

# Deciphering Protein Complexes and Protein Interaction Networks by Tandem Affinity Purification and Mass Spectrometry

ANALYTICAL PERSPECTIVE\*

Anna Shevchenko‡, Daniel Schaft§, Assen Roguev§, W. W. M. Pim Pijnappel§, A. Francis Stewart§¶, and Andrej Shevchenko‡||

**We employed a combination of tandem affinity purification and mass spectrometry for deciphering protein complexes and the protein interaction network in budding yeast. 53 genes were epitope-tagged, and their interaction partners were isolated by two-step immunoaffinity chromatography from whole cell lysates. 38 baits pulled down a total of 220 interaction partners, which are members of 19 functionally distinct protein complexes. We identified four proteins shared between complexes of different functionality thus charting segments of a protein interaction network. Concordance with the results of genome-wide two-hybrid screening was poor (14% of identified interactors overlapped) suggesting that the two approaches may provide complementary views on physical interactions within the proteome. *Molecular & Cellular Proteomics* 1:204–212, 2002.**

With the completion of genome sequencing for 60 prokaryotic organisms (see [www.tigr.org/tdb/mdb/mdbcomplete.html](http://www.tigr.org/tdb/mdb/mdbcomplete.html)) and several important eukaryotic organisms (1–3), including a draft of the human genome (4, 5), the challenging task now is to decipher relationships between individual genes and to understand the molecular organization of cellular networks. This requires the documentation of stable protein complexes, which are the functional units of cellular molecular machinery (6). Beyond this, some members of a protein complex may interact physically with other protein assemblies and link the complexes into a network (7). Further steps involve the identification of post-translational modifications that regulate protein engagement and quantitative description of their regulatory dynamics (8, 9).

The availability of the budding yeast genome (1) stimulated efforts toward global mapping of a comprehensive “circuit diagram” of physical and functional protein interactions

through bioinformatics (10), genome-wide two-hybrid screening (11, 12), mRNA (13), and protein (14) arrays. Although even a preliminary sketch of a global gene interaction network is a remarkable achievement, these experiments revealed important limitations of the technologies. First, they did not document protein complexes but rather inferred their content from interaction data sets, which are, problematically, contaminated with false positives. In particular, two-hybrid screening only records interactions between pairs of genes and misses interactions stabilized by more than two partners. Thus it appears that there is no substitute for authentic biochemical characterization of protein complexes purified from the original host. Furthermore, native protein assemblages are formed in a complex milieu, which is regulated by folding, modification, limited proteolysis, transportation to specific cellular compartments, and assembly with non-proteinous co-factors such as RNAs, nucleotides, metal cations, etc.

Protein tagging presents a generic approach for the analysis of native protein complexes. The tagged protein is affinity-purified from a whole cell lysate, together with associated proteins, which are subsequently characterized by mass spectrometry (15). Because it is relatively straightforward to fuse affinity tags with target proteins in the budding yeast, the approach was successfully applied to the characterization of numerous assemblages of various molecular weight, localization, and biological function (reviewed in Ref. 16). Further developments were focused on improving protein identification by multidimensional liquid chromatography-tandem mass spectrometry (17, 18) and of protein-tagging methodology (19). The tandem affinity purification (TAP)<sup>1</sup> method (20, 21) utilizes two affinity tags spaced by a cleavage site of tobacco etch virus (TEV) proteinase. Compared with other epitope tags such as *myc*– or *HA*–, TAP gives better yields of affinity-purified proteins, along with lower background of non-specifically associated proteins (22, 23).

From the ‡Max Planck Institute of Molecular Cell Biology and Genetics (MPI CBG) and §Genomics, Technische Universität Dresden, c/o MPI CBG, Pfotenhauerstrasse 108, 01307 Dresden, Germany

Received, January 4, 2002, and in revised form, January 24, 2002  
Published, MCP Papers in Press, February 1, 2002, DOI 10.1074/mcp.M200005-MCP200

<sup>1</sup> The abbreviations used are: TAP, tandem affinity purification; CBI, codon bias index; IP, immunoprecipitation; MALDI MS, matrix-assisted laser desorption/ionization mass spectrometry; M(ox), methionine sulfoxide; nanoES MS, nanoelectrospray mass spectrometry; ORF, open reading frame; TEV, tobacco etch virus; 2HY, two-hybrid screening.

Two ways to use epitope tagging, immunoaffinity chromatography, and mass spectrometry in the analysis of protein complexes are evident. In one strategy, a large number of baits are processed in parallel by an established high throughput protein purification and identification routine. The biological significance of identified interactions is evaluated later and only for a selection of baits that yielded the most interesting patterns of associated proteins.

Alternatively, a sequential approach offers the advantage of systematic verification of identified interactions. In a first round, a gene is tagged, and its interaction partners are sought. Interacting proteins identified in the first round are subsequently tagged, and the procedure is repeated (24). In contrast to the parallel analysis, this approach, previously termed SEAM for sequential rounds of epitope tagging, immunoaffinity chromatography, and mass spectrometry (25), is better suited for addressing specific biological problems than for charting protein-protein interactions on a proteome scale. Importantly, the function of identified proteins is independently evaluated in biological experiments, which effectively navigate further IPs. It is therefore conceivable that at some point subunit(s) associated with core subunits of complexes with different functionality would be identified thus linking the complexes into a network. In this paper we employed sequential tagging to identify interaction partners for a selection of genes, which may potentially be involved in chromatin remodeling, RNA processing, and regulation of transcription. By identifying interactors of 48 TAP-tagged yeast genes, we assess the analytical perspectives of the technology as a generic functional proteomics tool.

#### EXPERIMENTAL PROCEDURES

*Epitope Tagging of Genes and Isolation of Protein Complexes*—Yeast transformations were performed as described (26). Genes of interest were tagged by in-frame fusion of the ORFs with a PCR-generated targeting cassette encoding for the TAP tag and a selectable marker (20, 21). Correct cassette integrations were confirmed by PCR and Western blot analysis.

The extraction of yeast cells was performed as described by Logie and Peterson (27). We found that the procedure of Rigaut *et al.* (20), which employs a French press for breaking cells, produced poor results. Reproducibility was improved significantly by using the glass bead beater protocol described by Logie and Peterson (27). TAP purification was performed according to Rigaut *et al.* (20) with the following modifications: 10 ml of supernatant collected after a 43,000-rpm centrifugation were allowed to bind to 200  $\mu$ l of IgG-Sepharose (Amersham Biosciences), equilibrated in Buffer E (27) for 2 h at 4 °C using a disposable chromatography column (Bio-Rad). 2–3 column volumes (the equivalent of 4–6 liters of yeast culture at  $A_{600} = 2$ –3) were used per purification. The IgG-Sepharose column was washed with 35 ml of Buffer E without proteinase inhibitors, followed by 10 ml of the TEV cleavage buffer (20). Cleavage with TEV was performed using 10  $\mu$ l (100 units) of rTEV (Invitrogen) in 1 ml of cleavage buffer for 2 h at 16 °C. Calmodulin-Sepharose (Stratagene) purification was performed as described (20). Purified proteins were concentrated according to Wessel and Flugge (28).

*Identification of Proteins by Mass Spectrometry*—Proteins were separated by electrophoresis using gradient (6–18%) one-dimen-

sional polyacrylamide gels and visualized by staining with Coomassie. Protein bands were excised and in-gel digested with trypsin (unmodified, sequencing grade; Roche Diagnostics) as described (29, 30). Proteins were identified by a combination of MALDI MS peptide mapping and nanoelectrospray tandem mass spectrometric sequencing as described (31). Briefly, 1- $\mu$ l aliquots were withdrawn from the in-gel digests and analyzed on a REFLEX III mass spectrometer (Bruker Daltonics) using a thin-layer probe preparation method (32). If no conclusive identification was achieved, the gel pieces were extracted with 5% formic acid and acetonitrile. Unseparated mixtures of recovered tryptic peptides were sequenced by nanoelectrospray tandem mass spectrometry (nanoES MS/MS) as described (33) on a triple quadrupole mass spectrometer API III or on a QSTAR quadrupole time-of-flight mass spectrometer (both from MDS Sciex, Concord, Canada). Database searching was performed against a comprehensive non-redundant database using MASCOT software (34).

#### RESULTS AND DISCUSSION

*Tagging of Proteins and Immunoaffinity Isolation of Protein Complexes*—The machinery for homologous recombination in *Saccharomyces cerevisiae* provides a simple and precise way to integrate DNA fragments into the genome so that endogenous in-frame fusions between any ORF of choice and PCR-generated epitope tag cassettes can be achieved readily. When chromosomal copies of genes are modified, the tagged proteins are expressed from their native promoters, and therefore the expression is typically close to its physiological levels. However, homologous recombination imposes a few practical limitations onto the epitope-tagging method. It is easier to fuse the tag into the C terminus of the protein backbone rather than into the N terminus. For C-terminal fusions the selectable marker gene can be placed further 3' of the coding region of the fused protein, whereas for N-terminal fusions the marker usually intervenes between the promoter and the targeted ORF and has to be subsequently removed. Consequently C-terminal fusions are technically simpler and were used in the present work. The alternative N-terminal tagging was applied only three times when a problem with C-terminal tagging had been observed.

Although a variety of epitope tags have been described (19), TAP tagging offers several important advantages. The TAP tag consists of two high affinity modules, a calmodulin binding peptide, and a double protein A epitope, which is separated by a cleavage site for TEV protease (20). Protein complex purification is achieved via two-step affinity chromatography, which is carried out under conditions that leave proteins intact. In addition to the high specificity of binding of each part of the tag, the tethered protein complex is cleaved off protein A-IgG-Sepharose beads by the highly specific TEV protease, leaving a bulk of nonspecifically associated proteins on the beads. Therefore the purification results in much less background compared with conventional IP methods.

Altogether, 53 genes were epitope-tagged using the TAP tag, and in 52 cases (98%) the tag was incorporated successfully, as assessed by Western blotting. In one case, homologous recombination was successful, but Western blotting

detected no fusion protein. Possibly this was because of an incorrect prediction of the ORF or the absence of noticeable protein expression under cell culturing conditions. Unexpectedly, in no case did we observe lethality or an obviously disturbed phenotype, caused by the tag. In a single case (C-terminal TAP tag fusion to Set1) we observed that the complex was not perturbed; however, the tag did interfere with its enzymatic activity (35). Fusion of the tag to the N terminus of Set1 led to retrieval of the same complex with no interference of enzymatic activity (data not shown).

In four of 52 cases the bait proteins were observed in Western blots; however, insufficient amounts of IPed proteins precluded their detection by mass spectrometry. Fusing the TAP tag to the N termini for two of these four proteins did not improve the yield, and scaling up the purification procedure did not improve results either. The codon bias index (CBI) (36) of each of these four proteins was lower than 0.1 suggesting they are of low abundance.

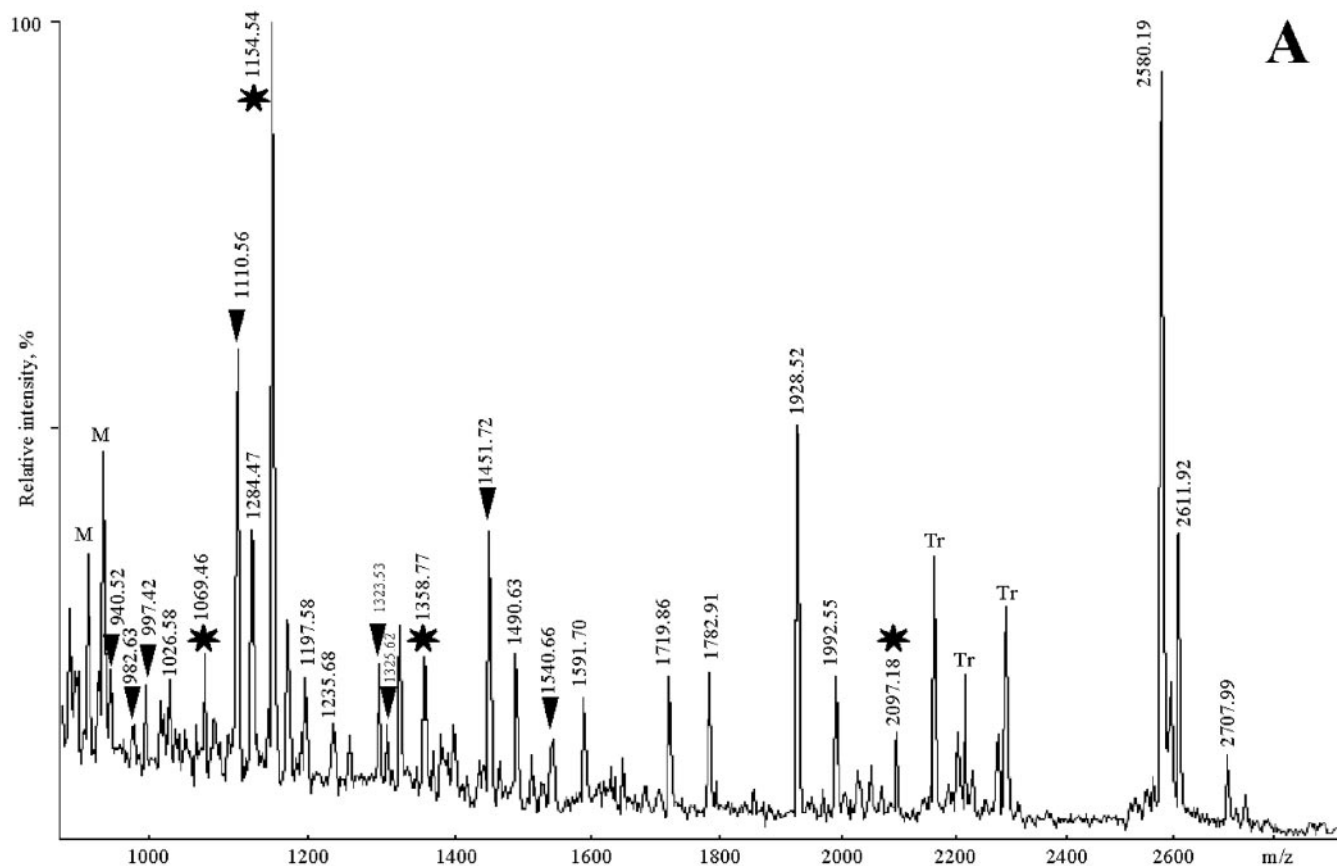
The successfully tagged genes encoded for proteins of a variety of physical properties with molecular mass between 9 and 175 kDa, calculated pI from 4.5 to 10.0, and CBI from -0.064 up to 0.16. Assuming that CBI represents reasonably a relative level of the protein expression (37), we concluded that TAP tagging was successful for low expressed proteins. Among the 52 successfully tagged genes 11 were essential, according to the YPD database (38). Altogether, 48 tagged genes (91%) were recovered by immunoaffinity chromatography in amounts sufficient for their reliable detection by mass spectrometry.

*Confident Identification of Proteins by Mass Spectrometry*—Deciphering protein complexes is a challenging task for mass spectrometry. First, protein complexes comprise subunits of various molecular weight, pI, and hydrophobicity, and therefore the number of peptides recovered from their in-gel digests markedly varies. Second, immunoaffinity isolation is performed typically under conditions that preserve relatively weak protein-protein interactions, and therefore co-isolation of nonspecifically associated proteins commonly occurs. Third, immunoaffinity purifications are usually difficult to scale up. If the yield of proteins of interest is low, their isolation from a larger volume of the cell culture results typically in loss of binding specificity and increased background. Although the TAP method reduces protein background considerably (see below), co-migration of two or more proteins within a single band occurs frequently, especially in the low molecular weight region. Therefore mass spectrometry is required to identify confidently the protein even if only a single peptide is recovered from its digest. Typical problems in confident protein identification are discussed here using an example of the analysis of an ~20-kDa Coomassie-stained protein band observed in the immunoprecipitate of tagged Set1 (35). Despite the low molecular mass of the protein, 25 prominent peptide peaks were detected in a MALDI MS spectrum of its digest (Fig. 1A), and their masses were used for database searching.

The search hit the 15-kDa 40 S ribosomal protein S24 (accession number P26782) with the score 115 (statistical significance threshold score was 52) by matching eight peptide ions with better than 100 ppm mass accuracy and better than 50% sequence coverage. However, the most intense peaks remained unaccounted for, and database searching with masses of unmatched ions did not result in any more hits, thus indicating that other yet unidentified protein(s) might be present in the sample. The digest was further analyzed by nanoES MS/MS, and another six ribosomal proteins, each of which matched a single unique sequenced peptide, were identified (Fig. 1B). Some of these proteins were among low confidence hits from MALDI MS analysis; however, none of the sequenced peptides identified S24 protein, the top hit. A single peptide sequence deduced from the spectrum acquired from a doubly charged ion with  $m/z$  382.0 matched the 15-kDa protein YBR258c (Fig. 1C). However, this hit could not be judged as confident as the retrieved sequence was short and degenerate. It is also known that large multiply charged peptides often undergo partial orifice fragmentation yielding abundant singly and doubly charged  $y$ -ions, and therefore database searching should not rely upon the cleavage specificity of trypsin. In fact, the peptide sequence (Leu/Ile)(Leu/Ile)Glu(Met(ox)/Phe)(Leu/Ile)Lys hits more than 200 proteins in a comprehensive database, including six proteins from the budding yeast. We retrospectively examined the MALDI MS map and found that the masses of another four peptides matched the sequence of YBR258c, and none of them matched other yeast protein candidates. We therefore concluded that although neither MALDI MS nor NanoES MS vouched for unambiguous identification, a combination of the two techniques produced a confident hit. Subsequent tagging of YBR258c confirmed that it is a *bona fide* subunit of the protein complex (35).

This example illustrates the need for complementary techniques to elucidate fully the composition of protein samples. NanoES MS and MALDI MS detect complementary peptide patterns, and their combination increases confidence in protein identification (39, 40). However, applying two ionization methods to the analysis of the same protein digest limits the throughput. An alternative is offered by emerging MALDI QqTOF mass spectrometry (41, 42) or MALDI TOF/TOF mass spectrometry (43), in which MALDI MS peptide mapping is complemented by an optional acquisition of high mass accuracy tandem mass spectra from multiple precursors detected in the spectrum of a digest, thus attaining high throughput and high confidence in the analysis of protein mixtures.

*Protein Background in the TAP Method*—Conventional IP experiments often result in complex patterns of co-isolated proteins (44). The immunoprecipitated proteins are usually separated on a one-dimensional polyacrylamide gel, and the pattern of proteins observed in the experiment lane is compared with the one in the control lane, and then only proteins detected selectively in the experiment are subjected to further



**FIG. 1. Mass spectrometric identification of a low molecular weight yeast protein.** A, MALDI MS peptide map of the tryptic digest of the 20-kDa protein band. Peaks of autolysis products of trypsin are designated with *Tr*, matrix peaks with *M*. Peptide peaks, which were retrospectively matched to YBR258c, are designated with asterisks; peaks matched to the ribosomal protein S24 are designated with filled triangles. B, nanoES mass spectrum of the same digest acquired in precursor ion scanning mode with the selected fragment ion at *m/z* 86. Peaks designated with *m/z* and charges were fragmented, and MS/MS spectra were matched to the sequences of tryptic peptides from various ribosomal proteins (see accession numbers at the corresponding peaks). The peak designated with the asterisk was the only peptide ion that matched yeast protein YBR258c. C, tandem mass spectrum acquired from the peptide ion with *m/z* 382.0 and the candidate peptide sequence, deduced by considering mass differences between *y*-ions.

identification by mass spectrometry. However, this approach is slow and prone to errors. We therefore excised and analyzed by mass spectrometry all bands detected in the experiment lane, effectively using no control. We identified a subset of proteins that were detected repeatedly using the TAP tag and hence are common background contaminants (Table I). Although these proteins vary in function, molecular weight and *pI*, they are all very abundant. In addition to these common proteins, we detected a few contaminants that were observed only occasionally (Table I) and are most likely because of small variations in the reagents used, in particular because of different batches of calmodulin beads or because of phenotypic alterations caused by the tag.

**Dissection of Protein Complexes**—The major difference in methodology between high throughput charting of protein-protein interactions and the more focused approach described here is that newly identified interactions were verified independently in a variety of experiments, including (but not

limited to) Western blotting, gene knockouts, and protein colocalization. For charting the proteomes anchored to selected proteins we chose a strategy of sequential protein tagging and immunoaffinity isolation, as described above. Details of the composition and biological role of some of the identified protein complexes are provided in Refs. 22, 23, 35, and 45.

The dissection of the subset of the proteome anchored at Set3 protein is discussed here as an illustrative example. Set3 possesses SET and PHD finger domains, which are hallmarks of proteins involved in chromatin regulation and epigenetics (45). Initially Set3 was TAP-tagged and immunoaffinity-isolated, and subsequent mass spectrometric analysis identified eight interacting proteins (Fig. 2). Kap95 and Kap60 belong to the family of importins; Hos2 and Hst1 were putative histone deacetylases; Sif2 and YIL112c contain multiple repeats of generic protein-protein interaction motifs, WD40 and ankyrin, respectively; YCR033w (Snt1) protein contains the putative DNA-binding SANT domain; and Cph1 (cyclophilin A) is a

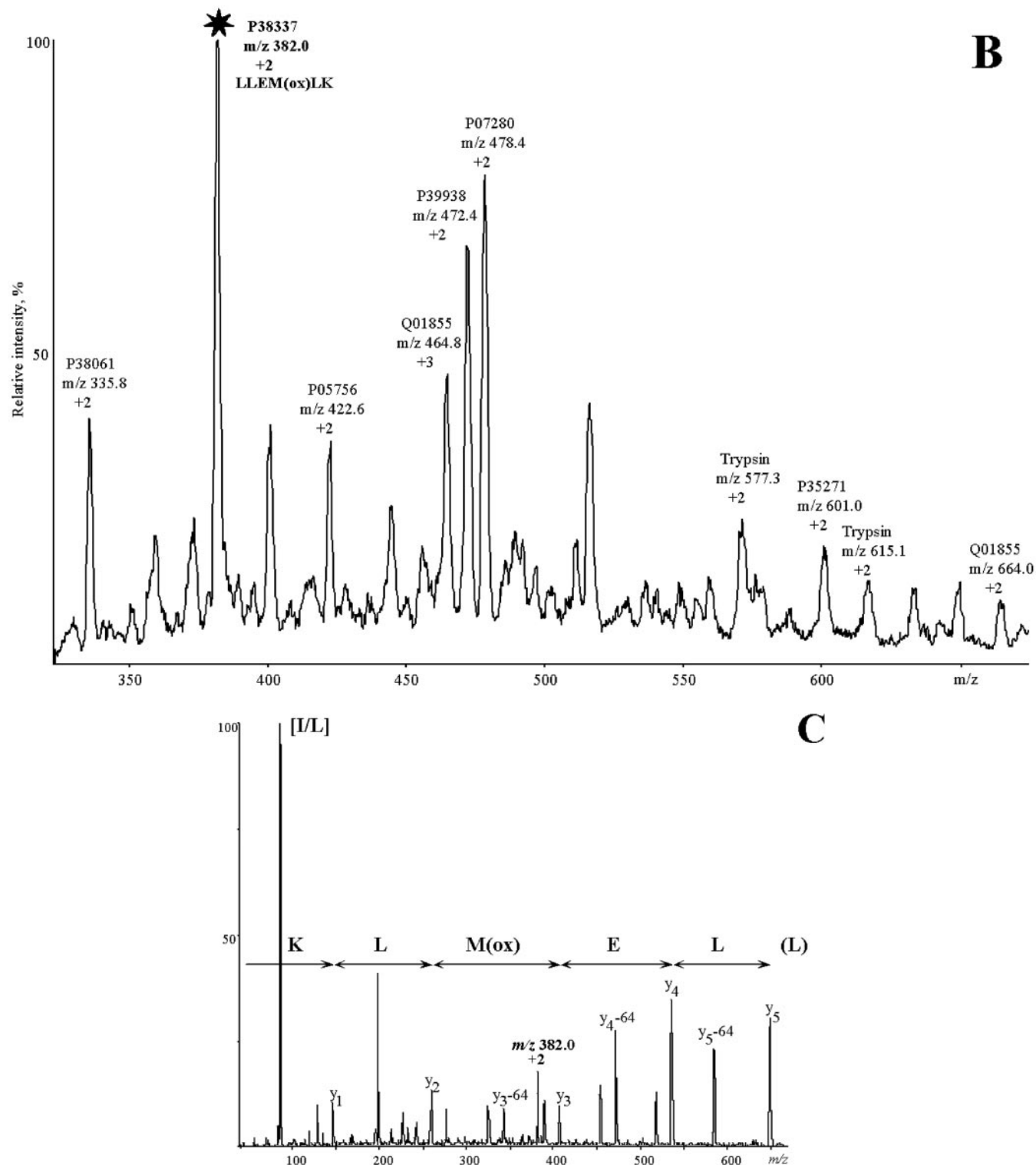


Fig. 1—continued

prolyl isomerase. The variability of plausible cellular functions of Set3 interactors prompted questions about the unity of the isolated complex. How many distinct protein assemblies were involved, and were artifacts included?

To address these and other questions, the Set3 interactors, Sif2, Snt1, Yil112c, Cph1, Hst1, and Hos2, were tagged and subjected to another round of the purification and identification protocol. Each time (with the exception of Cph1, whose

TABLE I  
 Background proteins in the TAP method

Protein	Gene	Molecular mass <i>kDa</i>	pI	Codon bias index	Localization
Heat shock protein homolog	SSE1	77.2	5.0	0.642	Cytoplasmic
Major coat protein of LA virus	GAG	76.3	5.5		Cytoplasmic
Heat shock proteins	SSA1/SSA2	69.4	4.9	0.893	Cytoplasmic
Heat shock proteins	SSB1/SSB2	66.5	5.2	0.890	Cytoplasmic
Pyruvate kinase 1	CDC19	54.1	7.7	0.965	Cytoplasmic
Elongation factor 1- $\alpha$	TEF1	49.9	9.2	0.929	Cytoplasmic
Tryptophanyl-tRNA synthetase	WRS1	49.2	6.6	0.350	Cytoplasmic
Enolase 2	ENO2	46.7	5.8	0.960	Cytoplasmic
Phosphoglycerate kinase	PGK1	44.6	7.1	0.906	Cytoplasmic
Glyceraldehyde 3-phosphate dehydrogenase 3	TDH3	35.6	6.7	0.988	Cytoplasmic
Vacuolar ATP synthase subunit E	VMA4	26.3	5.2	0.427	Cytoplasmic
Phosphoglycerate mutase	GPM1	27.6	8.9	0.811	Cytoplasmic
60 S ribosomal protein L3	MAK8	43.6	10.4	0.890	Cytoplasmic
Putative ATP-dependent RNA helicase <sup>a</sup>	DED1 <sup>a</sup>	65.4	7.9	0.376 <sup>a</sup>	Cytoplasmic
Heat shock protein 60 <sup>a</sup>	MIF4 <sup>a</sup>	58.2	4.9	0.382 <sup>a</sup>	Mitochondrial
Mitochondrial import inner membrane translocase <sup>a</sup>	TIM44 <sup>a</sup>	48.8	9.6	0.140 <sup>a</sup>	Mitochondrial membrane
Other ribosomal proteins <sup>b</sup>		10–30			
Mean				0.89 $\pm$ 0.22	

<sup>a</sup> Proteins were only observed in some immunoprecipitations.

<sup>b</sup> Background low molecular mass ribosomal proteins may vary. Proteins Rpl32, Rps13, Rps15, Rps26A/B, Rpl19A/B, Rps24A/B, Rps19A/B, Rps18A/B, Rpl17A/B, Rpl7A/B, Rpl4A/B, Rps11A/B, Rps14A/B, Rpl33A/B, Rpl10, Rps7A/B, Rps6A/B, Rpl2A/B were the most typical contaminants.

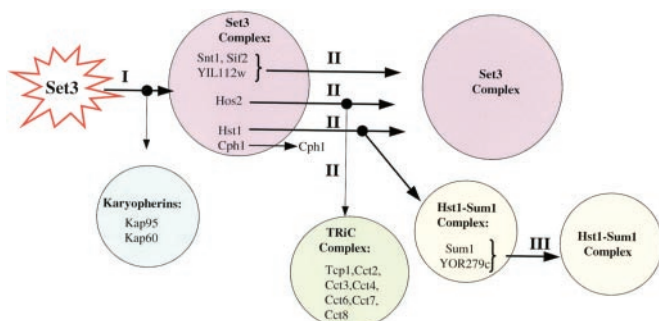


FIG. 2. Charting the proteome anchored at Set3 by sequential rounds of epitope tagging, immunoprecipitation, and mass spectrometry. Rounds of tagging are designated as I, II, and III. Individual protein complexes are circled.

purification yielded only the bait) the same set of proteins was co-immunoprecipitated, thus proving that they are *bona fide* members of a single histone deacetylase complex termed Set3C (Fig. 2).

The tagging of Set3C members indicated that three subunits were also engaged in specific interactions with other protein complexes. The interaction with importins Kap60 and Kap95 was only observed when Set3 was tagged. Similarly, tagged Hos2 pulled down seven of eight known subunits of the chaperonin complex, TRiC (which, unlike heat shock chaperones SSA and SSB, does not belong to common background proteins in TAP purifications (Table I)). No interaction with importins and/or TriC was detected when any other Set3C member was tagged.

Tagging Hst1 revealed that it is also engaged in another functionally distinct protein complex with YOR279c and Sum1. To validate its integrity, a third round of tagging and purification was performed on both Sum1 and YOR279c, thereby confirming the interactions among Hst1, YOR279c, and Sum1. Noticeably, no members of Set3C (other than Hst1) were detected in Sum1 and YOR279c precipitates. Thus by starting at a single entry gene, *set3*, sequential rounds of epitope tagging, immunoprecipitation, and mass spectrometry identified two novel functionally distinct protein complexes with plausible histone deacetylase activity, linked via a shared subunit, Hst1, and also indicated linkage of two Set3C subunits to other known complexes.

Analysis of these and other protein complexes allowed us to draw a few conclusions about strategies for the characterization of a protein “interactome,” *i.e.* of a network of interacting protein complexes (46). Apparently IPs can pool together members of different protein complexes and therefore the “guilty by association” concept of defining what proteins belong to the complex is inherently error-prone. This may become a severe limitation for high throughput “parallel” analysis, in which *bona fide* interaction partners are established neither by functional experiments nor by sequential tagging and purification of other candidate subunits of the complex.

It is also important to distinguish proteins that represent a core of the complex and that are essential for its integrity, from proteins whose interaction with the complex is transient. Therefore even approximate estimation of stoichiometry of protein interactions adds vitally important pieces of informa-

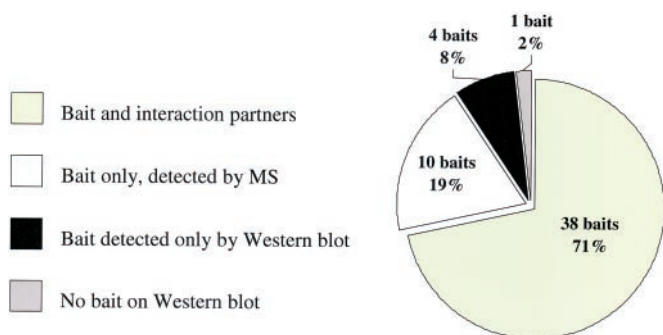


FIG. 3. Pie chart summarizing the results of immunoprecipitations of TAP-tagged proteins.

tion. Mass spectrometry is well suited to determine relative changes in the concentration of the same protein obtained under different experimental conditions (reviewed in Ref. 47). However, it is not straightforward to compare the concentration of different proteins present in the mixture. Amino acid composition of detected peptides strongly affects the signal intensity observed in a mass spectrum (48), and the pattern of peptide maps and the recovery of individual peptides depend on protein visualization and sample-processing protocols (49, 50). Gel electrophoresis and visualization of bands by Coomassie staining is less dependent on protein properties and is widely applied in expressional proteomic studies (51). Hst1, importins Kap95 and Kap60, and the TriC members were detected in apparently substoichiometric amounts compared with other subunits of Set3C, and thus their transient association with the core of the complex can be inferred. In the case of Hst1, this was confirmed by IP of intact Set3C from  $\Delta hst1$  strain (45). Similarly, semiquantitative information, taken together with IP patterns of other tagged Set3C subunits, assisted in charting the boundaries between individual protein complexes pulled down by IP of TAP-tagged Set3C members.

**Identified Protein Complexes and Segments of a Protein Interaction Network**—In 48 successful IPs, interaction partners were determined for 38 baits (71%), and in 10 IPs only the bait protein was detected (Fig. 3). Noticeably, these 10 idle baits were of average molecular weight and pI. Three of them were abundant proteins (CBI > 0.2), and one gene is essential. The 38 successful baits pulled down a total of 220 interaction partners, which are members of 19 functionally distinct protein complexes with the average of 5.8 interactors per protein, a value that agrees with bioinformatic estimates (10, 52). The complexes comprised from three to 16 subunits and varied in function and cellular localization; however, none of identified preys was a membrane protein. We underscore here that no protein complexes were defined solely on the basis of a single IP experiment.

At the time of these experiments, 16 protein complexes were either new or were known complexes to which new subunits were identified. 19 subunits of six complexes were

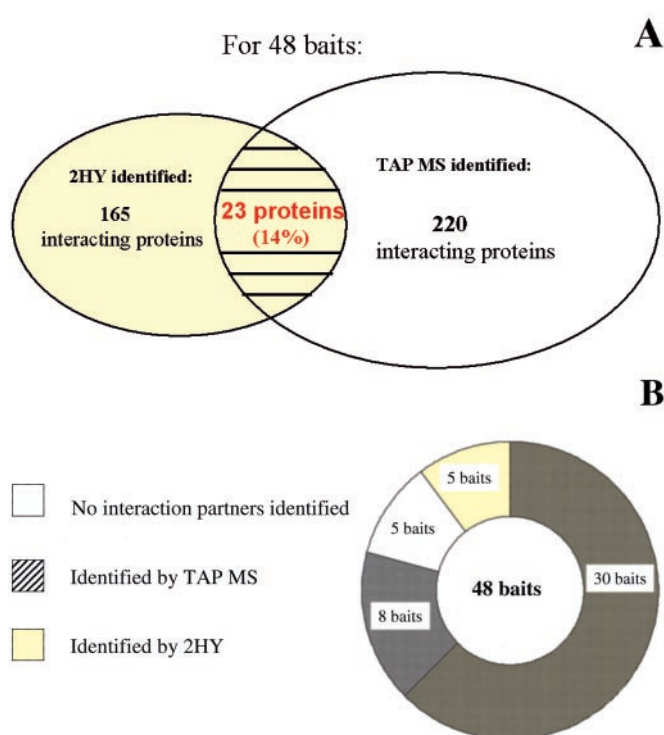


FIG. 4. Charting of protein-protein interactions by two-hybrid screening and mass spectrometry is shown as follows: A, from the perspective of preys. Pools of interacting proteins identified by both methods for 48 selected baits are shown. B, from the perspective of baits. Baits, interaction partners of which were identified by 2HY screening, TAP-MS, or by both methods, are shown.

subjected to verification by sequential tagging, as described in the previous section, and in four cases physical links to other multiprotein complexes were revealed.

We found our data in noticeable disagreement with the complexity of protein interaction networks suggested by alternative genome mining approaches, such as two-hybrid screening and bioinformatics (10–12). As similar discrepancies were observed previously in the analysis of affinity-purified protein complexes (25, 46), we believe the experiments point to some fundamental limitations, which should be further understood and accounted for in the elucidation of the molecular organization of the proteome.

**Interaction Partners Identified by TAP and by Two-hybrid Screening**—As mentioned above, a total of 48 proteins were successfully tagged, and interaction partners were identified in IPs of 38 baits. We compared further TAP-identified interactors with the ones suggested for the same baits by genome-wide two-hybrid screening (11, 12). Of the 48 baits, 2HY screening defined interaction partners for 35, with 165 interactors in total. Comparison to the set of 220 interactors identified by TAP and mass spectrometry revealed that only 23 proteins (14%) between these two sets overlapped (Fig. 4A).

We further checked whether 2HY or TAP-MS could identify interaction partners in those cases when the other method did

not (Fig. 4B). 2HY screening did not find interactors for 13 baits among the 48 TAP-tagged proteins. For eight of these 13 baits, interaction partners were identified by TAP. On the other hand, from the same selection of 48 genes, TAP failed to reveal interactors for 10 baits. For five of these 10 baits 2HY screening suggested 13 interaction partners. We also checked whether some of these 13 proteins are known *bona fide* members of previously identified complexes or interact with members of the complexes other than the baits and found that no such interaction has been described. We conclude that although the two technologies provide a complementary view on physical interactions within the proteome, they do not correlate well enough for further validation of identified interactions.

#### CONCLUSION AND PERSPECTIVES

We explored analytical perspectives of a combination of TAP and mass spectrometry for dissecting protein complexes and protein interaction networks in budding yeast. Although the sample selection is small compared with the total number of proteins encoded in the yeast genome, it allowed us to comment on strategies for elucidation of the genomic blueprint via mining the proteome. By accurate mapping of protein-protein interactions, we present currently the most detailed characterization of protein complexes, as well as of segments of a macronetwork that link them together. However, it has also become clear that understanding of protein complexes and their functional links requires substantially more work. As demonstrated above, rigorous validation of the inferred composition of protein complexes via multiple precipitations is not an excessive “purist” requirement but is rather a necessity. Similar to genomic sequencing, several readouts need to be obtained before the composition and links of the protein complex are charted accurately.

At present characterization of protein complexes by mass spectrometry is mostly limited to qualitative description of their composition and interactions. However, it is becoming apparent that methods should be developed to describe quantitatively the stoichiometry of protein-protein interactions.

These problems are challenging, but recent developments both in mass spectrometry and in gene manipulation technology suggest these goals are within reach. Deciphering of protein complexes will produce unique information for the understanding of functional organization of genomes of higher eukaryotes, including the human genome.

*Acknowledgments*—We are grateful to Dr. Matthias Wilm (EMBL) and to other members of the Stewart and Shevchenko groups for useful discussions and technical support.

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence may be addressed. E-mail: stewart@mpi-cbg.de.

¶ To whom correspondence may be addressed. Tel.: 49-351-210-2615; Fax: 49-351-210-2000; E-mail: shevchenko@mpi-cbg.de.

#### REFERENCES

- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996) Live with 6000 Genes. *Science* **274**, 546–567
- The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G. et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921
- Venter, J. C. (2001) The sequence of the human genome. *Science* **291**, 1304–1351
- Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294
- Eisenberg, D., Marcotte, E. M., Xenarios, I., and Yeates, T. O. (2000) Protein function in the post-genomic era. *Nature* **405**, 823–826
- Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934
- Becskei, A., Seraphin, B., and Serrano, L. (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535
- Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science* **285**, 751–753
- Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4569–4574
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297
- Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., and Snyder, M. (2001) Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105
- Lamond, A., and Mann, M. (1997) Cell biology and the genome projects - a concerted strategy for characterizing multiprotein complexes by using mass spectrometry. *Trends Cell Biol.* **7**, 139–142
- Mann, M., Hendrickson, R. C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **70**, 437–473
- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., III (1999) Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–682
- Washburn, M. P., Wolters, D., and Yates, J. R., III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247
- Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K., and Schiebel, E. (1999) Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15**, 963–972
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030–1032
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification.



- Methods* **24**, 218–229
22. Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Seraphin, B. (2000) A Sm-like protein complex that participates in mRNA degradation. *EMBO J.* **19**, 1661–1671
  23. Caspary, F., Shevchenko, A., Wilm, M., and Seraphin, B. (1999) Partial purification of the yeast U2 snRNP reveals a novel yeast pre-mRNA splicing factor required for pre-spliceosome assembly. *EMBO J.* **18**, 3463–3474
  24. Shevchenko, A., Zachariae, W., and Shevchenko, A. (1999) A strategy for the characterization of protein interaction networks by mass spectrometry. *Biochem. Soc. Trans.* **27**, 549–554
  25. Seol, J. H., Shevchenko, A., Shevchenko, A., and Deshaies, R. J. (2001) Skp1 forms multiple protein complexes, including RAVE, a regulator of V-ATPase assembly. *Nat. Cell Biol.* **3**, 384–391
  26. Soni, R., Carmichael, J. P., and Murray, J. A. (1993) Parameters affecting lithium acetate-mediated transformation of *Saccharomyces cerevisiae* and development of a rapid and simplified procedure. *Curr. Genet.* **24**, 455–459
  27. Logie, C., and Peterson, C. L. (1999) Purification and biochemical properties of yeast SWI/SNF complex. *Methods Enzymol.* **304**, 726–741
  28. Wessel, D., and Flugge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143
  29. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858
  30. Shevchenko, A., Chernushevich, I., Wilm, M., and Mann, M. (2000) *De Novo* peptide sequencing by nano-electrospray tandem mass spectrometry using triple quadrupole and quadrupole/time-of flight instrument. *Methods Mol. Biol.* **146**, 1–16
  31. Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., and Mann, M. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14440–14445
  32. Vorm, O., Roepstorff, P., and Mann, M. (1994) Matrix surfaces made by fast evaporation yield improved resolution and very high sensitivity in MALDI TOF. *Anal. Chem.* **66**, 3281–3287
  33. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469
  34. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567
  35. Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W. M., Wilm, M., Aasland, R., and Stewart, A. F. (2001) The *S. cerevisiae* Set1 complex includes an Ash2 homolog and methylates histone 3 lysine 4. *EMBO J.* **20**, 7137–7148
  36. Kurland, C. G. (1991) Codon bias and gene expression. *FEBS Lett.* **285**, 165–169
  37. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9390–9395
  38. Costanzo, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Olsen, P., Robertson, L. S., Skrzypek, M. S., Braun, B. R., Hopkins, K. L., Kondu, P., Lengieza, C., Lew-Smith, J. E., Tillberg, M., and Garrels, J. I. (2001) YPD, PombePD and WormPD: model organism volumes of the Bio-Knowledge library, an integrated resource for protein information. *Nucleic Acids Res.* **29**, 75–79
  39. Medzihradszky, K. F., Leffler, H., Baldwin, M. A., and Burlingame, A. L. (2001) Protein identification by in-gel digestion, high-performance liquid chromatography, and mass spectrometry: peptide analysis by complementary ionization techniques. *J. Am. Soc. Mass Spectrom.* **12**, 215–221
  40. Shevchenko, A., Loboda, A., Ens, W., Schraven, B., Standing, K. G., and Shevchenko, A. (2001) Archived polyacrylamide gels as a resource for proteome characterization by mass spectrometry. *Electrophoresis* **22**, 1194–1203
  41. Shevchenko, A., Loboda, A., Shevchenko, A., Ens, W., and Standing, K. G. (2000) MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal. Chem.* **72**, 2132–2141
  42. Krutchinsky, A. N., Zhang, W., and Chait, B. T. (2000) Rapidly switchable matrix-assisted laser desorption/ionization and electrospray quadrupole-time-of-flight mass spectrometry for protein identification. *J. Am. Soc. Mass Spectrom.* **11**, 493–504
  43. Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L., and Burlingame, A. L. (2000) The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **72**, 552–558
  44. Shevchenko, A., and Mann, M. (1999) in *Mass Spectrometry in Biology and Medicine* (Burlingame, A. L., Carr, S. A., and Baldwin, M., eds) pp. 237–269, Humana Press, Totowa, NJ
  45. Pijnappel, W. W., Schaft, D., Roguev, A., Shevchenko, A., Tekotte, H., Wilm, M., Rigaut, G., Seraphin, B., Aasland, R., and Stewart, A. F. (2001) The *S. cerevisiae* SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. *Genes Dev.* **15**, 2991–3004
  46. Deshaies, R. J., Seol, J. H., McDonald, W. H., Cope, G., Lyapina, S., Shevchenko, A., Shevchenko, A., Verma, R., and Yates, J. R., III (2002) Charting the protein complexome in yeast by mass spectrometry. *Mol. Cell. Proteom.* **1**, 3–10
  47. Griffin, T. J., and Aebersold, R. (2001) Advances in proteome analysis by mass spectrometry. *J. Biol. Chem.* **276**, 45497–45500
  48. Krause, E., Wenschuh, H., and Jungblut, P. R. (1999) The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Anal. Chem.* **71**, 4160–4165
  49. Shevchenko, A., and Shevchenko, A. (2001) Evaluation of the efficiency of in-gel digestion of proteins by peptide isotopic labeling and MALDI mass spectrometry. *Anal. Biochem.* **296**, 279–283
  50. Lauber, W. M., Carrol, J. A., Dunfield, D. R., Kiesel, J. R., Radabaugh, M. R., and Malone, J. P. (2001) Mass spectrometry compatibility of two-dimensional gel protein stains. *Electrophoresis* **22**, 906–918
  51. Rabilloud, T. (2000) Detecting proteins separated by 2-D gel electrophoresis. *Anal. Chem.* **72**, 48A–55A
  52. Huynen, M., Snel, B., Lathe, W., III, and Bork, P. (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res.* **10**, 1204–1210