2 Applicability of TAP–MudPIT to Pathway Proteomics in Yeast

This chapter describes the exploration of the use of multidiemensional protein identification technology for the analysis of modereately complex polypeptide mixtures as resulting from affinity purification of protein complexes in a nonspecialized accademic laboratory setting. It was published as

> Graumann, J., Dunipace, L. A., Seol, J. H., McDonald, W. H. and Yates III, J. R. et al. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics*, 3(3):226–37.

 The supplementary material referred to is available at http://www.mcponline.org/cgi/data/M300099-MCP200/DC1/1.

2.1 Summary

A combined multidimensional chromatography–mass spectrometry approach known as "MudPIT" enables rapid identification of proteins that interact with a tagged bait while bypassing some of the problems associated with analysis of polypeptides excised from SDS–polyacrylamide gels. However, the reproducibility, success rate, and applicability of MudPIT to the rapid characterization of dozens of proteins have not been reported. We show here that MudPIT reproducibly identified bona fide partners for budding yeast Gcn5p. Additionally, we successfully applied MudPIT to rapidly screen through a collection of tagged polypeptides to identify new

protein interactions. Twenty–five proteins involved in transcription and progression through mitosis were modified with a new TAP tag. TAP–MudPIT analysis of 22 yeast strains that expressed these tagged proteins uncovered known or likely interacting partners for 21 of the baits, a figure that compares favorably with traditional approaches. The proteins identified here comprised 102 previously–known and 279 potential physical interactions. Even for the intensively studied Swi2p/Snf2p, the catalytic subunit of the Swi/Snf chromatin remodeling complex, our analysis uncovered a new interacting protein, Rtt102p. Reciprocal tagging and TAP–MudPIT analysis of Rtt102p revealed subunits of both the Swi/Snf and RSC complexes, identifying Rtt102p as a common interactor with, and possible integral component of, these chromatin remodeling machines. Our experience indicates it is feasible for an investigator working with a single ion trap instrument in a conventional molecular/cellular biology laboratory to carry out proteomic characterization of a pathway, organelle, or process (i. e. "pathway proteomics") by systematic application of TAP–MudPIT .

2.2 Introduction

To understand the function of a protein, it is crucial to characterize its physical environment: what other proteins is it interacting with under various conditions? Traditionally, this question has been addressed by biochemical fractionation of cell extracts under mild conditions and subsequent identification of the members of a purified protein complex by immunoblotting or peptide sequencing.

Primed by the dawning of the postgenomic era, genome–wide yeast two–hybrid interaction screens (Ito et al. 2001; and Uetz et al. 2000) and protein chip based

methods (Zhu et al. 2001) have supplemented traditional purification and identification techniques, allowing broader insight into the interaction networks that constitute a functional cell. Both of these approaches require the creation and maintenance of libraries of tagged proteins and in the case of protein chips the daunting task of purifying and spotting them under conditions that preserve their activity. The potential for detecting nonphysiological protein–protein interactions and the necessity to piece together interaction networks from a catalog of resulting binary interactions further complicate these approaches.

Developed in parallel with two–hybrid and protein chip technologies, mass spectrometry of protein complexes purified through single or tandem affinity steps eliminates the need for complex–specific immunochemicals and enables analysis of very small amounts of sample on a proteome wide scale (Gavin et al. 2002; and Ho et al. 2002). This approach can be performed under more physiological conditions and substitutes whole complex analysis for the reconstruction of interaction networks from binary interaction data. However, the Gavin et al. (2002) and Ho et al. (2002) studies employed SDS–PAGE to separate affinity–purified protein mixtures prior to mass spectrometric analysis, thereby encountering the problems linked to this technique including: limitations of dynamic range of detection, considerable sample parallelization, variable elution efficiency of peptides from the polyacrylamide matrix, and potential selection against proteins with properties that impede analysis by SDS–PAGE (e. g., unusually high or lowmolecular weight, diffuse migration, comigration with contaminants, and poor binding to stain).

To circumvent these problems McCormack et al. (1997) demonstrated the possibility of analyzing digested protein complexes directly using single dimensional liquid chromatography. An improvement of this method—multidimensional protein

identification technology (MudPIT; Link et al. 1999)—extended its applicability to large protein complexes and is a bona fide alternative to gel–based protein separation. MudPIT relies on digestion in solution of the protein mixture to be analyzed, and separation of the resulting complex peptide mixture by multidimensional capillary chromatography connected in–line to an ion trap mass spectrometer. Owing to its unique advantages, MudPIT is an attractive alternative to traditional methods for the rapid identification of protein–protein interactions for stoichiometric and substoichiometric partners. MudPIT can also be applied to deconvolve complex sets of proteins related by a common property. For example, Peng et al. (2003) applied a multidimensional approach similar to MudPIT to identify hundreds of candidate ubiquitinated proteins in budding yeast cells.

Despite its considerable power, some potential limitations to MudPIT remain to be addressed. For example, it is unclear how reproducible such analyses are. This is of particular concern for analysis of samples that contain many proteins, like that reported by Peng et al. (2003). Second, since only individual analyses have been reported to date, it remains unclear what the likelihood of success is for any given MudPIT experiment. The success rate of individual experiments, in turn, is important for the question of whether it will be profitable to scale the MudPIT approach to the rapid analysis of multiple baits. Third, because the issues of reproducibility and scalability have not been addressed, it is not known if the parallel application of MudPIT to multiple proteins will enable filtering approaches to separate bona fide interactors from nonspecific contaminants. Finally, it remains unclear how feasible it will be to transfer cutting–edge proteomic technologies like MudPIT from specialized environments to a conventional cell biology laboratory.

In this study, we address these various issues. We show that the combination

of a bipartite affinity tag with MudPIT allows for the rapid analysis of protein complexes. Pilot experiments with Gcn5p confirmed the reproducibility of the technique. Application of MudPIT to a set of 22 expressed baits revealed a success rate comparable to conventional approaches, and confirmed the scalability of the approach. Comparison of proteins identified across all MudPIT analyses, comprising diverse baits from different subcellular compartments and pathways, also enabled a filtering strategy to cull nonspecific contaminants. Our experience indicates that multidimensional chromatography in combination with mass spectrometry technology can be readily transferred from a specialized analytical chemistry environment to a traditional molecular cell biology laboratory. Routine application of MudPIT may thus enable cell biologists to dissect dynamic changes in protein interactions in response to specific chemical or biological ligands, environmental perturbations, or mutations.

2.3 Experimental Procedures

2.3.1 Construction of a Bipartite Affinity Purification Tag

To construct pJS–HPM53H, a 940 bp fragment was PCR amplified from pJS–TM53H (RDB1344, Seol et al. 2001) with the primers HTM A and B (see supplementary material, table 1). This was used as a template to PCR amplify a HPM tag containing 670 bp fragment with the primers HPM C and D (see supplementary material, table 1), which replaced the XhoI–EcoRI restriction fragment of pJS–TM53H.

2.3.2 Strain Construction

The bipartite affinity purification tags were amplified by PCR from pJS–HPM53H (HPM tag) or pKW804 (modified TAP tag, Cheeseman et al. 2001) with primers conferring sequence homology to the 3' end of targeted open reading frames (see supplementary material, table 1), using Expand High Fidelity PCR System (Roche, Indianapolis, IN). The resulting PCR products were transformed into the *Saccharomyces cerevisiae* strain RJD 415 (W303 background, *MAT*a, *can1*–100, *leu2*–3,–112, *his3*–11,–15, *trp1*–1, *ura3*–1, *ade2*–1, *pep4*Δ::*TRP1*, *bar1*Δ::*HISG*; see supplementary material, table 2) with a modified Lithium acetate method (Ito et al. 1983). Integration and expression of the tagged gene product were checked by anti–myc western blotting of whole cell lysate using 9E10 monoclonal antibodies (Evan et al. 1985). Strain RJD 2067, carrying a TAP tagged (Rigaut et al. 1999) *GCN5* allele was a gift from Erin O'Shea, UCSF.

To knock out *SNF2*, *ARP9* and *RTT102*, an *HIS3* carrying cassette was PCR amplified from pFA6a–His3MX6 (Longtine et al. 1998) and transformed into the strain RJD 415. The primers used (see supplementary material, table 1), allowed for complete replacement of the respective open reading frames by homologous recombination.

2.3.3 Preparation of Protein Complexes by Dual–step Affinity Purification

2.3.4 HPM–Tag

Yeast cells carrying a HPM–tagged gene were grown in 2.5 l YPD (1 % yeast extract, 2 % bacto–peptone, 2 % glucose) to $OD_{600\ nm} \approx 1.5$. Cell extract was prepared by

glass beading in TNET (20 mM Tris·HCl pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 0.2 % Triton X–100), supplemented with 10 μg/ml Aprotinin, 10 μg/ml Leupeptin, 10 μg/ml Chymostatin and 2 μg/ml Pepstatin A. The extract was cleared by centrifugation at 100,000 g and 4°C for 30 min. Crude extract (300 mg total protein in 14 ml volume) was incubated with 200 μl 9E10 α–myc (Evan et al. 1985) coupled protein A sepharose beads (Sigma, St. Louis, MO) for 1.5 h at 4°C. The beads were washed three times in 50 bead volumes cold TNET, resuspended in 300 μl TNET and adjusted to 1 mM DTT. Protein complexes were eluted for 25 min at room temperature by addition of 10 units of GST–tagged PreScission Protease (Amersham, Piscataway, NJ) and protease carryover was reduced by 10 min further incubation with 1/10 9E10 bead volumes of glutathione sepharose 4B beads (Amersham, Piscataway, NJ).

For the second affinity purification step 20 μl of Ni–NTA agarose beads (Qiagen, Valencia, CA) were added to 200 μl supernatant from the first step and the sample was rotated for 1 h at 4°C. The beads were washed three times with 25 bead volumes of cold TNET and twice with 25 bead volumes of cold TNE (20 mM Tris·HCl pH 7.5, 150 mM NaCl, 0.1 mM EDTA). Proteins were eluted by addition of 50 μl 100 mM EDTA and the resulting supernatant lyophillized.

2.3.5 TAP Tag

Purification of TAP–tagged Gcn5p was modified from Rigaut et al. (1999). Protein extractions for strain RJD 2067 (see supplementary material, table 2), carrying a TAP tagged *GCN5* allele was performed as described for HPM tagged strains, substituting IPP150 (10 mM Tris·HCl pH 8.0, 150 mM NaCl, 0.1 % NP40) for TNET.

After protein extraction, 200 μl of IgG sepharose (Amersham, Piscataway, NJ) was added to 300 mg total protein in a volume of 14 ml. This slurry was incubated at 4°C, rotating for 2 h. After incubation, the resin was washed 3 times with 50 bead volumes of IPP150, and once with 50 bead volumes of TEV protease cleavage buffer (10 mM Tris·HCl pH 8.0, 150 mM NaCl, 0.1 % NP40, 0.5 mM EDTA, 1 mM DTT). The IgG sepharose was resuspended in 300 μl TEV protease cleavage buffer containing 100 U TEV protease (Invitrogen, Carlsbad, CA), and incubated at room temperature, rotating, for 45 min. The bead supernatant (280 μl) was then retrieved and mixed with 840 μl of calmodulin binding buffer (10 mM β–mercaptoethanol, 10 mM Tris·HCl pH 8.0, 150 mM NaCl, 1 mM Mg–acetate, 1 mM imidazole, 2 mM $CaCl_2$, 0.1 % NP40), 0.84 μl 1 M $CaCl_2$, and 200 μl calmodulin beads (Stratagene, La Jolla, CA). This mixture was incubated for 1 h at 4°C, with rotating. After incubation, the beads were washed 3 times with 5 bead volumes of calmodulin binding buffer and eluted 2 times with 250 μl of calmodulin elution buffer (10 mM β–mercaptoethanol, 10 mM Tris·HCl pH 8.0, 150 mM NaCl, 1 mM Mg–acetate, 1 mM imidazole, 2 mM EGTA, 0.1 % NP40). The eluate was TCA precipitated, and the pellet was washed two times with ice cold acetone.

2.3.6 Modified TAP–Tag

The protocol for affinity purification of Gcn5p tagged with the modified TAP tag was adapted from Cheeseman et al. (2001) and was identical to the TAP protocol up through the TEV protease treatment. After TEV protease digestion 50 μl of S protein agarose (Novagen, Madison, WI) was added to 280 μl of the supernatant and the slurry was incubated, rotating, at 4°C for 1.5 h. The beads were washed 3

times with 10 volumes of IPP150, once with IPP150 without NP40, and then with 50 mM Tris·HCl pH 8.5, 5 mM EGTA, 1 mM EDTA, 75 mM KCl. The protein was eluted in 50 μl 100 mM Tris·HCl pH 8.5, 8 M urea for 30 min at room temperature.

2.3.7 Proteolytic Digest

Protein samples were proteolytically digested as follows: lyophillized protein mixtures were resolubilized in 40 μl 8 M urea, 100 mM Tris·HCl pH 8.5 and reduced by incubation at a final concentration of 3 mM T–CEP (Pierce, Rockford, IL) for 20 min at room temperature. Reduced cysteines were subsequently alkylated by addition of iodoacetamide (10 mM final concentration) and incubation for 15 min at room temperature. Proteolysis was initiated with 0.1 μg endoproteinase Lys–C (Roche, Indianapolis, IN) and allowed to proceed for 4 h at 37°C. The sample was then diluted fourfold by addition of 100 mM Tris·HCl pH 8.5 and adjusted to 1 mM $CaCl_2$. Next, 0.5 μg of sequencing grade trypsin (Roche, Indianapolis, IN) were added and the mixture incubated overnight at 37°C. The digest was quenched with the addition of formic acid to 5 % and stored at −20°C.

2.3.8 Multidimensional Protein Identification Technology (MudPIT)

The peptide mixtures were separated utilizing a triphasic microcapillary column as described in McDonald et al. (2002). A fused silica capillary with an inner diameter of 100 μm (PolyMicro Technology, Phoenix, AZ) and a 5 μm diameter tip pulled with a P–2000 capillary puller (Sutter Instrument Company, Novato, CA) was packed with 6.5 cm 5 μm Aqua C_{18} reverse phase material (Phenomenex, Ventura,

CA), 3.5 cm 5 μm Partisphere strong cation exchanger (Whatman, Clifton, NJ) and another 2.5 cm 5 μm Aqua C_{18} (in this order from the tip). The sample was pressure loaded onto the column.

In the event of irreversible column clogging, the 6.5 cm 5 μm Aqua C_{18} separation phase was replaced by an inline microfilter assembly (UpChurch Scientific, Oak Habour, WA) and a 250 μm ID fused silica collection capillary to reduce the overall back pressure. A 6.5 cm 5 μm Aqua C_{18} separation phase was spliced onto the setup after completion of loading. We noted that the presence of EDTA in the sample may increase the risk of clogging events.

The sample–loaded column was placed in line between a HP–1100 quaternary HPLC pump (Agilent, Palo Alto, CA) and a LCQ–DecaXP electrospray ion trap mass spectrometer (ThermoElectron, Palo Alto, CA). Sample separation was achieved with a six step chromatography program modified according to McDonald et al. (2002). Solutions used were 5 % acetonitrile/0.1 % formic acid (buffer A), 80 % acetonitrile/0.1 % formic acid (buffer B) and 500 mM ammonium acetate/5 % acetonitrile/0.1 % formic acid (buffer C). Step 1 consisted of an 80 min gradient to 40 % buffer B, followed by a 10 min gradient to 100 % buffer B and 10 min of 100 % buffer B. Chromatography steps 2 to 5 followed the same pattern: 3 min of 100 % buffer A followed by a 2 min buffer C pulse, a 10 min gradient to 15 % buffer B and a 100 min gradient to 45 % buffer B. The buffer C percentages used were 5 %, 12.5 %, 25 % and 40 %, respectively, for the steps. The terminal step consisted of 3 min 100 % buffer A, 20 min 100 % buffer C, a 10 min gradient to 15 % buffer B and a 100 min gradient to 55 % buffer B. The flow rate through the column was approximately 150 nl/min.

Eluting peptides were electrosprayed into the mass spectrometer with a distally

applied spray voltage of 2.4 kV. The column eluate was continuously analyzed during the whole six step chromatography program. One full range mass–scan (400–1400 m/z) was followed by three data dependent MS/MS spectra at 35 % collision energy in a continuous loop.

Both HPLC pump and mass spectrometer where controlled by the `Xcalibur` software (ThermoElectron, Palo Alto, CA).

2.3.9 Data Analysis

In a first step, MS/MS spectra recorded by `Xcalibur` were analyzed for their charge state and controlled for data quality by `2to3` (Sadygov et al. 2002). The data were then searched by `SEQUEST` (Eng et al. 1994) against the translated *Saccharomyces* Genome Database (SGD; Cherry et al. 1998; release time stamped 05/23/03) supplemented with common contaminants (e. g., Keratins) on a `Linux` cluster composed of twenty 1.8 GHz Athlon CPUs (Racksaver, San Diego, CA). `DTASelect` (Tabb et al. 2002) filtered the `SEQUEST` results according to the following parameters: minimum `XCORR`s of 1.8, 2.5 and 3.5 for singly, doubly and triply charged precursor ions, respectively, minimum ΔC_n of 0.08, and a minimum requirement of two peptides per protein.

The resulting data was annotated and sorted with the `Python` script `RAYzer`. Annotation was added from SGD annotation tables (Cherry et al. 1998; table release time stamped 06/07/03) and interaction data curated by the MIPS Comprehensive Yeast Genome Database (MIPS CYGD; Mewes et al. 1997, 2002; release time stamped 04/29/03), the General Repository for Interaction Datasets (GRID; Breitkreutz et al. 2003; release 1.0) and the Yeast Protein Database (YPD; Garrels 1996;

as of 06/09/03). Based on known interaction annotation and the frequency of appearance in a reference data set containing one representative experiment for every tagged open reading frame in this study ($n = 22$), the data were then sorted into three tables: previously reported interactors retrieved in the experiment, potential new interacting proteins detected and likely contaminants (see supplementary online material). Proteins recovered in greater than 20 % of the experiments in the reference data set were automatically considered contaminants (see section 2.5).

2.4 Results

2.4.1 HPM–Tag

We constructed a bipartite affinity tag composed of nine histidines and nine myc–epitopes separated by two PreScission protease (Cordingley et al. 1990; and Walker et al. 1994) cleavage sites (HPM tag, figure 2.1; see section 2.3). Homologous recombination enables chromosomal integration of the PCR–amplified cassette in *Saccharomyces cerevisiae his3* strains at the 3' end of open reading frames targeted for affinity purification.

Using this cassette we tagged a test set of 25 gene products involved in transcription and progression through mitosis (see supplementary material, table 2) and established a variant of the "tandem affinity purification" (TAP) protocol (Rigaut et al. 1999) that employs affinity chromatography on a 9E10 monoclonal antibody resin followed by elution with PreScission Protease and adsorption to Ni–NTA resin (see section 2.3). For simplicity's sake we refer to our protocol as "TAP," even though

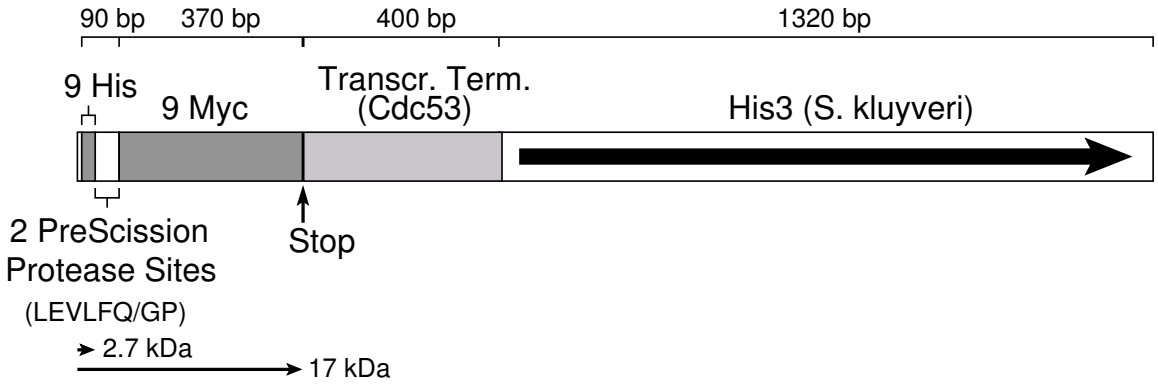

Figure 2.1 Schematic Representation of the HPM Tag. Nine histidines are separated from nine consecutive myc epitopes by two PreScission protease cleavage sites. The transcriptional terminator downstream of the Stop codon is from the *CDC53* locus. Chromosomal integration of the cassette can be selected for by restoring histidine prototrophy to *his3* mutant *S. cerevisiae* strains.

our tandem tag design requires a different purification protocol. Preliminary mass spectrometrical analyses showed that the eluates from the 9E10 resin still retained a high level of contaminating protein background (data not shown), and thus subsequent analyses were performed only on samples that were subjected to the complete TAP protocol. A representative SDS–PAGE analysis of the purification of four gene products is shown in figure 2.2.

The effectiveness and reproducibility of our overall approach was evaluated by analyzing the intensively studied histone acetyltransferase (HAT) Gcn5p (see fig. 2.3). Of the 23 previously reported interactors that were identified here, our experiments captured 15 (65 %) in all three replicates and an additional 5 (22 %) in two out of three attempts, including 18 known members of the SAGA/SLIK and ADA–HAT complexes (Sanders et al. 2002; Eberharter et al. 1999; Grant et al. 1998; and Pray-Grant et al. 2002). The majority of these validated partners ranked at the top of the list when the recovered proteins were sorted based on the size–normalized number of unique peptides sequenced per protein. These data indicate that TAP–MudPIT shows a high degree of reproducibility and robustness independent of fluctuations in

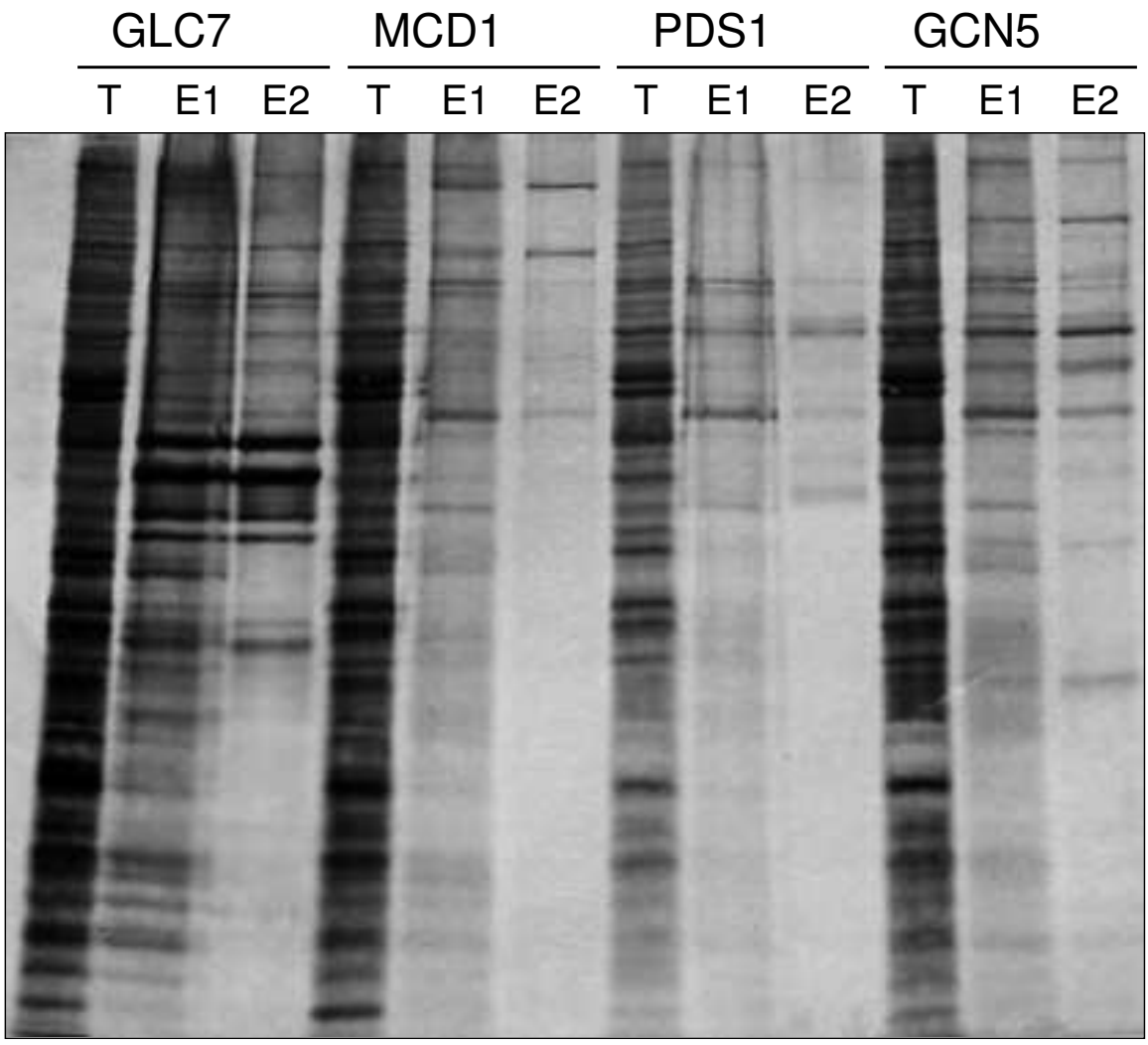

Figure 2.2 SDS–Polyacrylamide Gel Analysis of Glc7p–HPM, Mcd1p–HPM, Pds1p–HPM and Gcn5p–HPM Affinity Purifications. T: 2.5 μg total cell extract protein. E1: 7 % of material eluted by PreScission protease digest from α–myc antibody beads. E2: 20 % of EDTA eluate from the second affinity purification resin (Ni–NTA).

the sample quality of the individual experiment (see, e. g., varying peptide recovery for the bait in fig. 2.3).

Previous reports employed the original bipartite TAP tag and a modified TAP tag for tandem affinity purification (Rigaut et al. 1999; Gavin et al. 2002; and Cheeseman et al. 2001). A direct comparison of Gcn5p–TAP, Gcn5p–modified TAP,

Figure 2.3 Reproducibility of Results Between Independent Gcn5p–HPM TAP–MudPIT Experiments. Samples were prepared and analyzed as described in 2.3. Column "Known Interactor" indicates whether the gene product is a previously known Gcn5p interactor according to MIPS, GRID and YPD. Column "Gene Product" represents the name of the protein according to SGD. Red, yellow and plain background indicate recovery of the protein in three, two or one experiment out of three, respectively. Column "Frequency in Reference Set" lists the frequency with which the gene product was retrieved in the complete data set ($n = 22$). Column "Length" represents the length of the ORF in amino acids according to SGD. Columns "Exp. 1–3" list the number of unique and total peptide hits assigned to the ORF for each of the three experiments. Gene products are listed in descending order starting with the highest average length–normalized number of unique peptide identifications. Data for highly homologous ORFs with identical length, identical peptide representation across experiments and identical frequency in the reference set have been merged. Ty–element related ORFs have been excluded from the analysis.

Known Interactor?	Gene Product	Frequency in Reference Set	Length (AA)	Exp. 1 Unique (All)	Exp. 2 Unique (All)	Exp. 3 Unique (All)
Bait	Gcn5p	4.55%	439	41 (41)	25 (25)	65 (65)
	YCR082Wp	4.55%	128	13 (13)	5 (5)	16 (16)
×	Ada2p	4.55%	434	22 (22)	15 (15)	36 (36)
×	Ngg1p	4.55%	702	30 (30)	29 (29)	53 (53)
×	Sgf29p	4.55%	259	19 (19)	4 (4)	17 (17)
	Rps22A/Bp	90.91%	130	6 (6)	3 (3)	6 (6)
	Rpl28p	68.18%	149	6 (6)	3 (3)	6 (6)
	Rpl2A/Bp	63.64%	254	7 (7)	5 (5)	10 (10)
×	Ahc1p	4.55%	566	16 (16)	7 (7)	22 (22)
×	Taf5p	4.55%	798	16 (16)	10 (10)	35 (35)
	Rps18A/Bp	68.18%	146	2 (2)	6 (6)	3 (3)
×	Hfi1p	4.55%	488	9 (9)	3 (3)	23 (23)
×	Taf9p	4.55%	157	2 (2)	2 (2)	7 (7)
×	Spt8p	4.55%	602	11 (11)	6 (6)	23 (23)
×	Spt20p	4.55%	604	8 (8)	6 (6)	26 (26)
×	Taf6p	4.55%	516	6 (6)	5 (5)	23 (23)
×	Taf10p	4.55%	206	–	4 (4)	9 (9)
	Rpl32p	63.64%	130	2 (2)	2 (2)	4 (4)
	Npl3p	95.45%	414	9 (9)	8 (8)	8 (8)
	Rps4A/Bp	59.09%	261	2 (2)	3 (3)	10 (10)
×	Taf12p	4.55%	539	7 (7)	6 (6)	16 (16)
×	Adh1p	86.36%	348	6 (8)	10 (14)	2 (4)
	Tef1/2p	90.91%	458	5 (5)	11 (11)	7 (7)
	Rps3p	54.55%	240	4 (4)	3 (3)	5 (5)
	Rps16A/Bp	27.27%	143	2 (2)	2 (2)	3 (3)
×	Sgf73p	4.55%	657	10 (10)	4 (4)	18 (18)
×	Spt3p	4.55%	337	4 (4)	–	12 (12)
	Rpp2Ap	18.18%	106	2 (2)	–	3 (3)
	Rps13p	22.73%	151	–	2 (2)	5 (5)
	Rps24A/Bp	22.73%	135	–	3 (3)	3 (3)
	Rpl19A/Bp	50.00%	189	–	6 (6)	2 (2)
	Rps15p	54.55%	142	–	3 (3)	3 (3)
×	Fba1p	68.18%	359	6 (6)	9 (9)	–
	Rps20p	31.82%	121	–	3 (3)	2 (2)
	Cts2p	81.82%	511	6 (6)	4 (4)	11 (11)
	Ugp1p	59.09%	499	8 (8)	12 (12)	–
	Rps6A/Bp	36.36%	261	3 (3)	2 (2)	5 (5)
	Rpl30p	22.73%	105	2 (2)	2 (2)	–
	Rpl14A/Bp	27.27%	138	2 (2)	–	3 (3)
×	Spt7p	4.55%	1332	9 (9)	10 (10)	28 (28)
	Sro9p	95.45%	466	4 (4)	6 (6)	5 (5)
	Rpl3p	40.91%	387	4 (4)	3 (3)	5 (5)
	YPL047Wp	4.55%	99	–	–	3 (3)
	Shm1p	63.64%	565	8 (8)	9 (9)	–
×	Ubp8p	4.55%	471	–	–	13 (13)
	Rpl10p	40.91%	221	–	–	6 (6)
	Sds22p	13.64%	338	–	–	9 (9)
	Cdc19p	50.00%	500	2 (2)	9 (10)	2 (2)
	Rpl38p	4.55%	78	2 (2)	–	–
	Rps8A/Bp	27.27%	261	2 (2)	2 (2)	2 (2)
	Rpl11A/Bp	9.09%	174	2 (2)	2 (2)	–
	Hta1/2p	4.55%	132	3 (3)	–	–
	Glc7p	13.64%	312	–	–	7 (7)
	Rpl25p	54.55%	142	–	2 (2)	1 (1)
	Rpl9Ap	95.45%	191	1 (10)	1 (4)	2 (10)
	Mis1p	90.91%	975	4 (4)	2 (2)	14 (16)
	Ssb1/2p	59.09%	613	2 (2)	–	10 (10)
	Psa1p	45.45%	361	3 (3)	–	4 (4)
	Rpp0p	45.45%	312	–	2 (2)	4 (4)
	Hyp2p	27.27%	157	3 (3)	0 (2)	–
	Rps25A/Bp	13.64%	108	–	–	2 (2)
	Rpp2Bp	9.09%	110	–	–	2 (2)
	Rps29Bp	18.18%	56	–	–	1 (2)
	Rpl31Ap	95.45%	113	1 (7)	0 (3)	1 (8)
	Rpl9Bp	90.91%	191	1 (10)	1 (4)	1 (9)
	Ura7p	13.64%	579	–	–	9 (10)
	Vma2p	22.73%	517	3 (3)	–	5 (5)
	Pfk1p	63.64%	987	4 (4)	7 (7)	4 (4)
	Rps17A/Bp	22.73%	136	–	2 (2)	–
×	Tra1p	4.55%	3744	11 (11)	7 (7)	35 (35)
	Rps31p	31.82%	152	2 (4)	0 (4)	0 (4)
	Svp26p	0.00%	228	–	3 (3)	–
	Tdh1p	95.45%	613	2 (6)	6 (13)	0 (3)
	Ypi1p	13.64%	155	–	–	2 (2)
	Rps11A/Bp	22.73%	156	–	–	2 (2)
	Rib4p	13.64%	169	–	2 (2)	–
	Rps2p	9.09%	254	3 (3)	–	–
	Acs2p	40.91%	683	–	8 (8)	–
×	Eno2p	22.73%	437	3 (7)	2 (3)	–
	Tdh3p	100.00%	613	2 (8)	3 (19)	2 (9)
	Msn4p	4.55%	630	2 (2)	–	5 (5)
	Rpl17Ap	45.45%	184	0 (2)	1 (5)	1 (3)

Continued . . .

Known Interactor?	Gene Product	Frequency in Reference Set	Length (AA)	Exp. 1 Unique (All)	Exp. 2 Unique (All)	Exp. 3 Unique (All)
	Rpl18A/Bp	31.82%	186	–	–	2 (2)
	Adh3p	50.00%	375	1 (2)	3 (4)	–
	Ilv1p	45.45%	576	–	3 (3)	3 (3)
	Rpl16Bp	27.27%	198	–	–	2 (2)
	Rpl16Ap	13.64%	199	–	–	2 (2)
	Rpl15Ap	36.36%	204	0 (2)	1 (2)	1 (3)
×	Pgk1p	0.00%	416	4 (4)	–	–
	Kcs1p	31.82%	1050	–	–	10 (10)
	Gua1p	9.09%	525	–	5 (5)	–
	Cdc33p	13.64%	213	–	2 (2)	–
	Pnc1p	18.18%	216	–	2 (2)	–
	Rpl1A/Bp	27.27%	217	–	–	2 (2)
	Rpl31Bp	90.91%	113	0 (6)	0 (3)	1 (8)
	YOR283Wp	4.55%	230	–	2 (2)	–
	Yef3p	54.55%	1044	–	9 (13)	–
	Sod2p	4.55%	233	–	–	2 (2)
	Shm2p	13.64%	469	–	4 (4)	–
	Trp5p	22.73%	707	–	6 (6)	–
	Tfp1p	36.36%	1071	2 (2)	7 (7)	–
	Prb1p	31.82%	635	2 (2)	3 (3)	–
	Rps1Ap	40.91%	255	2 (3)	0 (2)	0 (7)
	Rps1Bp	45.45%	255	1 (2)	0 (2)	1 (8)
	Ssa2p	100.00%	639	2 (12)	1 (8)	2 (15)
	Rpl8Ap	50.00%	256	0 (3)	0 (2)	2 (6)
	YJR023Cp	0.00%	133	–	1 (1)	–
	Ura2p	63.64%	2214	2 (2)	2 (2)	12 (12)
	Tub2p	13.64%	457	–	–	3 (3)
	Sec23p	18.18%	768	–	5 (5)	–
	Rpl21Ap	36.36%	160	–	0 (3)	1 (5)
	Trp3p	13.64%	484	–	–	3 (3)
	Sik1p	4.55%	504	3 (3)	–	–
	Adh2p	72.73%	348	2 (4)	0 (4)	0 (2)
	Hos3p	22.73%	697	4 (4)	–	–
	Rpl6Bp	27.27%	176	1 (3)	0 (2)	0 (3)
	Rpl6Ap	31.82%	176	1 (3)	0 (2)	0 (3)
	Act1p	18.18%	375	–	2 (2)	–
	Aro2p	9.09%	376	–	2 (2)	–
×	Pfk2p	36.36%	959	2 (2)	3 (3)	–
	Rpl13Bp	36.36%	199	0 (2)	0 (3)	1 (3)
	Yhb1p	18.18%	399	–	–	2 (2)
	Eno1p	18.18%	437	1 (5)	1 (2)	–
	Hsm3p	0.00%	480	–	2 (2)	–
×	Clu1p	45.45%	1277	5 (5)	–	–
	Rrb1p	13.64%	511	–	2 (2)	–
	Asn2p	22.73%	572	–	2 (5)	–
	Ssa1p	100.00%	642	2 (12)	0 (7)	0 (12)
	Tfc1p	4.55%	649	–	–	2 (2)
	Ppz2p	13.64%	710	–	–	2 (2)
	Gfa1p	9.09%	717	–	–	2 (2)
	Pbp1p	9.09%	722	2 (2)	–	–
	Rpl4Ap	68.18%	362	0 (5)	1 (8)	0 (6)
	Ysh1p	4.55%	779	–	–	2 (2)
	Imd1p	22.73%	403	0 (2)	0 (2)	1 (2)
	Eft1/2p	18.18%	842	–	–	2 (2)
	Sec24p	4.55%	926	–	2 (2)	–
	Imd3p	27.27%	523	0 (4)	0 (2)	1 (2)
	Asn1p	9.09%	572	–	1 (4)	–
	Tdh2p	100.00%	613	0 (6)	1 (17)	0 (7)
	YPL137Cp	9.09%	1276	–	–	2 (2)
	YHL035Cp	0.00%	1592	2 (2)	–	–
	Glt1p	13.64%	2145	–	2 (2)	–
	Sth1p	9.09%	1359	1 (2)	–	–
	Ssa4p	72.73%	642	0 (4)	0 (2)	0 (7)
	Imd2p	22.73%	523	0 (4)	0 (2)	–
	Rpl8Bp	50.00%	256	0 (3)	0 (2)	0 (4)
	Rps14Bp	22.73%	138	0 (2)	–	0 (3)
	Rps14Ap	22.73%	137	0 (2)	–	0 (3)
	Rpl21Bp	36.36%	160	–	0 (3)	0 (4)
	Anb1p	0.00%	157	–	0 (2)	–
	Rpl20Ap	45.45%	180	0 (3)	0 (3)	0 (8)
	Rpl17Bp	40.91%	184	0 (2)	0 (4)	0 (2)
	Imd4p	9.09%	524	0 (3)	–	–
	Rpl13Ap	36.36%	199	0 (2)	0 (3)	0 (2)
	Rpl20Bp	45.45%	174	0 (3)	0 (3)	0 (8)
	Rpl15Bp	36.36%	204	0 (2)	–	0 (2)
	Ade3p	9.09%	946	–	–	0 (2)
	Ssa3p	81.82%	649	0 (4)	0 (3)	0 (6)
	Rpl4Bp	68.18%	362	0 (5)	0 (7)	0 (6)
	Hef3p	18.18%	1044	–	0 (4)	–
	Rpl40A/Bp	18.18%	128	0 (2)	0 (5)	0 (4)
	Ubi4p	22.73%	381	0 (2)	0 (4)	0 (4)

and Gcn5p–HPM revealed that the set of previously known interactors identified with the different tags are well within the margins of variability between independent experiments performed with the HPM tag (table 2.1).

Remarkably, our comparative analysis of Gcn5p purifications yielded strong candidates for six new Gcn5p interactors. YCR082W, a nonessential gene product (Winzeler et al. 1999; and Giaever et al. 2002) with unknown function, was found in all five Gcn5p purifications but was not recovered with any of the other baits that we analyzed. YCR082W exhibits a two–hybrid interaction with Ahc1p (Uetz et al. 2000; and Ito et al. 2001), which together with Gcn5p is a member of the ADA histone acetyltransferase complex (Eberharter et al. 1999). Another candidate is Msn4p, a nonessential (Estruch and Carlson 1993; and Winzeler et al. 1999) major transcriptional regulator of stress responses (Treger et al. 1998). Msn4p was recovered in four of the five Gcn5p pull down experiments but was not recovered with any of the other baits. This finding is interesting in the light of evidence that promoters activated by Msn4p and its partner Msn2p show increased histone H4 acetylation (Deckert and Struhl 2001). Other potential interaction partners include YPL047W (present in two of the HPM purifications and the TAP purification), histones Hta1p/Hta2p and Imd4p (in TAP, modified TAP and one HPM pulldown). Other gene products recovered in more than two of the experiments are mostly ribosomal proteins that are likely contaminants. Finally, the interaction observed between Gcn5p and Swi1p in the TAP tag experiment was previously proposed only on the basis of their synthetically lethal genetic interaction (Pollard and Peterson 1997).

Table 2.1 Comparison of TAP–MudPIT Analyses Using Different Bipartite Affinity Tags to Gcn5p. Samples were prepared and analyzed as described in section 2.3. Column "Gene Product" represents the name of the gene product recovered and known to interact with Gcn5p according to GRID, MIPS and YPD. "Exp. 1–3" represent three independent affinity purifications of Gcn5p–HPM. "TAP tag" and "Mod. TAP tag" represent tandem affinity purification–MudPIT experiments performed with strains in which the *GCN5* locus was tagged with either the TAP (Rigaut et al. 1999) or modified TAP tag from Cheeseman et al. (2001). The numbers of unique peptides from each ORF that were sequenced are shown (with the total number of sequenced peptides in parentheses). The last column lists the frequency with which the gene product is found in the entire data set ($n = 22$). For example, a gene product found in association with a single bait has a frequency of 4.55 % (1/22). The GRID, MIPS, and YPD interaction databases contain 83 additional gene products classified as interacting with Gcn5p, but not recovered in our analyses.

Gene Product	HPM tag			TAP tag	Mod. TAP tag	Frequ. in Ref. Set
	Exp. 1	Exp. 2	Exp. 3			
Gcn5p	41 (41)	25 (25)	65 (65)	19 (19)	21 (21)	4.55 %
Ada2p	22 (22)	15 (15)	36 (36)	22 (22)	36 (36)	4.55 %
Adh1p	6 (8)	10 (14)	2 (4)	—	—	86.36 %
Ahc1p	16 (16)	7 (7)	22 (22)	7 (7)	31 (31)	4.55 %
Clu1p	5 (5)	—	—	—	—	45.45 %
Eno2p	3 (7)	2 (3)	—	—	—	22.73 %
Fba1p	6 (6)	9 (9)	—	—	—	68.18 %
Hfi1p	9 (9)	3 (3)	23 (23)	20 (20)	23 (23)	4.55 %
Ngg1p	30 (30)	29 (29)	53 (53)	43 (43)	68 (68)	4.55 %
Pfk2p	2 (2)	3 (3)	—	—	—	36.36 %
Pgk1p	4 (4)	—	—	—	—	0.00 %
Rpg1p	—	—	—	—	5 (5)	0.00 %
Sgf29p	19 (19)	4 (4)	17 (17)	21 (21)	32 (32)	4.55 %
Sgf73p	10 (10)	4 (4)	18 (18)	25 (25)	29 (29)	4.55 %
Spt20p	8 (8)	6 (6)	26 (26)	26 (26)	29 (29)	4.55 %
Spt3p	4 (4)	—	12 (12)	12 (12)	8 (8)	4.55 %
Spt7p	9 (9)	10 (10)	28 (28)	49 (49)	52 (53)	4.55 %
Spt8p	11 (11)	6 (6)	23 (23)	18 (18)	20 (20)	4.55 %
Swi1p	—	—	—	3 (3)	—	9.09 %
Taf10p	—	4 (4)	9 (9)	7 (7)	11 (11)	4.55 %
Taf12p	7 (7)	6 (6)	16 (16)	28 (28)	23 (23)	4.55 %
Taf5p	16 (16)	10 (10)	35 (35)	46 (46)	37 (37)	4.55 %
Taf6p	6 (6)	5 (5)	23 (23)	24 (24)	26 (26)	4.55 %
Taf9p	2 (2)	2 (2)	7 (7)	7 (7)	15 (15)	4.55 %
Tra1p	11 (11)	7 (7)	35 (35)	82 (82)	99 (99)	4.55 %
Ubp8p	—	—	13 (13)	17 (17)	18 (18)	4.55 %
Yap1p	—	—	—	6 (6)	12 (12)	0.00 %

Table 2.2a Potential New Interactors for a Test Set of HPM Tagged Proteins. Samples were prepared and analyzed as described in 2.3. Column "Known interactors—Total" lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. "Known interactors—Recovered" represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked " * " are reported to interact physically as well as genetically. Column "Potential new interactors" contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20 % of the baits analyzed ($n = 22$).

Bait	Known interactors			Potential new interactors
	Total	Recovered		
	phys./gen.	phys.	genet.	
Bim1p–HPM	6/57	1	—	Rpb2p, Rpl12A/Bp, Rpl22Ap, Rps25A/Bp, Rps29Ap, Rps5p, YGR161C–Cp
Cdc20p–HPM	12/3	6	—	Bub3p, Cct4p, Cct6p, Cct7p, Cct8p, Hef3p, Ilv6p, Pnc1p, Rfa1p
Chk1p–HPM	16/0	—	—	Act1p, Car2p, Gpd2p, Hht1p, Hht2p, Htb2p, Htb1p, Htz1p, Pnc1p
Cla4p–HPM	15/77	—	Rpl17Bp, Rpl17Ap, Rpl19Bp, Rpl19Ap	Pbp1p, Pre8p, Rpl36Ap, Rpl36Bp, Rpl7Ap, Rpl7Bp, Rpp2Ap, Rps2p, Sec23p, Skm1p, YBR225Wp, Yhb1p
Dbf2p–HPM	27/9	3	Dbf20p, Mob1p*	Adh5p, Caf20p, Car2p, Cdc33p, Emi2p, Gfa1p, Gly1p, Gpd2p, Hsp42p, Ilv6p, Pnc1p, Pro1p, Rib4p, Sec23p, Shm2p, Snf1p, Trp3p, Tub2p

2.4.2 Screening for Interactions

Having established the relative reproducibility of TAP–MudPIT and the comparability of the HPM tag to other available bipartite affinity tags, we set out to address three issues. First, we wished to determine what fraction of TAP–MudPIT experiments yield usable results. Second, we hoped to determine whether the parallel application of MudPIT to numerous baits would enable us to cull nonspecific contaminants by comparing protein identifications across multiple experiments. Third, we wanted to test whether it will be feasible for an investigator in a cell biology laboratory to work at the scale needed to dissect a biological pathway or process by systematic application of MudPIT to a few dozen gene products. To addresses these questions, we screened for new protein–protein interactions in a test set of 25 gene products involved in transcription and progression through mitosis. Table 2.2 summarizes the results and gives an overview of potential new interactors. The

Table 2.3b Potential New Interactors for a Test Set of HPM Tagged Proteins. Samples were prepared and analyzed as described in 2.3. Column "Known interactors—Total" lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. "Known interactors—Recovered" represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked " * " are reported to interact physically as well as genetically. Column "Potential new interactors" contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20 % of the baits analyzed ($n = 22$).

Gcn5p–HPM	99/12	18	Ngg1p*	Ade3p, Eft2p, Eft1p, Gfa1p, Glc7p, Msn4p, Ppz2p, Rpl16Ap, Rpp2Ap, Rpp2Bp, Rps25Ap, Rps25Bp, Rps29Bp, Sds22p, Sod2p, Tfc1p, Trp3p, Tub2p, Ura7p, YCR082Wp, Yhb1p, Ypi1p, YPL047Wp, YPL137Cp, Ysh1p
Glc7p–HPM	177/9	28	Ppz2p*, Ppz1p*, Reg1p*	Abf1p, Ade16p, Ade17p, Ahp1p, Bmh1p, Bmh2p, Ccr4p, Cka2p, Eno1p, Fun21p, Gal83p, Hsp60p, Imp2p, Mor1p, Pdc1p, Pgk1p, Pol2p, Rpp2Ap, Snf1p, Sol1p, Sol2p, YBR225Wp, YDR474Cp, YER158Cp, YGR237Cp, YHR097Cp, YPL137Cp
Ino4p–HPM	52/0	1	—	Act1p, Mdn1p, Pmd1p, Xrs2p

primers to tag CDC5 and ESS1 while TAP–MudPIT experiments for Bir1p–HPM and Nbp1p–HPM resulted in little or no recovery of the tagged baits themselves. Of the 21 "successful" purifications that yielded sequence coverage for the tagged bait, 20 of the experiments (95 %) yielded interacting proteins that are either true binding partners validated by other direct approaches, probable binding partners that display genetic interaction with the bait, or candidate binding partners that were found in association with only one bait. The Pho2p–HPM experiment yielded 'hits' only from proteins that were found associated with other, unrelated baits or were otherwise deemed to be likely contaminants.

The set of bait proteins evaluated in this study overlaps considerably with the Ho et al. (2002) effort. Figure 2.4 compares the retrieval of physical interactors for 13 gene products used as baits in both studies. Notably, in each case our approach identified at least as many or more of the previously–known binding partners of the

Table 2.4c Potential New Interactors for a Test Set of HPM Tagged Proteins. Samples were prepared and analyzed as described in 2.3. Column "Known interactors—Total" lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. "Known interactors—Recovered" represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked " * " are reported to interact physically as well as genetically. Column "Potential new interactors" contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20 % of the baits analyzed ($n = 22$).

Lte1p–HPM	48/12	5	—	Ade4p, Aro2p, Asc1p, Asn1p, Bcy1p, Bmh1p, Caf20p, Car2p, Cdc33p, Eft2p, Eft1p, Emi2p, Eno1p, Eno2p, Flo8p, Gad1p, Glk1p, Glt1p, Gly1p, Gpm1p, Gua1p, Hef3p, Hem1p, Hsp60p, Ilv6p, Lpd1p, Mkt1p, Nfs1p, Pbi2p, Pdc1p, Pgk1p, Pnc1p, Pro1p, Rax2p, Rib4p, Rpl23Ap, Rpl23Bp, Rps23Ap, Rps23Bp, Rps29Ap, Rps29Bp, Rps5p, Sec23p, Sec24p, Shm2p, Sod1p, Tpi1p, Tps3p, Vps1p, YDR348Cp, Yhb1p, YHL021Cp
Mad2p–HPM	11/10	2	—	Apl4p, Caf20p, Eno1p, Eno2p, Pdc1p, Pgk1p, Rrb1p, Trx2p, Ura7p, YOR283Wp
Mcd1p–HPM	17/8	3	Smc1p*, Trf4p	Bdf1p, Csm1p, Nuf2p, Not5p, Pom152p, Srm1p, Stu2p, YBL005W–Ap, YDR170W–Ap, YMR045Cp, YNL284C–Bp, YNL284C–Ap, YMR046Cp
Pds1p–HPM	4/1	1	Esp1p*	Azr1p, Ire1p, Mss1p, Swi3p
Pds5p–HPM	0/1	—	Mcd1p	Aro4p, Chs5p, Hal5p, Kem1p, Mss1p, Pbp1p
Pho2p–HPM	4/1	—	—	Rpl35Bp, Rpl35Ap, Rps5p, YBL005W–Ap, YDR170W–Ap, YDR261W–Bp, YGR161C–Cp, YJR026Wp, YOL103W–Ap, YML040Wp, YLR256W–Ap ,YLR227W–Ap, YLR157C–Ap, YJR028Wp, YMR045Cp, YNL284C–Bp

13 bait proteins. For eight of the baits, Ho et al. (2002) identified more putative interacting partners. However, since Ho et al. (2002) utilized single–step affinity purification of overproduced bait protein, additional interactions revealed only in that study should be considered as tentative, pending verification by independent methods.

The second issue that we addressed was the feasibility of using a filtering approach to cull nonspecific contaminants from the list of proteins identified in each

Table 2.5d Potential New Interactors for a Test Set of HPM Tagged Proteins. Samples were prepared and analyzed as described in 2.3. Column "Known interactors—Total" lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. "Known interactors—Recovered" represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked " * " are reported to interact physically as well as genetically. Column "Potential new interactors" contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20 % of the baits analyzed ($n = 22$).

Pho4p–HPM	11/1	—	—	Ade16p, Ade3p, Ape3p, Aro2p, Aro4p, Asn1p, Bbc1p, Bcy1p, Cct4p, Cct8p, Cdc33p, Cdc73p, Chs5p, Dbp2p, Dbp3p, Dig1p, Eap1p, Eft2p, Eft1p, Fas1p, Fun12p, Glk1p, Glt1p, Gly1p, Gua1p, Hef3p, Hom3p, Hrb1p, Hsp60p, Imd4p, Kem1p, Kri1p, Lys21p, Lys20p, Myo5p, Nfs1p, Nma1p, Nop1p, Nop58p, Nsr1p, Pab1p, Rpa135p, Rpa34p, Rpl11Bp, Rpl11Ap, Rpl12Bp, Rpl12Ap, Rpl16Ap, Rpl23Ap, Rpl23Bp, Rpl24Ap, Rpl24Bp, Rpl26Bp, Rpl26Ap, Rpl29p, Rpl34Ap, Rpl34Bp, Rpl35Bp, Rpl35Ap, Rpl36Ap, Rpl36Bp, Rpl38p, Rpl43Bp, Rpl43Ap, Rpl5p, Rpl7Ap, Rpl7Bp, Rpp1Ap, Rpp2Ap, Rpp2Bp, Rps12p, Rps19Bp, Rps19Ap, Rps2p, Rps23Ap, Rps23Bp, Rps25Ap, Rps25Bp, Rps27Bp, Rps27Ap, Rps29Ap, Rps29Bp, Rps5p, Rps7Ap, Rps7Bp, Rps9Ap, Rps9Bp, Rrb1p, Rrp5p, Rsp5p, Sec23p, Ses1p, Shm2p, Sik1p, Sin3p, Snf1p, Srm1p, Ste11p, Ste50p, Stm1p, Tsr1p, Tub1p, Tub2p, Tub3p, Ura7p, Utp7p, Vip1p, Vps1p, Vrp1p, YAR075Wp, YBL101W–Bp, YGR161W–Bp, YFL002W–Ap, YDR210W–Bp, YDR034C–Dp, YCL019Wp, YDR261W–Bp, YGL068Wp, YHR121Wp, YIL137Cp, YMR045Cp, YMR050Cp, YMR237Wp, YNL054W–Bp

TAP–MudPIT experiment. The idea is that nonspecific proteins should show up in a high fraction of experiments, whereas specific interactors should only show up in one or a small number of experiments (depending upon the degree of functional relatedness of the tagged genes in the query set). We found that proteins that were identified in five or more TAP–MudPIT experiments tended to have a high codon

Table 2.6e Potential New Interactors for a Test Set of HPM Tagged Proteins. Samples were prepared and analyzed as described in 2.3. Column "Known interactors—Total" lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. "Known interactors—Recovered" represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked " * " are reported to interact physically as well as genetically. Column "Potential new interactors" contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20 % of the baits analyzed ($n = 22$).

Rtt102p–HPM	2/0	—	—	Aro4p, Arp7p, Arp9p, Fyv6p, Gsy2p, Hsl1p, Hta2p, Hta1p, Htl1p, Ldb7p, Nfi1p, Nfs1p, Npl6p, Rim1p, Rpl35Bp, Rpl35Ap, Rpl36Ap, Rpl36Bp, Rpl43Bp, Rpl43Ap, Rps2p, Rps29Bp, Rrb1p, Rsc1p, Rsc2p, Rsc3p, Rsc4p, Rsc58p, Rsc6p, Rsc8p, Rsc9p, Sfh1p, Snf12p, Snf2p, Snf5p, Snf6p, Sth1p, Swi1p, Swi3p, Taf14p, YFL049Wp, YHR097Cp
Sds22p–HPM	45/0	4	—	Nip100p, Ppz1p, Snf1p, Stu1p, Vps8p, YBL010Cp
Snf2p–HPM	164/13	11	—	Chs5p, Pab1p, Rpl11Bp, Rpl11Ap, Rpl16Ap, Rpl26Ap, Rpl26Bp, Rpl34Ap, Rpl34Bp, Rpl35Bp, Rpl35Ap, Rpl36Ap, Rpl36Bp, Rps12p, Rps2p, Rtt102p, Sth1p, Stm1p, YDL053Cp, YGR161C–Cp
Spo12p–HPM	18/5	1	—	Act1p, Ado1p, Ahp1p, Ald6p, Azr1p, Bmh1p, Cpr1p, Cys3p, Eft2p, Eft1p, Eno1p, Eno2p, Gpm1p, Hsp12p, Hsp42p, Hxk2p, Pdc1p, Pgi1p, Pgk1p, Rhr2p, Rps12p, Rps19Bp, Rps19Ap, Tif2p, Tif1p, Trp3p, Trx2p, Yhb1p, YNL134Cp, YPL257W–Bp
Yak1p–HPM	75/0	3	—	Caf20p, Glt1p, Gly1p, Hef3p, Kem1p, Nfs1p, Rib4p, YJL206Cp
YHR115Cp–HPM	17/0	8	—	Dbp3p, Gcd11p, Jip5p, Mkt1p, Sec16p, YBL101W–Bp, YLR410W–Bp, YGR161W–Bp, YFL002W–Ap, YDR210W–Bp, YDR034C–Dp, YCL019Wp, YJR026Wp, YOL103W–Ap, YML040Wp, YLR256W–Ap, YLR227W–Ap, YLR157C–Ap, YJR028Wp

adaptation index (Sharp and Li 1987), which is a rough measure of abundance (Jansen et al. 2003, data not shown). Based on this correlation, we automatically considered proteins found in more than five experiments to be probable contaminants. A similar filtering approach was employed by Gavin et al. (2002) and Ho

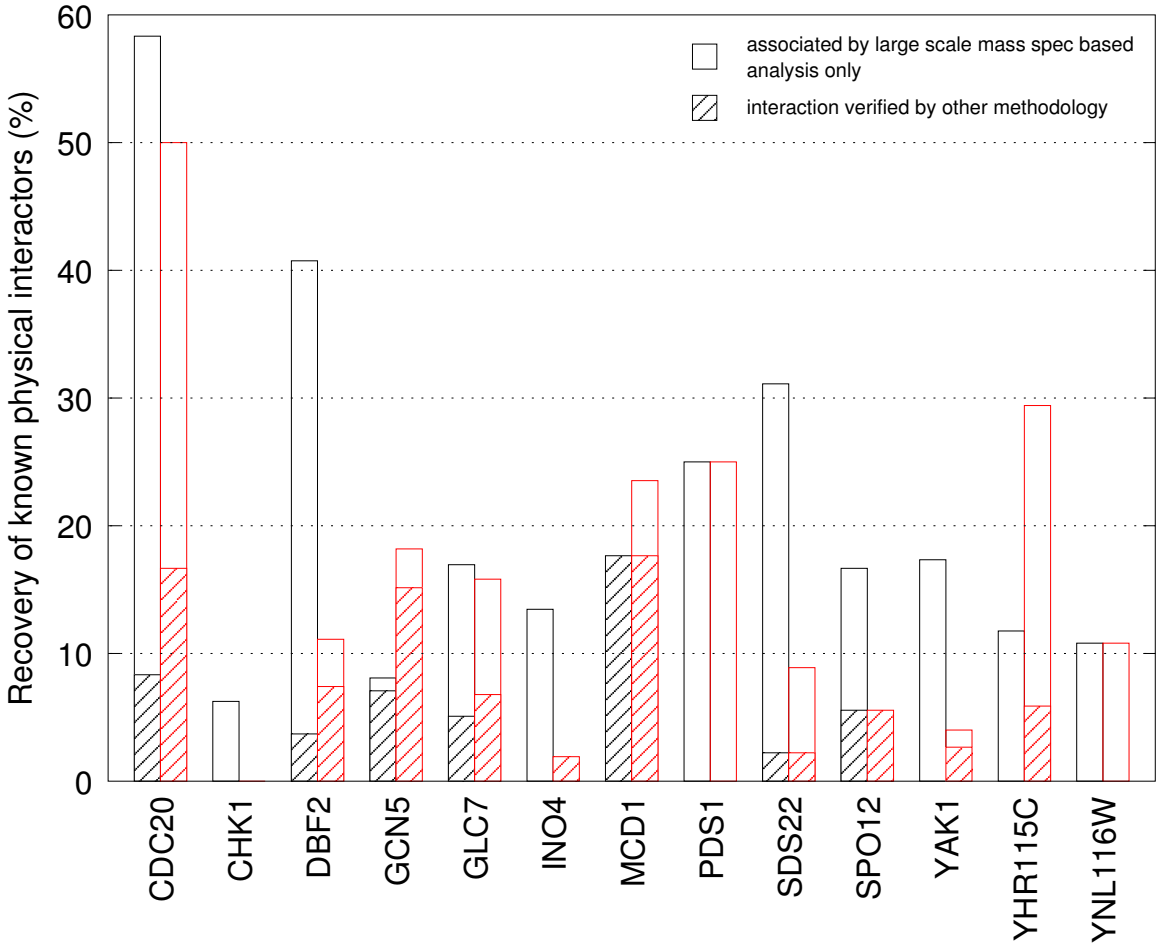

Figure 2.4 Comparison with Ho et al. (2002). Comparison of the data set presented here (red) with that of Ho et al. (2002) (black). ORFs listed were used as bait proteins in both studies. Bars represent the percentage of previously known interacting partners (as reported in MIPS CYGD, GRID and YPD) that was recovered in each experiment. Note that the set of interacting partners listed in these databases includes those reported by Ho et al. (2002). Empty bars represent percentage of gene products reported as interactors only by large scale mass spectrometric analysis whereas hatched bars represent interactions established or verified by other methods.

et al. (2002), but since their data–sets were much larger they were able to employ lower thresholds.

To showcase the possibility of identifying new potential interacting partners in any given TAP–MudPIT experiment, we analyzed in more detail our results for Snf2p–HPM. Snf2p is a subunit of the Swi/Snf complex and founding member of the

ATP–dependent family of chromatin remodeling factors (Fry and Peterson 2001). TAP–MudPIT analysis of Snf2p–HPM yielded eight of the nine known members of this complex (Cote et al. 1994; Henry et al. 1994; and Cairns et al. 1998; Arp7p, Arp9p, Snf5p, Snf6p, Swi1p, Swi3p, Snf12p, Taf14p; missing: Snf11p) as well as YFL049W, a protein of unknown function reported to copurify with Snf2p via its interaction with Snf5p (Gavin et al. 2002). A prominent Snf2p–HPM copurifying protein that was not commonly retrieved by other baits was Rtt102p, a protein of unknown function, whose inactivation results in a slight increase in Ty1 retrotransposon mobility (Scholes et al. 2001). To check whether the interaction of Snf2p with Rtt102p was reciprocal, we tagged the Rtt102p locus with sequences encoding the HPM epitope, and performed TAP–MudPIT analysis for Rtt102p–HPM. This experiment yielded all of the subunits of the Swi/Snf chromatin remodeling complex that copurified with Snf2p–HPM (see above), as well as all subunits of the RSC chromatin remodeling complex (Scholes et al. 2001; Npl6p, Rsc1p, Rsc2p, Rsc3p/Rsc30p, Rsc4p, Rsc58p, Rsc6p, Rsc8p, Rsc9p, Sfh1p, Sth1p). YFL049W copurified with Rtt102p–HPM as well as with Snf2p–HPM, further strengthening the case that it is a bona fide Swi/Snf component. These results suggest that Rtt102p, like Arp7p and Arp9p (Cairns et al. 1998; and Peterson et al. 1998), is specifically associated with the Swi/Snf and RSC chromatin remodeling complexes, and may be an integral component of both.

Knockouts of Swi/Snf complex members show reduced growth on sucrose/antimycin, galactose/antimycin and glycerol (Peterson et al. 1998). When tested for growth on these carbon sources, a *rtt102*Δ strain grew similar to wild type on glucose, sucrose/antimycin and galactose/antimycin, but exhibited a severe growth phenotype on glycerol (see fig. 2.5), further supporting a functional Rtt102p–Swi/Snf

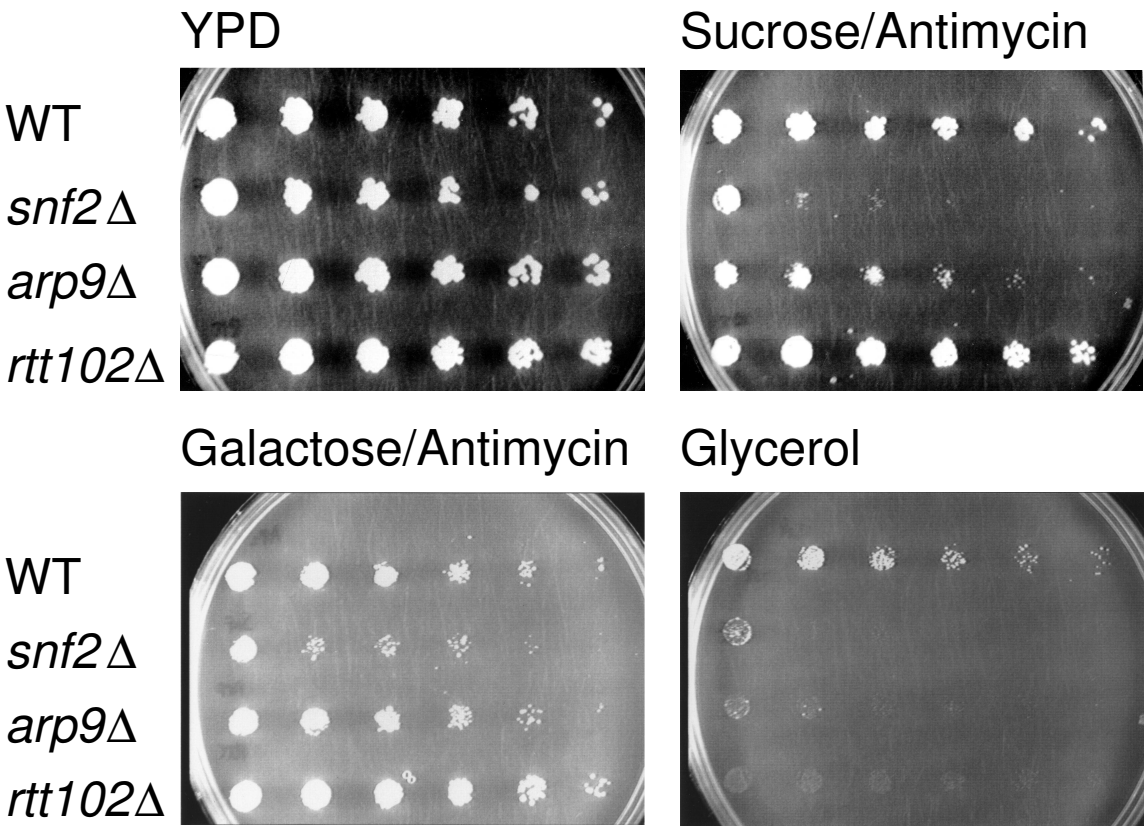

Figure 2.5 An *rtt102*Δ Strain Partially Recapitulates the Phenotype of Mutants Lacking Swi/Snf Complex. "WT" is W303 *pep4*Δ::*TRP1*, *bar1*Δ::*HISG* (RJD 415). "snf2Δ" is RJD 415, *snf2*Δ::*HIS3* (RJD 2566). "arp9Δ" is RJD 415, *arp9*Δ::*HIS3* (RJD 2567). "rtt102Δ" is RJD 415, *rtt102*Δ::*HIS3* (RJD 2568). Media compositions are: 1 % yeast extract, 2 % peptone and 2 % final concentration of glucose, sucrose, galactose or glycerol. Sucrose and galactose containing media were supplemented with 1 μg/ml antimycin.

connection.

2.5 Discussion

A key goal of proteomics research is to identify and characterize protein interaction networks. Several approaches have been taken to achieve this goal, including genome–wide two–hybrid analyses and protein chip–based approaches (Uetz et al. 2000; Ito et al. 2001; and Zhu et al. 2001). A limitation of both of these methods

is that they primarily reveal binary interactions. Large–scale mass spectrometric analyses of affinity–purified protein complexes have been reported by two different groups (Gavin et al. 2002; and Ho et al. 2002). Whereas this approach bypasses some of the key limitations of two–hybrid and protein chip assays, the efforts reported so far were based on gel separation of purified proteins, which both greatly increased the number of mass spectrometry runs required to analyze each bait and limited the dynamic range to proteins that could be stained and visualized on the same gel. Indeed, both efforts were carried out in an industrial context that can not be readily adapted to a conventional molecular/cellular biology laboratory. We believe this is an important issue, because unlike the genomic sequence, the protein interactions that exist in a cell or organism are not a finite and bounded set that can be determined as a complete "reference" knowledge set. Rather, their most important feature is that they change as a function of intracellular and extracellular signals and learning how they change is essential for probing the cellular processes of interest. Thus, to characterize fully the protein interaction networks in a cell and their dynamic changes over time, it will be necessary to perform multiple analyses under different conditions and in different genotypes. In this sense, mass spectrometry–based proteomics resembles microarray–based transcriptomics. This fact underscores the need for simple, reproducible, rapid, portable (i. e. can be performed outside of a specialized mass spectrometry environment), yet powerful methods for exploring protein interaction networks.

We show here that a combination of double affinity purification and multidimensional capillary chromatography in line to mass spectrometry (TAP–MudPIT) fulfills these criteria. TAP–MudPIT can be applied to rapidly identify interacting proteins for any given bait in a single mass spectrometry analysis. Using this

approach, a single investigator working with a single mass spectrometer and performing the complete protocol from affinity purification to data analysis can readily screen 20 samples per month (i. e. 20 different baits or one bait evaluated under 20 different conditions). Thus, it is feasible for a single investigator to perform, in a reasonable time frame, a thorough analysis of a focused collection of baits that define a particular organelle, pathway, or process.

It should also be noted that in addition to protein identification, the TAP–MudPIT approach enables the parallel analysis of posttranslational modifications (Cheeseman et al. 2002).

Although an exhaustive analysis of every one of the 22 TAP–MudPIT experiments that we performed (21 from the original collection of baits plus Rtt102p) is beyond the scope of this paper, we wish to highlight several interesting points. First, our analysis of Swi2p/Snf2p identified a new interacting partner, Rtt102p, which is remarkable given the large body of work that has already been performed on this extensively characterized protein and its interacting partners. Second, we uncovered TRF4 as a candidate partner of the cohesin Mcd1p/Scc1p. Trf4p was originally reported to function as an alternative DNA polymerase that mediates sister chromatid cohesion (Wang et al. 2000), but this proposal has been the subject of controversy following the report that Trf4p can catalyze polymerization of poly(A) tails on mRNA transcripts (Saitoh et al. 2002). Third, Bub3p was found as a Cdc20p–associated protein and Mcd1p/Scc1p was found as a Pds5p–associated protein. Although these pairs of proteins were already known to function together in mitotic checkpoint signaling and sister chromatid cohesion, respectively, a physical association of the yeast proteins has not been reported. Finally, in addition to Trf4p, Mcd1p/Scc1p retrieved the Csm1p subunit of monopolin and the Nuf2p

subunit of the Tid3p/Nuf2p/Spc24p/Spc25p centromere–binding complex (Janke et al. 2001). Both interactions are excellent candidates to subserve a role in chromosome segregation given the known functions of the proteins involved.

Analysis of Rtt102p, identified here as a Swi2p/Snf2p interactor, illustrated the power of this system for making fast and simple first–order interaction validation. This was accomplished by a reciprocity test, in which Rtt102p was shown to specifically retrieve Swi2p and other known components of the Swi/Snf complex. Because this is an independent determination, it provides a more convincing confirmation for an interaction than a mere repetition of the initial measurement. The experiment also illustrates how TAP–MudPIT can be used for directed interaction "walks" (Seol et al. 2001), in this case showing that Rtt102p also interacts with, or is a component of, the RSC chromatin remodeling complex.

Whereas TAP–MudPIT is sufficiently robust to be applied in a nonspecialized environment, two substantial problems remain to be addressed. First, the interpretation of the data that is generated would benefit from improvement. The combination of `2to3` (Sadygov et al. 2002), `SEQUEST` (Eng et al. 1994) and `DTASelect` (Tabb et al. 2002) enables analysis and display of raw mass spectrometrical data. What are missing, however, are tools that simplify interpretation of the massive amount of data generated by the analysis of even a protein interaction network of even modest size. In particular, seperating good candidates for novel interaction partners from the contaminating chaff is a major challenge. We followed the approach used by Gavin et al. (2002) and Ho et al. (2002), by excluding from consideration any protein that was found associated with more than 20 % of the baits analyzed (the comparable thresholds were 3 % in Ho et al. (2002), 3.5 % in Gavin et al. 2002). When applied to the proteins found in all three independent Gcn5p–HPM TAP–

MudPIT analyses shown in fig. 2.3, our filter threshold retains only the previously known interactors and the potential new Gcn5p–interacting protein YCR082Wp. A problem with excluding candidates by this criterion is that we were not using an unbiased reference data set. Since the proteins that we analyzed are all involved in either transcription or mitosis, it is possible that some true interacting proteins were improperly excluded.

The complete data set contains a total of 464 potential interactions passing the requirement of being associated with less then 20 % of the baits analyzed. However, this subset includes ribosomal, cytoskeletal and other proteins, that, due to their abundance, have a high probability of being contaminants. Discarding Ty–Element related proteins and applying a filter that allows a maximum CAI of 0.6 eliminates these problematic candidates and reduces the number of potential new interaction partners identified to 279.

In addition to "post hoc" approaches, honing the purification protocol and making it more stringent may lessen the problem posed by contaminating proteins. However, this comes at the possible expense of disrupting specific interactions. When analyzing a single bait under varying conditions, optimizing the purification may greatly improve the specificity of the purification, but optimization becomes a daunting task when dealing with multiple baits.

The second major problem arises from the databases used to biologically annotate the gene products identified by MudPIT. Given the amount of data produced by a MudPIT experiment, machine readability of data bases is of great value. Unfortunately, of the data bases used in this study only the regularly updated data in SGD and MIPS CYGD are readily accessible in an automated manner (`ftp`). GRID data can be manually downloaded in a tab delimited file, but YPD does not

allow any such access, and thus requires manual merging of its annotation data into a computationally annotated data set.

As more and more large–scale analyses are performed, an issue that looms large for the future is how to evaluate the quality of the datasets. Even relatively small–scale analyses like the one reported here are prone to produce false positives (e. g., the large number of ribosomal proteins classified as potential interactors for Pho4p in table 2.2). As a specific example of this problem, consider Adh1p (alcohol dehydrogenase). Adh1p is annotated in YPD as a protein in complex with Gcn5p and Snf2p, because Adh1p was reported to copurify with these proteins in TAP experiments using Spt15p and Med2p (Gcn5p) or Enp1p (Snf2p) as bait proteins (Gavin et al. 2002). However, given that we found Adh1p associated with 86 % of our baits, it is most likely a common contaminant that nevertheless cleared the filter imposed by Gavin et al. (2002). An important challenge is to generate databases that express the likelihood that a protein–protein interaction is relevant based on the number of independent analyses (and methods) upon which the conclusion is based.

In conclusion, we report the application of TAP–MudPIT—tandem–affinity purification coupled with multidimensional capillary chromatography in line to mass spectrometry—to identify binding partners for a set of 22 budding yeast proteins involved in gene regulation or progression through mitosis. Our analysis uncovered 102 previously known and 279 potential physical interactions. TAP–MudPIT is simple, rapid, reproducible and can be carried out in a traditional cell biology laboratory. The simplicity and power of this method enables a depth of analysis that will facilitate thorough characterization of protein interaction networks.

2.6 Acknowledgements

This work was supported by Howard Hughes Medical Institute (R. J. D.), Beckman Institute (B. J. W.), National Institute of Health grant RR11823 to J. R. Y. and W. H. M. and Department of Energy grant DE–FG03–01ER63231 to B. J. W. and R. J. D. We thank Erin O'Shea (UCSF) for the gift of RJD 2067. J. G. thanks all members of the Yates lab (The Scripps Research Institute) and Eberhard Dürr (Sidney Kimmel Cancer Center) for introduction into mass spectrometry and `SEQUEST` and D. A. Gold, E. E. Griffin and A. Mah (all Caltech) for critical reading of the manuscript.

2.7 Note Added in Proof

The reproducibility of MudPIT applied to whole cell extracts was recently reported by Washburn et al. (2003).

2.8 References

Breitkreutz, B. J., Stark, C. and Tyers, M. (2003). The GRID: The General Repository for Interaction Datasets. *Genome Biol*, 4(3):R23.

Cairns, B. R., Erdjument-Bromage, H., Tempst, P., Winston, F. and Kornberg, R. D. (1998). Two actin–related proteins are shared functional components of the chromatin–remodeling complexes RSC and SWI/SNF. *Mol Cell*, 2(5):639–51.

Cheeseman, I. M., Anderson, S., Jwa, M., Green, E. M. and Kang, J. et al. (2002). Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. *Cell*, 111(2):163–72.

Cheeseman, I. M., Brew, C., Wolyniak, M., Desai, A. and Anderson, S. et al. (2001). Implication of a novel multiprotein Dam1p complex in outer kinetochore function. *J Cell Biol*, 155(7):1137–45.

Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A. and Dwight, S. S. et al. (1998). SGD: *Saccharomyces* genome database. *Nucleic Acids Res*, 26(1):73–9.

Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M. and Colonno, R. J. (1990). Substrate requirements of human rhinovirus 3C protease for peptide cleavage *in vitro*. *J Biol Chem*, 265(16):9062–5.

Cote, J., Quinn, J., Workman, J. L. and Peterson, C. L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*, 265(5168):53–60.

Deckert, J. and Struhl, K. (2001). Histone acetylation at promoters is differentially affected by specific activators and repressors. *Mol Cell Biol*, 21(8):2726–35.

Eberharter, A., Sterner, D. E., Schieltz, D., Hassan, A. and Yates III, J. R. et al. (1999). The ADA complex is a distinct histone acetyltransferase complex in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 19(10):6621–31.

Eng, J. K., McCormack, A. L. and Yates III, J. R. (1994). An approach to correlate tandem mass–spectral data of peptides with amino–acid sequences in a protein database. *J Am Soc Mass Spectr*, 5(11):976–989.

Estruch, F. and Carlson, M. (1993). Two homologous zinc finger genes identified by multicopy suppression in a SNF1 protein kinase mutant of *Saccharomyces cerevisiae*. *Mol Cell Biol*, 13(7):3872–81.

Evan, G. I., Lewis, G. K., Ramsay, G. and Bishop, J. M. (1985). Isolation of monoclonal antibodies specific for human c–myc proto–oncogene product. *Mol Cell Biol*, 5(12):3610–6.

Fry, C. J. and Peterson, C. L. (2001). Chromatin remodeling enzymes: Who's on first?. *Curr Biol*, 11(5):R185–97.

Garrels, J. I. (1996). YPD—A database for the proteins of *Saccharomyces cerevisiae*. *Nucleic Acids Res*, 24(1):46–9.

Gavin, A. C., Bosche, M., Krause, R., Grandi, P. and Marzioch, M. et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868):141–7.

Giaever, G., Chu, A. M., Ni, L., Connelly, C. and Riles, L. et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, 418(6896):387–91.

Grant, P. A., Schieltz, D., Pray-Grant, M. G., Yates III, J. R. and Workman, J. L. (1998). The ATM–related cofactor Tra1 is a component of the purified SAGA complex. *Mol Cell*, 2(6):863–7.

Graumann, J., Dunipace, L. A., Seol, J. H., McDonald, W. H. and Yates III, J. R. et al. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics*, 3(3):226–37.

Henry, N. L., Campbell, A. M., Feaver, W. J., Poon, D. and Weil, P. A. et al. (1994). TFIIF–TAF–RNA polymerase II connection. *Genes Dev*, 8(23):2868–78.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D. and Moore, L. et al. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, 415(6868):180–3.

Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. *J Bacteriol*, 153(1):163–8.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M. and Hattori, M. et al. (2001). A comprehensive two–hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA*, 98(8):4569–74.

Janke, C., Ortiz, J., Lechner, J., Shevchenko, A. and Shevchenko, A. et al. (2001). The budding yeast proteins Spc24p and Spc25p interact with Ndc80p and Nuf2p at the kinetochore and are important for kinetochore clustering and checkpoint control. *EMBO J*, 20(4):777–91.

Jansen, R., Bussemaker, H. J. and Gerstein, M. (2003). Revisiting the codon adaptation index from a whole–genome perspective: Analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Res*, 31(8):2242–51.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. and Mize, G. J. et al. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, 17(7):676–82.

Longtine, M. S., McKenzie III, A., Demarini, D. J., Shah, N. G. and Wach, A. et al. (1998). Additional modules for versatile and economical PCR–based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast*, 14(10):953–61.

McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S. and Barnes, G. et al. (1997). Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low–femtomole level. *Anal Chem*, 69(4):767–76.

McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J. and Yates III, J. R. (2002). Comparison of three directly coupled hplc ms/ms strategies for identification of proteins from complex mixtures: Single–dimension lc–ms/ms, 2–phase mudpit, and 3–phase mudpit. *Int J Mass Spectrom*, 219(1):245–251.

Mewes, H. W., Albermann, K., Heumann, K., Liebl, S. and Pfeiffer, F. (1997). MIPS: A database for protein sequences, homology data and yeast genome information. *Nucleic Acids Res*, 25(1):28–30.

Mewes, H. W., Frishman, D., Guldener, U., Mannhaupt, G. and Mayer, K. et al. (2002). MIPS: A database for genomes and protein sequences. *Nucleic Acids Res*, 30(1):31–4.

Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C. and Cheng, D. et al. (2003). A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol*, 21(8):921–6.

Peterson, C. L., Zhao, Y. and Chait, B. T. (1998). Subunits of the yeast SWI/SNF complex are members of the actin–related protein (ARP) family. *J Biol Chem*, 273(37):23641–4.

Pollard, K. J. and Peterson, C. L. (1997). Role for ADA/GCN5 products in antagonizing chromatin–mediated transcriptional repression. *Mol Cell Biol*, 17(11):6212–22.

Pray-Grant, M. G., Schieltz, D., McMahon, S. J., Wood, J. M. and Kennedy, E. L. et al. (2002). The novel SLIK histone acetyltransferase complex functions in the yeast retrograde response pathway. *Mol Cell Biol*, 22(24):8774–86.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M. and Mann, M. et al. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*, 17(10):1030–2.

Sadygov, R. G., Eng, J., Durr, E., Saraf, A. and McDonald, H. et al. (2002). Code developments to improve the efficiency of automated MS/MS spectra interpretation. *J Proteome Res*, 1(3):211–5.

Saitoh, S., Chabes, A., McDonald, W. H., Thelander, L. and Yates III, J. R. et al. (2002). Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. *Cell*, 109(5):563–73.

Sanders, S. L., Jennings, J., Canutescu, A., Link, A. J. and Weil, P. A. (2002). Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. *Mol Cell Biol*, 22(13):4723–38.

Scholes, D. T., Banerjee, M., Bowen, B. and Curcio, M. J. (2001). Multiple regulators of Ty1 transposition in *Saccharomyces cerevisiae* have conserved roles in genome maintenance. *Genetics*, 159(4):1449–65.

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(4):384–91.

Sharp, P. M. and Li, W. H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*, 15(3):1281–95.

Tabb, D. L., McDonald, W. H. and Yates III, J. R. (2002). DTASelect and Contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res*, 1(1):21–6.

Treger, J. M., Magee, T. R. and McEntee, K. (1998). Functional analysis of the stress response element and its role in the multistress response of *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun*, 243(1):13–9.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A. and Judson, R. S. et al. (2000). A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403(6770):623–7.

Walker, P. A., Leong, L. E., Ng, P. W., Tan, S. H. and Waller, S. et al. (1994). Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Biotechnology (N Y)*, 12(6):601–5.

Wang, Z., Castano, I. B., De Las Penas, A., Adams, C. and Christman, M. F. (2000). Pol kappa: A DNA polymerase required for sister chromatid cohesion. *Science*, 289(5480):774–9.

Washburn, M. P., Ulaszek, R. R. and Yates III., J. R. (2003). Reproducibility of quantitative proteomic analyses of complex biological mixtures by multidimensional protein identification technology. *Anal Chem*, 75(19):5054–61.

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H. and Anderson, K. et al. (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, 285(5429):901–6.

Zhu, H., Bilgin, M., Bangham, R., Hall, D. and Casamayor, A. et al. (2001). Global analysis of protein activities using proteome chips. *Science*, 293(5537):2101–5.