in the gel by the protease (*i.e.*, trypsin) followed by an MS analysis. Alternatively, the protein complex can be digested in the solution followed by a 2-D-LC separation through an SCX and C18 column. Separated peptides are then ready for the MS analysis. After MS analysis, protein identification is reached by the use of an algorithm against a sequence database.

The immunoprecipitated protein complex can then be subjected to an MS analysis for protein identification. For example, 39 novel interacting proteins were coimmunoprecipitated with Hsp90 and identified by MS in human cells [43]. Although protein–protein interaction has been widely studied by use of the yeast-two hybrid system [1, 44], the coimmunoprecipitation provides an alternative strategy to discover interacting proteins. The first advantage of this approach is that it can isolate endogenous interacting proteins *in vivo*. The second advantage over yeast-two hybrid system is that it can resolve multiple interacting clients one at a time. Normally in a yeast-two hybrid system, positive interacting proteins to a certain bait protein can be multiple. Whether these proteins are physically interacting in the same complex remains questionable. Nevertheless, in the coimmunoprecipitation, client proteins are associated in the same complex and can be examined by a Western blot. The disadvantage of the coimmunoprecipitation is that the antibody may still have a chance to crossreact with other nonspecific proteins. In order to overcome this problem, a good antibody with high degree of specificity to the target protein is crucial to reduce the probability of nonspecific binding. However, generating a high quality antibody (*i.e.*, mAb) will also be more expensive.

3 The epitope-tag affinity purification of protein complex – modern approach

Since a proteomic study in yeast published in 1999 [45], the epitope-tag affinity purification has become widely used in isolating protein complexes *in vivo*. Several reviews have commented this approach [4, 46]. In 2002, two independent groups utilized the epitope-tag affinity purification to isolate protein complex in yeast followed by an MS analysis, and identified hundreds of protein complexes *in vivo* [47, 48]. In particular Gavin *et al.* [47] utilized tandem affinity purification (TAP), and identified 232 protein complexes. This novel approach, the MS-based proteomic analysis of the epitope-tag purified protein complex, opens a new window to study protein–protein interaction. Table 2 lists several representative protein complexes in eukaryotes that have been isolated and characterized by use of the epitope-tag affinity purification so far. This approach involves genetically engineering an oligopeptide, an epitope-tag (*i.e.*, hexahistidine (His_6), hemaglutinin (HA), Myc, GST, FLAG, calmodulin-binding protein (CBP), maltose-binding protein (MBP), or protein A) (Table 1) [49], fused solely into either the N- or C-terminus of the full-length coding region of the target protein in an expression vector (*i.e.*, Gateway vectors) [50] for protein overexpression in the cells (Fig. 2A). If there are more than two tags to be used, they can be fused together into either the N- or C-terminus of the target protein (Fig. 2B). This is called tandem affinity tags.

Purification of the tagged fusion protein can be achieved through either one- or two-step purification from the stably transformed tissues or cells by the affinity (*i.e.*, His, GST,

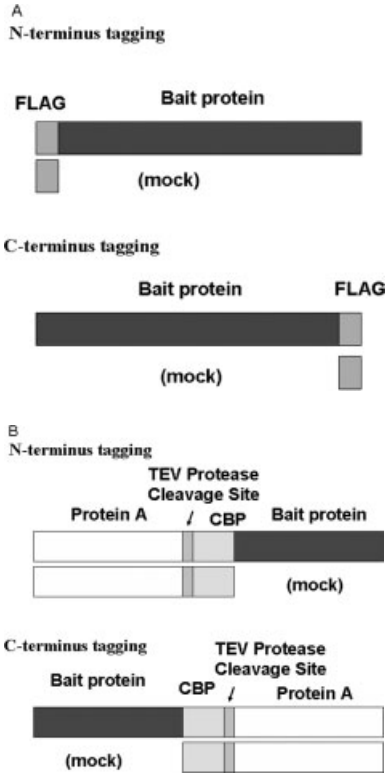

Figure 2. The tagging strategy for one- or two-step protein purification. The epitope tagging strategy is dependent on either (A) one-step affinity purification or (B) two-step affinity purification. The epitope tag can be fused solely or tandemly to the N- or C-terminus of the bait protein. For a TAP-tag construct, a TEV cleavage site is designed between the protein A and CBP tag.

Table 1. Commonly used epitope-tags for the MS-based proteomic study

Epitope tag	Amino acid length	Purification method	Binding agent	Elution agent	Kit available
His	6 aa	Affinity	Ni-NTA	Imidazole	Qiagen
Strep II	8 aa (WSHPQFEK)	Affinity	Strep-Tactin	Desthiobiotin	IBA
FLAG	8 aa (DYKDDDDK)	Immunoaffinity	Anti-FLAG Ab	FLAG	Sigma
Protein A	14 kDa	Affinity	IgG	TEV cleavage	GE Healthcare
MBP	40 kDa	Affinity	Amylose	Maltose	Biolabs
CBP	4 kDa	Affinity	Calmodulin	EGTA	GE Healthcare

protein A, and CBP) or immunoaffinity (*i.e.*, c-Myc, HA, and FLAG) columns. One-step affinity purification purifies bait protein with one tag fused to it. With the affinity column or tag-specific antibody, the tagged protein from the cell lysate can bind to the affinity column or antibody, and be eluted by the epitope-tag analog. Figure 3A shows the purification steps for a FLAG-tag protein as an example. By use of the anti-FLAG mAb, the protein complex of the target protein can be immunoaffinity purified. After the application of three-fold FLAG peptides to compete binding to the antibody, the FLAG-tag protein is eluted, and the interacting proteins can thus be copurified. For example, Tyers group utilized FLAG-tag technology to purify and identify protein complexes in yeast for high-throughput MS analysis [48]. In another example, Westermarck and coworkers utilized Strep-tag system for a one-step purification of the PP2A complex in mammalian cells [51]. In addition, mRNA can also be fused to a tag. In a novel experiment by Reed group, a pre-mRNA was fused to an MBP tag for one-step purification by maltose beads. The purified splicesome was subjected to an MS analysis and 58 novel components were identified [25].

Nevertheless, for fusion proteins with more than two tags on it, the TAP technology allows two-step purification of protein complexes [45, 52, 53]. The two-step affinity purification purifies bait fusion protein with tandem tags on it. A classic TAP-tag construct system includes protein A and CBP as tandem tags with a TEV protease cleavage site in between [52]. These tags can be fused tandemly to the target protein on either the N- or C-terminus, and the fusion proteins can be overexpressed in the transformed cells. Once cell lysate go through the IgG column, protein A binds to the column. After the TEV protease cleavages, the CBP fusion portion is released. The flow-through then goes through a calmodulin column for CBP binding. After the addition of EGTA, the tagged protein and the interacting proteins can be coeluted (Fig. 3B). This system has been widely used and successfully purified several protein complexes in eukaryotes.

To ensure the stability of the protein complex isolated for MS analysis, crosslinkers may be an additional help. These crosslinkers chemically crosslink interacting proteins to the bait protein. Adding the crosslinkers may stabilize the transient or weak interactors. Fromm and coworkers [54] performed a TAP-tag protein purification of GVG transcription factor in *Arabidopsis*, and showed improved recovery of the interacting proteins by an additional crosslinking step. They successfully recovered and identified two interacting proteins, Hsp70 and Hsp90. In addition, Kaiser group crosslinked the ubiquitinated proteins in the Skp1 protein complex, and identified a handful of protein targets with ubiquitinated sites at lysine residue [55].

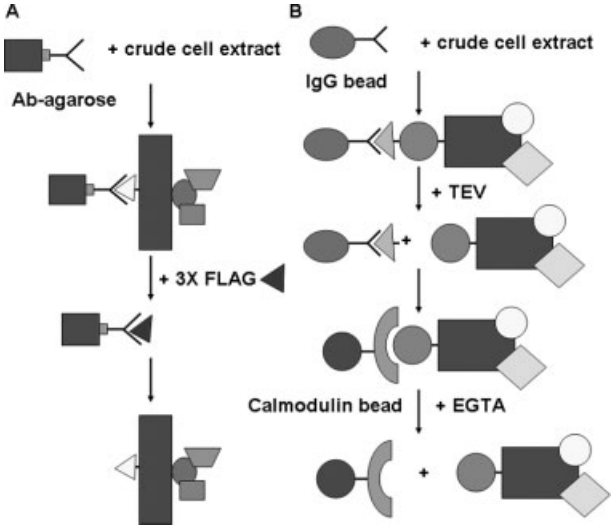

Figure 3. One- or two-step affinity purification of protein complex. The tagged protein complex is purified by either a (A) one-step or (B) two-step affinity purification using affinity columns. For the one-step purification (FLAG-tag as example), the purified cell extract binds to the anti-FLAG mAb. After the application of three-fold of FLAG peptides to compete binding, the target protein and client proteins are released. For the two-step purification (TAP), the purified cell extract binds to the IgG column first. TEV protease is then applied to cleave the TEV cleavage site, which leaves the protein A fragment bound. The remaining fragment then binds to the calmodulin column. After the application of EGTA, the target protein containing CBP and client proteins are released.

4 Comparison of the efficiency of the different tag systems

Although the efficiency of purification of these tag systems was compared and it appeared that the Strep-tag is better [56], frankly speaking there is no best tag system. Each tag has its advantage and disadvantage, depending on its cost, size, efficiency, and the host system. The general rule is that the tag should not interfere with the biological functions of the proteins to be tagged. Smaller tag such as FLAG and Strep might be a better choice, but it is not necessarily true. For example, the Na^+ transport activity of *Arabidopsis* HKT was inhibited by the addition of a peptide tag at the C-terminus [57]. Although protein A and CBP are the commonly used TAP-tag system (classic TAP-tag), the combination and number of tags to be used can be variable. As a matter of fact, accumulating studies suggest that CBP tag might not always work. Some modified tag systems were designed instead to give better efficiency of purification. These tags include FLAG [58], Myc [59], biotin [60], and protein C [61]. Either two of them can be used as a possible combination for tandem tag system. For example, Deng *et al*. [59] utilized three-tag (protein A, 6xHis, and 9xMyc) system (Table 2) to purify COP9 signalsome in *Arabidopsis* by three steps. A modified TEV protease was engineered and utilized to yield more specific cleavage. They also tested and confirmed that most of the purified protein complexes were biologically functional by a complementation assay. Similarly, Futcher and coworkers [62] performed a three-step purification of

Table 2. Representative epitope-tag affinity purified protein complexes in eukaryotes

Protein complex	Bait protein	Epitope tag utilized	Organism	Proteins identified	Functional pathway	Reference
14-3-3	14-3-3 gamma	FLAG	Human cells	170	Cytoskeleton	[66]
14-3-3	14-3-3 sigma	Protein A, CBP	Human cells	117	Cancer cell cycle	[67]
PP2A	PR65 alpha subunit	Strep	Mammalian cells	4	Dephosphorylation	[51]
26S proteasome[a)]	Rpn11, Rpt5	His, biotin	Yeast	64	Protein degradation	[17]
26S proteasome	multiple	Myc or HA	Yeast	>24	Protein degradation	[16]
Cyclin-CDK	Clb2	CBP, His, HA	Yeast	31	Cell cycle	[62]
mRNP	Scp160p	FLAG	Yeast	2	mRNA metabolism	[76]
Polyribosome	r-Protein L25	FLAG	Yeast	2	Protein synthesis	[19]
Polyribosome	r-Protein L18	FLAG	*Arabidopsis*	>50	Protein synthesis	[75]
Glucocorticoid receptor	r-GVG transcriptional factor	Protein A, CBP	*Arabidopsis*	2	Steroid-dependent gene induction	[54]
COP9 signalsome	CSN3	His, Myc, protein A	*Arabidopsis*	2	Proteolytic regulation	[59]
Insulin growth factor-like	IRS-1	Protein A, CBP, biotin	Mouse cells	18	Insulin receptor substrate signaling	[60]
RNA polymerase	RNA polymerase II	Protein A, CBP	Human cells	13	Transcription	[77]
Swi/Snf chromatin-remodeling	snf6	Protein A, CBP	Yeast	12	Chromatin remodeling	[78]
Histone acetyl transferase	Ada2	Protein A, CBP	Yeast	18	Chromatin remodeling	[78]
SMAD3	SMAD3	Protein A, FLAG	Human cells	1	TGF receptor signaling	[58]
SnRNP	Sun71p	Protein A, CBP	Yeast	12	RNA splicing	[45]
HNF4 nuclear receptor	dHNF4	FLAG, His	Drosophila	2	Lipid metabolism	[63]
Many	Multiple	Protein A, CBP	Yeast	Many	Multiple	[47]
Many	Multiple	FLAG	Yeast	Many	Multiple	[48]
Many	Multiple	Protein A, CBP	Rice	Many	Kinase signaling	[79]
Many	Multiple	Protein A, CBP	Yeast	Many	RNA metabolism	[52]
TATA box binding	TBP	Biotin	Rice	86	mRNA transcription	[80]
Translation initiation complex	EIF 5A	Protein A, CBP	Yeast	20	Protein synthesis	[81]
Poly(A) polymerase	Trf5p	Protein A, CBP	Yeast	2	Polyadenylation	[82]
Pre-mRNA retention and splicing	Sun17p	Protein A, CBP	Yeast	2	Pre-mRNA splicing	[83]
Many	38 proteins	Protein A, CBP	Yeast	220	Multiple	[84]
Many	21 proteins	His, Myc, His	Yeast	Many	Multiple	[42]
Many	4562 proteins	Protein A, CBP	Yeast	2708	Multiple	[65]
Splicesome	Adml pre-mRNA (RNA)	MBP	Human cells	145	Pre-mRNA splicing	[25]
Ribosome	Multiple	Protein A, CBP	Yeast	77	mRNA translation	[85]

a) Quantative proteomic analysis was performed in the study.

CBP–$(His)_6$–$(HA)_3$-tag protein complex in yeast. Graumann *et al*. [42] compared the result of using classic TAP-tag to using their TAP-tag ($(Myc)_9$ and $(His)_3$), and showed that their strategy yielded higher coverage of the identified proteins. They also identified more interacting proteins for some of the tagged proteins than Ho *et al*. [48] previously did. Yang *et al*. [63] also compared the purification difference between protein A, CBP and FLAG, His. He found FLAG, His as TAP-tag yielded better purification. In addition, Baserga and coworkers [60] introduced additional tag, biotin, other than protein A and CBP as tag combination in the purification of several protein complexes in mammalian cells. On the other hand, even using a single tag may work the same as TAP-tag does. A study in purifying plant calcium-dependent protein kinase showed that using Strep II-tag worked as well as TAP-tag did [64].

5 Successful rates in the purification and identification of interacting proteins

In 2002, Gavin *et al*. [47], utilized TAP-tag affinity purification, and identified 232 protein complexes in yeast. The successful rate of the purification is 67 and 78% for protein identification. A different group utilized TAP-tag affinity purification in yeast, and had a 52% success rate of the purification [65]. In addition, Fromm and coworkers [79] utilized the same approach and identified 23 kinase protein complexes in rice. The successful rate of the purification is 95 and 58% for protein identification. Therefore, the successful rate of TAP-tag purification is at least 50% so far in eukaryotes, which means that still about less than half of the gene construct cannot be expressed and purified. Even though the protein complexes are successfully purified, the MS identification of these complexes is not 100%. These statistic data suggests the limitations of the complex purification and protein identification, which will be discussed in later chapters.

6 Choices of MS analysis method

The choices among different MS analyses can be dependent on the complexity of the interactors. If the interactors are not many (less than 50–100), a gel-based MS analysis is suitable. The gel image can provide quantitative information of how interactors are enriched by the affinity purification. The stoichiometry of the binding relationship can be deduced by the protein band intensity. This will help discerning the true interactors or the contaminants. After finding protein bands or spots enriched in the gel samples from the affinity purified fractions, an in-gel digestion followed by an MS analysis can identify the interactors. With limited antibody available, the confirmation of the protein identification can be achieved through a Western blotting. Nevertheless if there are numerous interactors (more than 100), a nongel-based MS analysis is probably proper. Due to the loading capacity and detection limitation of the 1-D or 2-D gels by staining reagents, these interactors may not all be resolved and detected. Especially low abundant proteins are harder to be detected and identified. Therefore an MudPIT analysis [42] following an in-solution digestion of the proteins may give better identifications. For example, 14-3-3 proteins potentially bind more than a hundred proteins [66, 67]. It is easily to miss the identification of any low abundant proteins by use of the gel-based method. Using a gel-free system [66, 67] will probably give more coverage of the proteome. In this way, a quantitative proteomic analysis is crucial to determine the enrichment of the interactors.

7 Validation of the interaction

The ways to validate the interacting relationship vary. A very common strategy is the coimmunoprecipitation. Specific antibody against the bait protein can be used to precipitate the protein complex. Once the candidate interactors are identified, another antibody against the candidate interactor can be used to detect copurification. If the interacting relationship is real, the interactor will be detected in the same protein complex. In addition, a reciprocal coimmunoprecipitation can be performed by use of the antibody against the client protein. If the original bait protein is in the complex, this provides additional evidence of interaction. A third approach to validate the interacting relationship is to make a fusion protein and test for binding *in vitro*. By engineering a fusion protein containing an affinity tag fused to the coding region of the interactor, the fusion protein can be overexpressed in the *Escherichia coli*. Fusion protein can be affinity purified and tested for binding through a column containing the bait protein. If the protein stays to the column, the interacting relationship is real.

8 Advantages, limitations, and challenges of the epitope-tag purification technique

The epitope-tag purification provides multiple advantages. The first advantage is that much purer protein complex purification can be achieved than the conventional approaches. For example, after TAP-tag purification, purified protein complexes are much cleaner than just by one-step purification as described in [45]. Kaneko et al. [68] also showed that in their TAP-tag purification, the protein gel looked cleaner after their second-step purification than only one-step purification. A mock construct (control) with only tags but without ORF of the target protein to be fused can be engineered. Whatever binds to the mock fusion protein or the beads can be the potential contaminants, which should be eliminated from the binding protein hit list. The second advantage is that the purification step is more efficient than the conventional protein purification methods. Most of the tag purifica-

tion methods or kits are commercially available (Table 1). This provides a uniform system for tagging and purifying different protein for the comparison one at a time. The antibody against the tag is also commercially available. A Western blot experiment to determine the enrichment of the protein purification is handy.

However, the epitope-tag purification method has limitations. The first limitation of this modern approach is that protein purification yield is bait-expression dependent. If the expression of the tag-protein is low, protein purification may not be efficient to get enough proteins for MS analysis. Instead of using endogenous promoter, a 35S promoter can be another choice to drive the gene expression. Having a good reporter (*i.e.*, green florescence protein) fused with the fusion protein is also helpful. The fluorescence character of the reporter can screen for high expression transgenic lines. In addition, a complementation analysis transferring overexpressed genes in the mutant lines may also help determining the expressed lines [59]. These will make sure the tag-protein to be purified is biologically functional. The second limitation is that the protein purification is topology dependent. In order to be accessible to the affinity column or antibody, the epitope tag has to be exposed out of the protein complex. If the tag-protein happens to be buried inside the complex, the protein complex purification will not be successful. In addition, PTMs of either the N- or C-terminus might disrupt the subcellular localization, which affect the real client protein binding. For example, the myristoylation and palmitoylation of the N-terminus of the calcium-dependent protein kinase is essential for the membrane localization in plants [69]. If the N-terminus is tagged, the membrane localization may be disrupted and the interacting proteins may be totally different. In Deng and coworkers [59] study, they also showed differential expression between the N- and C-terminus tagged lines, which suggested the tagging position does matter. The third limitation is the possible competition between the tag-protein and the endogenous target protein for binding to the interacting proteins. If a large gene family encodes the tagged protein, the competition among different family members (isoforms) for binding is extremely unavoidable. This competition can be abolished if the endogenous target protein is genetically knockout or knockdown. For example, Forler *et al*. [70] introduced RNAi to knockout endogenous gene expression to avoid competition with tagged protein. The fourth limitation is the conformational change by the epitope-tag. The tag may also change the charge state of the tag-protein, which makes the interacting protein binding difficult. For example, addition of a peptide tag at the C-terminus inhibited the Na^+ transport activity of *Arabidopsis* HKT [57].

The first challenge for the epitope-tag purification of protein complex is to get high-throughput performance. So far this technique is still time-consuming and requires lots of human effort. An exception is a high-throughput analysis of TAP-tag collections by use of a 96-well plate screening in yeast [71]. The second challenge is to purify cell type-specific or organelle-specific protein complex. Although it provides a possibility to purify cell-type specific protein complex by engineering protein expression driven by a cell-type specific promoter or endogenous promoter, very few studies have shown successful purification of cell type-specific protein complex. Even though membrane proteins can be epitope-tag or TAP-tag affinity purified solely [64, 72], the isolation of membrane protein interacting proteins is still a challenge. A successful case is that Jones and coworkers [73] have utilized a TAP-tag purification approach and isolated a Cf-9 protein complex (~420 kDa) in tobacco. By crosslinking the bait protein to the interacting clients may be a way to purify the intact membrane protein complex. The third challenge is to perform a quantitative proteomics. So far only limited studies used this approach (Table 2). The quantitative analysis will allow characterizing the protein enrichment after complex purification to avoid false positives.

9 Application of the epitope-tag affinity purification of protein complex – eukaryotic cytosolic ribosomes as an example

The cytosolic ribosome is the major molecular machinery for protein translation. The eukaryotic cytosolic ribosome (80S) is composed of four ribosomal RNAs, and about 80 different types of ribosomal proteins (r-proteins) on two (40S and 60S) subunits. In 2000 Bailey-Serres and coworkers [74] identified 249 genes encoding 80 possible cytosolic r-proteins in the *Arabidopsis* genome. Cytosolic ribosomes were then isolated from *Arabidopsis* suspension cultures by ultracentrifugation. r-Proteins were fractionated by 2-D gels followed by the in-solution protease digestion and MS analysis. Cytosolic ribosomes were also isolated from *Arabidopsis* suspension cultures by the sucrose gradient centrifugation as well. In summary, this proteomic analysis of the ultracentrifugation-based purification of ribosomes yielded identification of 74 cytosolic r-proteins in *Arabidopsis* [22]. In addition to the expected ribosome complex present in the ribosome-containing fractions, several nonribosomal proteins were also identified which include a receptor of activated kinase C (RACK1), vacuolar, mitochondrial, and several proteasome subunit proteins [22]. Interestingly, a WD (Trp-Asp)-domain containing protein RACK1 was found to be associated with the 40S subunit and polyribosomes in *Arabidopsis* culture cells.

To determine whether these nonribosomal proteins are true interactors or just contaminants, the cytosolic ribosome complex was purified by the epitope-tag affinity purification method in stead. A FLAG-tag strategy was designed by Bailey-Serres group to affinity-purify ribosomes from rosette leaves of a transgenic *Arabidopsis* line. The DNA sequence of the FLAG-tag was cloned in fusion with an r-protein L18 at the N-terminus in a constitutive CaMV 35S promoter expression cassette, and the ribosome complex was isolated

from leaf extracts of plants expressing the tagged r-protein L18 [75]. RACK1 was further confirmed to be associated with polysomes in the FLAG-tag purified polysomal fractions in *Arabidopsis*. In addition to RACK1, some RNA-binding proteins were identified as well by Zanetti. However, the mitochondrial, vacuolar, and proteasome proteins previously identified in the ribosome pellet [22] were all absent in the epitope-tag purified polysomes [75]. It appeared that without the epitope-tag affinity purification approach, RACK1 could never be identified and confirmed to be a ribosome-interacting protein merely by a genome surfing or a microarray clustering analysis. This demonstrates that the epitope-tag affinity purification provides a good tool and a better chance to identify *bona fide* interacting proteins than the conventional ultracentrifugation approach.

10 Conclusions

As introduced above, the epitope-tag affinity purification has been widely used in the MS-based proteomic analysis. This technology allows one-step or multiple-step purification of the protein complex *in vivo* through the affinity columns. This technology provides efficient purification to yield purer proteins for MS analysis, but reduce the chances to identify nonspecific binding proteins. Once all the protein complexes in the cells are characterized, the protein interacting network can be constructed as done previously in the yeast model [47, 48]. This may elucidate the dynamics and functions of known and unknown proteins. However, this technology is still not high-throughput. Many optimizations can still be performed in the future to improve this system, including the use of the MS-compatible crosslink technology, quantitative proteomic analysis, and cell-type specific complex purification.

The authors greatly appreciate Dr. Kangling Zhang (University of California, Riverside) for his critical review of the manuscript.

11 References

[1] Cusick, M. E., Klitgord, N., Vidal, M., Hill, D. E., *Hum. Mol. Genet.* 2005, *14*, R171–181.

[2] Aebersold, R., Mann, M., *Nature* 2003, *422*, 198–207.

[3] Droit, A., Poirier, G. G., Hunter, J. M., *J. Mol. Endocrinol.* 2005, *34*, 263–280.

[4] Gingras, A. C., Aebersold, R., Raught, B., *J. Physiol.* 2005, *563*, 11–21

[5] Monti, M., Orrù, S., Pagnozzi, D., Pucci, P., *Biosci. Rep.* 2005, *25*, 45–56.

[6] Yates, J. R., *J. Mass Spectrom.* 1998, *33*, 1–19.

[7] Li, J., Assmann, S. M., *Plant Physiol.* 2000, *123*, 807–809.

[8] Pandey, A., Mann, M., *Nature* 2000, *405*, 837–846.

[9] Aebersold, R., Goodlett, D. R., *Chem. Rev.* 2001, *101*, 269–295.

[10] Godovac-Zimmermann, J., Brown, L. R., *Mass Spectrom. Rev.* 2001, *20*, 1–57.

[11] Mann, M., Hendrickson, R. C., Pandey, A., *Annu. Rev. Biochem.* 2001, *70*, 437–473.

[12] Patterson, S. D., Aebersold, R., Goodlett, D. R., in: Pennington, S. R., Dunn, M. J. (Eds.), *Proteomics: From Protein Sequence to Function*, Springer-Verlag New York Inc., NY 2001, pp. 87–130.

[13] Patterson, S. D., Aebersold, R. H., *Nature* 2003, *33*, 311–323.

[14] Figeys, D., McBroom, L. D., Moran, M. F., *Methods* 2001, *24*, 230–239.

[15] Dreger, M., *Eur. J. Biochem.* 2003, *270*, 589–599.

[16] Verma, R., Chen, S., Feldman, R., Schieltz, D. *et al.*, *Mol. Biol. Cell* 2000, *11*, 3425–3439.

[17] Guerrero, C., Tagwerker, C., Kaiser, P., Huang, L., *Mol. Cell. Proteomics* 2006, *5*, 366–378.

[18] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. *et al.*, *Nat. Biotechnol.* 1999, *17*, 676–682.

[19] Inada, T., Winstall, E., Tarun, S. Z., Jr., Yates, J. R. *et al.*, *RNA* 2002, *8*, 948–958.

[20] Manuell, A., Yamaguchi, K., Haynes, P. A., Milligan, R. A., Mayfield, S. P., *J. Mol. Biol.* 2005, *351*, 266–279.

[21] Giavalisco, P., Wilson, D., Kreitler, T., Lehrah, H. *et al.*, *Plant Mol. Biol.* 2005, *57*, 577–591.

[22] Chang, I. F., Szick-Miranda, K., Pan, S., Bailey-Serres, J., *Plant Physiol.* 2005, *137*, 848–862.

[23] Taylor, S. W., Fahy, E., Zhang, B., Glenn, G. M. *et al.*, *Nat. Biotechnol.* 2003, *21*, 281–286.

[24] Kruft, V., Eubel, H., Jänsch, L., Werhahn, W., Braun, H. P., *Plant Physiol.* 2001, *127*, 1694–1710.

[25] Zhou, Z., Licklider, L. J., Gygi, S. P., Reed, R., *Nature* 2002, *419*, 182–185.

[26] Peltier, J. B., Friso, G., Kalume, D. E., Roepstorff, P. *et al.*, *Plant Cell* 2000, *12*, 319–341.

[27] Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B. *et al.*, *Plant Cell* 2004, *16*, 478–499.

[28] Wang, S. B., Hu, Q., Sommerfeld, M., Chen, F., *Proteomics* 2004, *4*, 692–708.

[29] Fukao, Y., Hayashi, M., Nishimura, M., *Plant Cell Physiol.* 2002, *43*, 689–696.

[30] Wu, C. C., MacCoss, M. J., Mardones, G., Finnigan, C. *et al.*, *Mol. Biol. Cell* 2004, *15*, 2907–2919.

[31] Bae, M. S., Cho, E. J., Choi, E. Y., Park, O. K., *Plant J.* 2003, *36*, 652–663.

[32] Marmagne, A., Rouet, M. A., Ferro, M., Rolland, N. *et al.*, *Mol. Cell. Proteomics* 2004, *3*, 675–691.

[33] Nielsen, P. A., Olsen, J. V., Podtelejnikov, A. V., Andersen, J. R. *et al.*, *Mol. Cell. Proteomics* 2005, *4*, 402–408.

[34] Nelson, C. J., Hegeman, A. D., Harms, A. C., Sussman, M. R., *Mol. Cell. Proteomics* 2006, *5*, 1382–1395.

[35] Koc, E. C., Burkhart, W., Blackburn, K., Moseley, A. *et al.*, *J. Biol. Chem.* 2000, *276*, 32585–32591.

[36] Saveanu, C., Fromont-Racine, M., Harington, A., Richard, F. *et al.*, *J. Biol. Chem.* 2001, *276*, 15861–15867.

[37] Koc, E. C., Burkhart, W., Blackburn, K., Moyer, M. B. *et al.*, *J. Biol. Chem.* 2001, *276*, 43958–43969.

[38] Carter, C., Pan, S., Zouhar, J., Avila, E. L. *et al.*, *Plant Cell* 2004, *16*, 3285–3303.

[39] Yamaguchi, K., von Konblauch, K., Subramanian, A. R., *J. Biol. Chem.* 2000, *275*, 28455–28465.

[40] Yamaguchi, K., Subramanian, A. R., *Eur. J. Biochem.* 2003, *270*, 190–205.

[41] Griffin, T. J., Aebersold, R., *J. Biol. Chem.* 2001, *49*, 45497–45500.

[42] Graumann, J., Dunipace, L. A., Seol, J. H., McDonald, W. H. *et al.*, *Mol. Cell. Proteomics* 2004, *3*, 226–237.

[43] Falsone, S. F., Gesslbauer, B., Tirk, F., Piccinini, A. M., Kungl, A. J., *FEBS L.* 2005, *579*, 6350–6354.

[44] Toby, G. G., Golemis, E. A., *Methods* 2001, *24*, 201–217.

[45] Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M. *et al.*, *Nat. Biotechnol.* 1999, *17*, 1030–1032.

[46] Bauer, A., Kuster, B., *Eur. J. Biochem.* 2003, *270*, 570–578.

[47] Gavin, A. C., Bosche, M., Krause, R., Grandi, P. *et al.*, *Nature* 2002, *415*, 141–147.

[48] Ho, Y., Gruhler, A., Hellbut, A., Bader, G. D. *et al.*, *Nature* 2002, *415*, 180–183.

[49] Stevens, R. C., *Structure* 2000, *8*, R177–R185.

[50] Earley, K. W., Haag, J. R., Pontes, O., Opper, K. *et al.*, *Plant J.* 2006, *45*, 616–629.

[51] Junttila, M. R., Saarinen, S., Schmidt, T., Kast, J., Westermarck, J., *Proteomics* 2005, *5*, 1199–1203.

[52] Puig, O., Caspary, F., Rigaut, G., Rutz, B. *et al.*, *Methods* 2001, *24*, 218–229.

[53] Kumar, A., Snyder, M., *Nature* 2002, *415*, 123–124.

[54] Rohila, J. S., Chen, M., Cerny, R., Fromm, M. E., *Plant J.* 2004, *38*, 172–181.

[55] Tagwerker, C., Flick, K., Cui, M., Guerrero, C. *et al.*, *Mol. Cell. Proteomics* 2006, *5*, 737–748.

[56] Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., Tan, S., *Protein Expr. Purif.* 2005, *41*, 98–105.

[57] Kato, Y., Hazama, A., Yamaguchi, M., Uozumi, N., *Biosci. Biotechnol. Biochem.* 2003, *67*, 2291–2293.

[58] Knuesel, M., Wan, Y., Xiao, Z., Holinger, E. *et al.*, *Mol. Cell. Proteomics* 2003, *2*, 1225–1233.

[59] Rubio, V., Shen, Y., Saijo, Y., Liu, Y. *et al.*, *Plant J.* 2005, *41*, 767–778.

[60] Drakas, R., Prisco, M., Baserga, R., *Proteomics* 2005, *5*, 132–137.

[61] Schimanski, B., Nguyen, T. N., Günzl, A., *Eukaryot. Cell* 2005, *4*, 1942–1950.

[62] Honey, S., Schneider, B. L., Schieltz, D. M., Yates, J. R., Futcher, B., *Nucleic Acids Res.* 2001, *29*, E24.

[63] Yang, P., Sampson, H. M., Krause, H. M., *Proteomics* 2006, *6*, 927–935.

[64] Witte, C. P., Noël, L. D., Gielbert, J., Parker, J. E., Romeis, T., *Plant Mol. Biol.* 2004, *55*, 135–147.

[65] Krogan, N. J., Cagney, G., Yu, H., Zhong, G. *et al.*, *Nature* 2006, *440*, 637–643.

[66] Jin, J., Smith, D., Stark, C., Wells, C. D. *et al.*, *Curr. Biol.* 2004, *14*, 1436–1450.

[67] Benzinger, A., Muster, N., Koch, H., Yates, J. R., Hermeking, H., *Mol. Cell. Proteomics* 2005, *4*, 785–795.

[68] Kaneko, A., Umeyama, T., Hanaoka, N., Monk, B. C. *et al.*, *Yeast* 2004, *21*, 1025–1033.

[69] Martin, M. L., Busconi, L., *Plant J.* 2001, *24*, 429–435.

[70] Forler, D., Köcher, T., Rode, M., Gentzel, M. *et al.*, *Nat. Biotechnol.* 2003, *21*, 89–92.

[71] Howson, R., Huh, W., Ghaemmaghami, S., Falvo, J. V. *et al.*, *Comp. Funct. Genomics* 2005, *6*, 2–16.

[72] Li, Q., Dai, X. Q., Shen, P. Y., Cantiello, H. F. *et al.*, *FEBS L.* 2004, *576*, 231–236.

[73] Rivas, S., Romeis, T., Jones, J. D. G., *Plant Cell* 2002, *14*, 689–702.

[74] Barakat, A., Szick-Miranda, K., Chang, I. F., Guyot, R. *et al.*, *Plant Physiol.* 2001, *127*, 398–415.

[75] Zanetti, M. E., Chang, I. F., Gong, F., Galbraith, D. W., Bailey-Serres, J., *Plant Physiol.* 2005, *138*, 624–635.

[76] Lang, B. D., Fridovich-Keil, J. L., *Nucleic Acids Res.* 2000, *28*, 1576–1584.

[77] Jeronimo, C., Langelier, M. F., Zeghouf, M., Cojocaru, M. *et al.*, *Mol. Cell. Biol.* 2004, *24*, 7043–7058.

[78] Lee, K. K., Prochasson, P., Florens, L., Swanson, S. K. *et al.*, *Biochem. Soc. Trans.* 2004, *32*, 899–903.

[79] Rohila, J. R., Chen, M., Chen, S., Chen, J. *et al.*, *Plant J.* 2006, *46*, 1–13.

[80] Zhong, J., Haynes, P. A., Zhang, S., Yang, X. *et al.*, *J. Proteome Res.* 2003, *2*, 514–522.

[81] Jao, D. L. E., Chen, K. U., *J. Cell. Biochem.* 2006, *97*, 583–598.

[82] Houseley, J., Tollervey, D., *EMBO Rep.* 2006, *7*, 205–211.

[83] Dziembowski, A., Ventura, A. P., Rutz, B., Caspary, F. *et al.*, *EMBO J.* 2004, *23*, 4847–4856.

[84] Shevchenko, A., Schaft, D., Roguev, A., Pijnappel, W. W. M. P. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 204–212.

[85] Fleischer, T. C., Weaver, C. M., McAfee, K. J., Jennings, J. L., Link, A. J., *Gene Dev.* 2006, *20*, 1294–1307.

