

The study of macromolecular complexes by quantitative proteomics

Jeffrey A. Ranish, Eugene C. Yi, Deena M. Leslie, Samuel O. Purvine¹, David R. Goodlett, Jimmy Eng & Ruedi Aebersold

Published online: 18 February 2003; doi:10.1038/ng1101

We describe a generic strategy for determining the specific composition, changes in the composition, and changes in the abundance of protein complexes. It is based on the use of isotope-coded affinity tag (ICAT) reagents¹ and mass spectrometry to compare the relative abundances of tryptic peptides derived from suitable pairs of purified or partially purified protein complexes. In a first application, the genuine protein components of a large RNA polymerase II (Pol II) preinitiation complex (PIC) were distinguished from a background of co-purifying proteins by comparing the relative abundances of peptides derived from a control sample and the specific complex that was purified from nuclear extracts by a single-step promoter DNA affinity procedure². In a second application, peptides derived from immunopurified STE12 protein complexes isolated from yeast cells in different states were used to detect quantitative changes in the abundance of the complexes, and to detect dynamic changes in the composition of the samples. The use of quantitative mass spectrometry to guide identification of specific complex components in partially purified samples, and to detect quantitative changes in the abundance and composition of protein complexes, provides the researcher with powerful new tools for the comprehensive analysis of macromolecular complexes.

One of the primary goals of proteomics is the description of the composition, dynamics and connections of the multiprotein modules that catalyze a wide range of biological functions in cells³. A number of recently described technologies have provided ways to approach these problems. The yeast 2-hybrid (Y2H) method is a genetic selection that is designed to detect binary interactions between proteins in the nucleus of a yeast cell⁴, and this method has provided the first large-scale protein linkage map for the yeast *Saccharomyces cerevisiae*⁵. One limitation of this technique is that it does not detect protein–protein interactions in the context of their physiologic environment.

Another widely used approach for the analysis of protein complexes involves the use of affinity chromatography for complex isolation or enrichment, followed by mass spectrometric identification of the constituent proteins. Two recently published large-scale efforts demonstrate the power of this approach, as the groups were able to detect co-precipitating proteins for roughly 460 and 493 bait proteins, respectively^{6,7}. Although this method is extremely powerful and allows, in principle, for the detection of many co-purifying proteins in a fraction, it remains difficult to distinguish specific from nonspecific interactions, and to detect quantitative changes in protein complex abundance and composition without direct visualization of the proteins in gels. This is mainly because mass spectrometry is not an inherently quantitative technique, and because the control and specific samples are

prepared for mass spectrometry separately and are therefore subject to different sample-handling errors.

In the method described here, stable isotope tagging and mass spectrometry were used: to guide the identification of the genuine components of a large RNA Pol II PIC (approximately 68 subunits) within a high background of co-purifying proteins following a simple one-step DNA affinity procedure; to detect quantitative changes in the abundance of STE12 protein complexes immunoprecipitated from yeast cells in different states; and to detect dynamic changes in the composition of STE12 immunoprecipitates from yeast cells in different cell states. The method is rapid, sensitive and comprehensive. In addition, since partially purified preparations can be used for the analysis, sample losses and potential losses of interacting proteins resulting from multiple chromatographic steps are avoided.

In a first application of the method, we attempted to distinguish specific components of a large, affinity-purified RNA Pol II PIC from a high background of co-purifying proteins after a one-step purification from yeast nuclear extracts. This was accomplished by comparing the sample containing the partially purified RNA Pol II PIC with a control sample in which the specific complex was not enriched. Proteins isolated from the specific and control purifications were labeled with either heavy or normal stable isotopes^{1,8,9} (Fig. 1). In this study, cysteine reactive ICAT reagents were used. Next, proteins were combined and

Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103-8904, USA. ¹Present address: Biotech, 19310 North Creek Parkway, Suite 115, Bothell, Washington 98011, USA. Correspondence should be addressed to R.A. (e-mail: raebersold@systemsbiology.org).

proteolyzed. To increase coverage by mass spectrometry, sample complexity was first reduced by strong cation exchange (SCX) chromatography and then ICAT-labeled peptides were isolated by avidin affinity chromatography¹⁰. Labeled peptides were analyzed by microcapillary reversed-phase liquid chromatography (μ LC) electrospray ionization (ESI) tandem mass spectrometry (MS/MS), and peptides were identified by sequence database searching using the search algorithm SEQUEST¹¹. The relative abundance of an identified ICAT-labeled peptide pair was determined from the ratio of the peptides' signal intensities. *Bona fide* components of the complex (or complexes) were identified by their increased abundance in the specific purification compared with the nonspecific purification.

In the yeast *S. cerevisiae*, the core Pol II preinitiation transcription machinery consists of approximately 68 subunits. The 12-subunit Pol II is responsible for the synthesis of all messenger RNAs in the nucleus. In addition to Pol II, the machinery includes six general transcription factors (GTFs; TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIH)¹². TFIID consists of the TATA binding protein (TBP) and 15 other proteins called TAFs (TBP-associated factors) that are recruited to a subset of promoters^{13,14}. In addition, a complex consisting of at least 24 factors, called Mediator, associates with Pol II and appears to be required for transcription of most promoters¹². In total, this core Pol II PIC (Pol II, GTFs, Mediator and TAFs) consists of at least 68 polypeptides. In addition to the core transcription machinery, a number of chromatin remodeling complexes (CRCs) are recruited to promoters, including SWI/SNF, RSC, ISW1, ISW2, SAGA and NuA4 (refs. 12, 15). The goals of this project were, first, to use a simple one-step

DNA-affinity isolation technique followed by quantitative mass spectrometry to comprehensively characterize a core Pol II PIC, and second, to catalog the protein composition of a eukaryotic Pol II promoter.

Pol II PICs were isolated from yeast nuclear extracts using an immobilized promoter DNA isolation scheme² (Fig. 2a). Briefly, a biotinylated template containing the TATA box region of the HIS4 promoter with a single Gal4 binding site located upstream was linked to streptavidin-coated magnetic beads and used as an affinity matrix. In previous experiments, it was established that active, promoter-dependent PICs could be isolated from nuclear extracts, and TBP and other components of the PIC were essential for full assembly of active complexes². One extract was prepared from a strain carrying a temperature-sensitive mutation in TBP (TBPI143N)¹⁶. TBP is a central component in transcription initiation because it can nucleate PIC formation by binding to the TATA element (see ref. 12 for review). When this extract was used in the immobilized promoter assay, recruitment of all core PIC factors probed for was defective, and this defect correlated with transcription inactivity². Importantly, addition of recombinant wildtype TBP (rTBP) to the reactions restored both the levels of all core PIC factors probed for and the transcription activity of the complexes to the levels observed with a wildtype extract.

Affinity-purified samples were isolated using a TBPI143N nuclear extract in the presence or absence of rTBP (Fig. 2). Consistent with previous results², the levels of known PIC components were increased by rTBP (Fig. 2b). Analysis of these fractions by SDS-PAGE followed by silver staining revealed that the isolated sample mixtures were so complex that it was difficult to detect specific components of the complex within the matrix of co-purifying proteins (Fig. 2c). Quantitative western blots and silver-stained SDS-PAGE gels (Fig. 2c and data not shown) indicated that the core Pol II components were present at 1/5–1/20 the levels of the most abundant proteins in the sample. Consistent with the background complexity, when these fractions were analyzed by mass spectrometry, 58% of the 326 unambiguously identified proteins had annotated functions in databases¹⁷ that were unrelated to Pol II transcription (see Supplementary Table 1 online). Included in the remaining 42% of proteins with annotated roles in Pol II transcription were components of the core Pol II transcription machinery as well as chromatin-remodeling factors and transcriptional regulatory factors. These included 64 of the 68 known components of the core Pol II PIC, 9 of the 9 SAGA subunits, 7 of the 7 NuA4 subunits, 7 of the

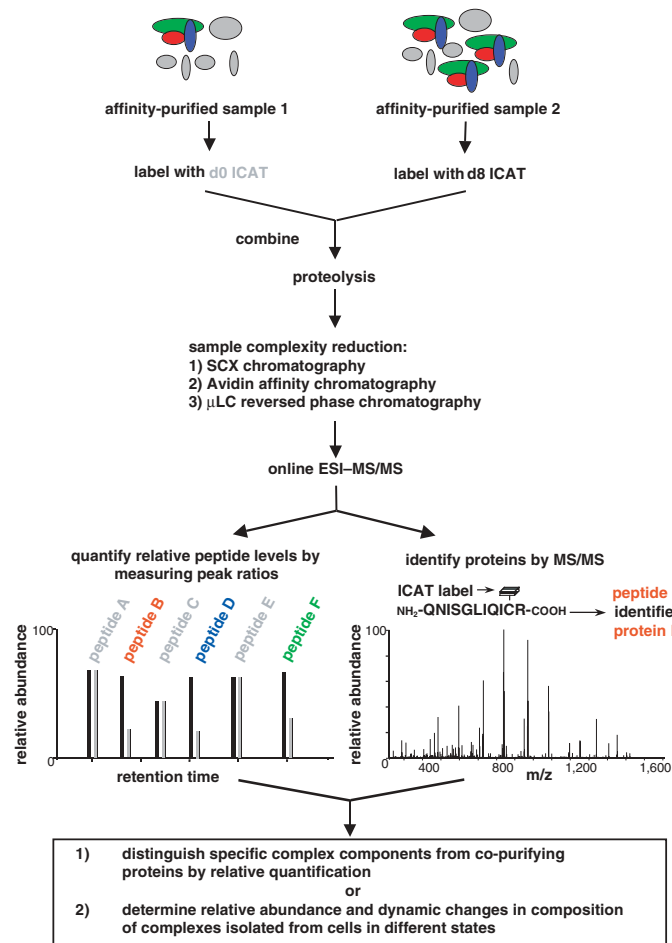
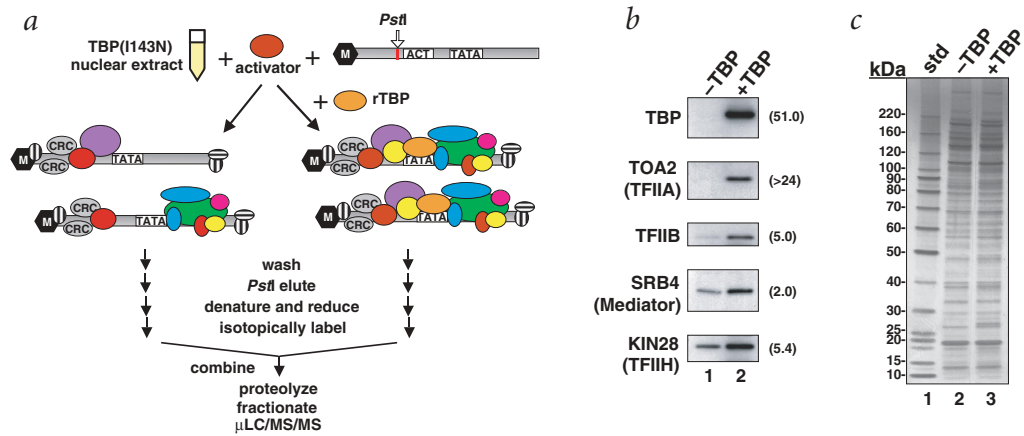


Fig. 1 Schematic representation of the quantitative proteomics approach for the analysis of affinity purified macromolecular complexes. To distinguish specific complex components from co-purifying proteins, a control purification (sample 1) is carried out in which the complex of interest is not enriched. To detect quantitative changes in the abundance and composition of a complex isolated from cells in different states, the samples are prepared identically. Affinity-purified proteins from both purifications are reduced, labeled with either the isotopically heavy or the normal version of the ICAT reagent and combined. After proteolysis, sample complexity is reduced in three sequential chromatographic steps, and then samples are subjected to ESI-MS/MS analysis. During this process, peptide pairs are quantified by measuring their peak ratios as they co-elute from the C18 column into the mass spectrometer. In every other scan, peptides are selected for fragmentation. The resulting MS/MS spectra are used to search sequence databases using SEQUEST to identify the peptides and thus the proteins from which they originated. Depending on the experiment, the relative quantification can be used to distinguish specific complex components from co-purifying proteins or to detect changes in the abundance and composition of complexes isolated from cells in different states.

Fig. 2 Single-step promoter DNA affinity purification of an RNA Pol II preinitiation complex from nuclear extracts. **a**, A nuclear extract from a strain carrying a temperature-sensitive allele of TBP (TBP1143N)¹⁶ was incubated with the activator Gal4-AH (ACT) and a fragment of the HIS4 promoter linked to magnetic beads in the presence or absence of rTBP. Extracts prepared from this strain are defective in the formation of active transcription complexes, but the addition of rTBP restores transcription activity². Templates were washed in transcription buffer and digested with *Pst*I, and promoter-bound proteins were recovered in the supernatant. Core PIC components are colored. Co-purification of chromatin remodeling complexes and nonspecific proteins (striped balls) are indicated. Proteins from each sample were labeled with either the isotopically heavy (+ rTBP) or the normal version of the ICAT reagent and then combined. Proteins were then analyzed as described in Figure 1. **b**, Western-blot analysis of proteins isolated from the single-step promoter DNA affinity purification. In this, 0.5% of the affinity purified proteins from **a** were electrophoresed on a 4–12% Bis-Tris gel, transferred to a membrane and probed with antibodies to known components of the core Pol II PIC. Proteins were detected by ECL. The factor of stimulation by TBP, as determined by quantitative western blotting, is listed to the right of the blots. **c**, Silver-stained SDS-PAGE analysis of proteins isolated from the single-step promoter DNA affinity purification. Here, 0.5% of the affinity-purified proteins from **a** were electrophoresed on a 4–12% Bis-Tris gel and detected by silver staining.



10 SWI/SNF subunits, 7 of the 11 RSC subunits, 3 of the 4 ISW1 subunits and 2 of the 2 ISW2 subunits.

Although many proteins were identified, it was not possible to distinguish specific components of the core Pol II PIC from co-purifying proteins from the protein identification results alone. However, this was possible by the application of ICAT reagent-labeling technology coupled with multidimensional chromatography (Table 1 and Supplementary Table 2 online). Figure 3 illustrates this point by showing the quantitative analysis of two peptide pairs identified in the same μLC-MS/MS run. Reconstruction of single ion chromatograms and SEQUEST searching identified TAF12 as a protein whose abundance was increased by rTBP (Fig. 3c). In contrast, the abundance of RFA2, a DNA-binding component of the replication machinery, was not affected by rTBP, and this protein was therefore identified as a contaminant. Importantly, the signal intensities for TAF12 peptides were approximately 1/5 the level of the signal intensities for RFA2 peptides (Fig. 3b,c). Thus, this quantitative approach is capable of distinguishing specific complex components from co-purifying proteins in complex samples, even if they are present at lower levels than the co-purifying proteins.

Out of 206 quantified proteins, 157 showed less than a 1.9-fold stimulation by rTBP (76%) (Table 1 and Supplementary Table 2 online). Of the 49 proteins that showed at least a 1.9-fold change in abundance, 45 (92%) were known components of the core Pol II PIC. These proteins included 17 of the 18 GTFs, 7 of the 7 nonshared subunits of Pol II, 9 of the 15 TAFs and 12 of the 24 Mediator subunits. One reason for the lack of quantification of some subunits was the absence of cysteine-containing tryptic

peptide ions with mass-to-charge (*m/z*) ratios (1+, 2+, 3+) that were within the mass range used in this study (10%, 7 of the 68 proteins). This limitation may be alleviated by the introduction of amino-terminal isotopic labeling strategies¹⁸. Another possible reason for the lack of quantification of some subunits was insufficient peak capacity of the multidimensional chromatography. In addition, since only approximately 10% of the complexes are functional in this system¹⁹, it is possible that a fraction of the complexes were partially assembled. The main source of false-negative results was Mediator components (12 of 24 quantitatively identified; 50%). Considering the GTFs and nonshared Pol II subunits, the rate of false negatives was only 4% (1 of 25), for which the one missed protein lacked cysteine.

A few core Pol II components showed less than a 1.9-fold stimulation by rTBP. Three subunits of the Pol II enzyme (RPB5, 8 and 12) were present in roughly equal ratios in the two samples. This is probably because these subunits are shared with Pol I and III, which were present at higher levels than Pol II in the samples and were unaffected by the TBP mutation. TAF3 and TAF11 showed a 1.7 and 1.4-fold stimulation, respectively. It is likely that the levels of these TAFs were

Table 1 • Summary of results

Proteins identified	326
Percentage of identified proteins with a role in Pol II transcription ^a	42
Number of components of the core Pol II PIC identified (Pol II, GTFs, SRB/Meds, and TAFs)	64
Percentage of core Pol II PIC identified	94
Number of proteins quantified	206
Number of components of the core Pol II PIC quantified	50
Percentage of quantified proteins that are components of the core Pol II PIC	24
Percentage of the core Pol II PIC identified ^b	82
Number of proteins stimulated by a factor of at least 1.9 by TBP	49
Number of components of the core Pol II PIC stimulated by a factor of at least 1.9 by TBP	45
Percentage of quantified core Pol II PIC components stimulated by a factor of at least 1.9 by TBP	96
Number of potential new core Pol II components (stimulated by a factor of at least 1.9 by TBP)	3
Number of proteins of unknown function ^c	30

^aIncludes proteins involved in maintaining or remodeling chromatin structure. ^bPercentage of quantified core Pol II components that have cysteine-containing tryptic peptide ions with *m/z* ratios (1+, 2+, 3+) that are within the mass range (400–1800 *m/z*) used in this study. ^cAccording to Yeast Protein Database and Saccharomyces Genome Database as of March 2002.

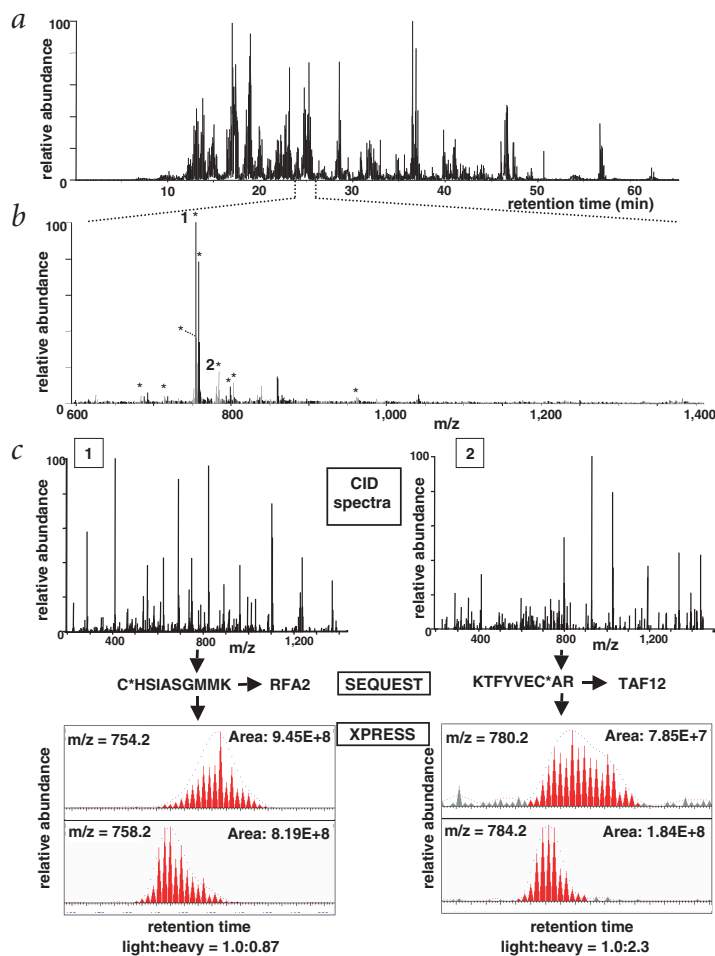


Fig. 3 Use of relative quantification to guide identification of specific complex components in complex mixtures. *a*, Analysis of auidin-purified SCX fraction 23 by μ LC-MS/MS. Ion chromatogram showing the base peak as a function of time. *b*, MS spectrum of ions eluting between 24 and 26 min during the μ LC-MS/MS run in *a*. Ions that were selected by the data-dependent routine for MS/MS analysis and identified as ICAT-labeled peptides are indicated by an asterisk. *c*, Example of post-MS processing of the data. Peaks 1 and 2 in *b* were selected for MS/MS analysis. The peptides (and the proteins from which they originated) were identified by sequence database searching using SEQUEST. Single-ion chromatograms were then reconstructed for each peptide pair using Xpress, and the relative abundance of each peptide pair was calculated after summing the signal intensities for each peptide. The relative abundance data reveal that TAF12, but not RFA2, was enriched in the reaction containing rTBP. C* in the peptide sequence indicates a cysteine residue that was labeled with the heavy form of the ICAT reagent.

cells exposed to different environmental conditions. The yeast transcription factor STE12 was protein A tagged at its native locus, and the complex of STE12 and associated proteins was immunopurified from whole-cell extracts prepared from mating type α (MAT α) cells grown in the presence or absence of α -mating factor. Immunopurified proteins were then labeled with either isotopically heavy or normal ICAT reagents and prepared for mass spectrometric analysis as described in Figure 1. Treatment of MAT α cells with α -factor causes cells to arrest in the G1 stage of mitosis and to express mating-specific genes in preparation for conjugation with cells of the opposite mating type²¹. STE12 mediates transcriptional induction of mating-specific genes in response to peptide pheromones. Using co-immunoprecipitation, *in vitro* binding or Y2H assays, STE12 has been shown to interact with a number of proteins. These include DIG1, DIG2, TEC1, FUS3, KSS1, ALPHA1, and MCM1 and PSE1 (see ref. 17 for references). It is

stimulated by approximately twofold by rTBP, like the other quantified TAF subunits, but because only one cysteine-containing peptide pair was quantified for TAF3 and 11 (1 of 1 for TAF3 and 1 of 2 for TAF11), it was difficult to assign high confidence to these measurements.

Of the four stimulated proteins that were not known components of the core Pol II PIC, three have been implicated in Pol II transcription by other studies. MOT1 is known to specifically interact with TBP on DNA²⁰. Additional experiments are being performed to test whether or not the other factors are *bona fide* components of the core Pol II PIC. Of particular interest was a protein corresponding to a previously uncharacterized open reading frame (ORF).

This is the first time that fully assembled RNA Pol II PICs have been comprehensively analyzed. The quantitative mass spectrometry technique therefore provided, also for the first time, a detailed description of the partially purified core Pol II complex and led to the detection of potential new components of this extensively studied complex. The technique is of general utility in any situation in which partially purified, macromolecular complexes are being analyzed. It obviates the need for extensive purification, which is often accompanied with general sample loss and, more importantly, with the specific loss of (weakly) interacting proteins. Therefore, the described method greatly extends the utility of mass spectrometry for the analysis of macromolecular complexes.

In the second application of the method, we attempted to detect quantitative changes in the abundance and composition of immunopurified STE12 protein complexes isolated from yeast

likely that the extent of some of these interactions change in response to α -factor treatment. In fact, co-immunoprecipitation experiments suggest a lower association of STE12 with FUS3 and DIG1 in response to α -factor treatment²².

Quantitative mass spectrometry allowed for the unambiguous identification and quantification of 13 proteins and the identification of four peptides derived from the GAG protein of Ty retroelements (Table 2). Ty retroelements are present in multiple copies in the yeast genome²³. Because of their redundancy, it is not possible to determine unambiguously the element from which each peptide was derived. Among the unambiguously identified proteins were STE12 and two known interactors, DIG1 and DIG2. Peptides derived from these proteins showed an approximate doubling in abundance in immunoprecipitates from α -factor treated cells relative to immunoprecipitates from untreated cells. By contrast, ten other proteins that were present in both samples showed no change in abundance in immunoprecipitates from treated and untreated cells. It is likely that these proteins co-purified nonspecifically because all (except for SCW4) have been previously identified as common contaminants in high-throughput proteomics studies^{6,7}. Results from western blots (see Supplementary Fig. 1 online) and DNA microarray experiments²⁴ concur with the observed increase in STE12 abundance after α -factor treatment. The coordinated increase in association of DIG1 and DIG2 with STE12 after α -factor treatment indicates an increase in a STE12 complex minimally consisting of STE12, DIG 1 and DIG2. This observation appears to be inconsistent with other results²² reporting a decrease in the amount STE12 that co-precipitated with DIG1 in

Table 2 • Quantitative characterization of STE12-pA immunoprecipitates from control and α -mating factor-treated cells

Locus	Protein	Number of peptide ions quantified	Number of unique peptides	d8/d0	STD	Percent error
YBL101W-A (+19)*	Ty2 Gag	2	1	>20		
YAR010C (+53)*	Ty1 Gag	3	1	>100		
YGR109W-A (+4)*	Ty3 Gag	1	1	13.9		
YBL005W-A (+6)*	Ty1 Gag	1	1	6.9		
YDR480W	DIG2	5	2	2.3	0.4	17
YPL049C	DIG1	1	1	2.2		
YHR084W	STE12	7	2	2.1	0.4	19
YBR118W	TEF1 or 2	1	1	1.3		
YLR180W	SAM1	2	1	1.2	0.0	0
YBL087C	RPL23A	2	1	1.1	0.2	18
YLR441C	RPS1A or B	1	1	1.1		
YDR502C	SAM1 or 2	1	1	1.0		
YGL135W	RPL1A or B	2	1	1.0	0.1	10
YGR279C	SCW4	1	1	1.0		
	Human IgGs	35	22	0.9	0.1	11
YNL064C	YDJ1	1	1	0.9		
YBR127C	VMA2	1	1	0.8		

*Numbers in parentheses indicate the number of open reading frames encoded in the *S. cerevisiae* genome containing the same peptide sequence.

response to a 30 minute α -factor treatment. This discrepancy is probably explained by the significantly different duration of α -factor exposure (30 minutes versus 3 hours) and in the different immunoprecipitation targets that were used in the respective studies. In the model proposed by Tedford *et al.*²², DIG1 and DIG2 bind to STE12 and repress the ability of STE12 to activate transcription, in the absence of α -factor. After exposure to α -factor, the mitogen-activated protein (MAP) kinase FUS3 phosphorylates DIG 1, DIG2 and STE12, which results in the release of STE12 from repression by DIG1 and DIG2, allowing STE12 to activate the transcription of genes involved in the mating response. Microarray data show, however, a decrease in the expression of a number of STE12-induced genes after exposure to α -factor for 2 hours²⁴. It is probable that, after prolonged exposure to α -factor treatment, the STE12–DIG1–DIG2 complex is re-established to downregulate STE12-induced gene expression. This model is consistent with our data as well as the previous data²². The results demonstrate the ability of the technique to detect changes in the abundance of an immunopurified protein complex isolated from cells exposed to different environmental conditions, thus greatly extending the usefulness of current mass spectrometry techniques.

In addition to detecting changes in the abundance of protein complexes, the method can detect dynamic changes in the composition of affinity-purified protein samples. Peptides derived from Ty 1, 2 and 3 retroelement GAG proteins show an approximately 7 to over 100-fold increase in abundance in STE12 immunoprecipitates from α -factor-treated cells compared with untreated cells (Table 2). Since the STE12 core complex shows only an approximately twofold increase in abundance after α -factor treatment, the increase in Ty GAG proteins indicates an increased association of these proteins with the STE12 affinity matrix after exposure to α -factor. Because, however, the transcription of Ty 1, 2 and 3 retroelements is strongly induced by exposure of cells to α -factor²⁴, we cannot rule out potential non-specific interactions of the increased amounts of Ty GAG proteins without further experiments. The results nonetheless demonstrate the ability of the technique to detect dynamic changes in the composition of immunopurified protein samples by direct mass spectrometric analysis, a type of information that has been difficult to obtain. By performing an appropriate

control experiment, the researcher can readily test the specificity of the interaction.

Having described three applications of the use of quantitative proteomics for the study of macromolecular complexes, it is important to comment on the accuracy of the method. Three different lines of evidence have been used to show the quantitative accuracy of the method. First, several previously published papers have shown that the method provides accurate relative quantification of proteins at known concentrations (with an accuracy in the region of 15%) in mixtures^{1,25,26}. Furthermore, a dose–response curve has been presented that demonstrated the linearity of the approach over a tenfold

range of relative abundance²⁶. Second, in cases in which, during the analysis of complex peptide mixtures, several observations of peptides from the same protein are made, the error inherent in the method can be assessed. An error rate of less than 20% was consistently found¹⁰. Analysis of the data from the Pol II complex experiment indicates an error rate of 25% or less for more than 90% of the quantitative measurements for which more than one peptide pair was quantified (see Supplementary Table 2 online and data not shown). Third, comparisons of abundance ratios measured in complex mixtures by isotope dilution and by western blotting consistently show comparable values²⁷ (Fig. 2b). Data also showed the coordinate increase (2–3-fold) in 20 ribosomal subunits in response to Myc expression, again reflecting the ability of the technique to accurately detect quantitative changes in the abundance of protein complexes²⁷. In every case, the ICAT reagent method was found to measure changes in abundance accurately.

There are two primary factors that potentially compromise the accuracy of quantification. First, variations in signal-to-noise ratio will affect the accuracy of the measurement, especially for low-intensity ions. Since different peptides ionize with different efficiencies, it is not possible to obtain consistent signal-to-noise data for all the peptides from a particular protein or from a protein complex. Second, although ICAT-labeled peptide pairs closely chromatograph, there is a slight delay in the elution of peptides labeled with the normal isotope. Fluctuations in electrospray ionization can produce variations in ion intensity during elution. This can lead to inaccuracies in quantification. These issues are being addressed by a number of technical advances. The use of a matrix-assisted laser desorption (MALDI) quadrupole time-of-flight mass spectrometer increases signal-to-noise ratios because the high mass accuracy and resolution afforded by these instruments reduces the noise level in the mass window of the target analyte, and the MALDI ionization method circumvents problems with unstable electrospray²⁵. In addition, a new automated, quantification algorithm has been developed which increases the accuracy of quantification by improved modeling of eluting peptide peaks and by taking into account background noise (X. Li, manuscript in preparation). Furthermore, second-generation ICAT reagents, which show precise co-elution of the heavy and normal form of a peptide during reversed-phase

high-performance liquid chromatography (HPLC), have been introduced (ABI). It can therefore be anticipated that the accuracy of quantification achieved with quantitative mass spectrometry will further increase in the near future.

Comprehensive analysis of macromolecular complexes is a formidable task due to the complexity of the samples, the propensity to co-purify unrelated (unspecific) proteins in even the most rigorous complex isolation schemes, the difficulty of purifying sufficient quantities for protein identification, the loss of interacting proteins during purification and the need to detect quantitative changes in complex abundance and composition. The quantitative proteomics strategies described here provide the researcher with powerful new ways to approach these problems. The technology provides a way of reliably distinguishing specific complex components from co-purifying proteins, even when the specific components are more than 20 times less abundant than the copurifying proteins. By applying the technology to the analysis of complexes isolated by simple one-step affinity purifications, protein losses resulting from multiple purification steps are avoided, the potential to identify weakly associated factors is increased, and time is saved. In addition, since the method provides better-controlled data than traditional mass spectrometry methods, it is expected to benefit high-throughput studies in which validation of the data is generally a huge problem, although this has yet to be shown. Finally, the ability to detect quantitative changes in the abundance and composition of protein complexes greatly extends the usefulness of data acquired with current mass spectrometric techniques. For these reasons, this approach constitutes a powerful new tool for the characterization of a wide range of macromolecular complexes.

Methods

Yeast strains and extracts. We used 18 l of yeast strain SHY70 [*MATa ade6 leu2-3,112 ura3-52 his4-519 Δspt15:HIS4/pSH254* (ARS CEN *LEU2 spt15* (TBPI143N))] grown at 26 °C in YPAD (1% yeast extract, 2% peptone, 40 mg l⁻¹ adenine and 2–3% glucose) to an A_{600} of 3 to prepare nuclear extracts for the Pol II PIC purification as described²⁸. We used yeast strain DF5a [*MATa leu2-3,112 his3-Δ200 trp1-1 lys2-801*] to create a C-terminal protein A-tagged version of genomic *STE12* by PCR-mediated homologous recombination as previously described²⁹. Primer sequences are available upon request. We grew two 1-l cultures in YPD to an A_{600} of 0.2 at 30 °C. One of the cultures was induced by the addition of α -mating factor (WHWLQLKPGQPMY-OH, Biopeptide) to 3 μ M, and both cultures were incubated at 30 °C for 3 h. Approximately 80% of cells had formed shmoo at the time of harvest. We collected the cells, washed them in water and resuspended them in 20 ml lysis buffer (LB; 20 mM NaHPO₄ (pH 7.5), 150 mM NaCl, 0.1 mM MgCl₂ and 50 mM β -glycerophosphate) containing protease inhibitors: 1 mM PMSF, 2 mM benzamide, 0.6 μ M leupeptin, 2 μ M pepstatin A and 3 μ M chymostatin. After lysing the cells using a microfluidizer (Microfluidics), we added dimethylsulfoxide and Triton X-100 to 10% and 1% (v/v), respectively, and incubated the extracts at 4 °C for 30 min. Insoluble material was removed by centrifuging extracts at 30,000g for 30 min. We added glycerol to 10% (v/v) and stored the extracts at –80 °C. We determined protein concentrations using the BioRad dye-binding assay.

Immobilized promoter templates. We prepared immobilized promoter templates as previously described². For these experiments, a 400-bp fragment containing 71 bp of the yeast *HIS4* promoter centered around the TATA box, one Gal4 binding site and upstream vector sequences was PCR amplified from pSH515 with biotinylated upstream primer p965 and downstream primer BKS8. Primer sequences are available upon request.

Isolation of RNA Pol II PICs. We prepared Pol II PICs as previously described² with the following modifications. Reactions were scaled up 215 \times compared with the reactions used to analyze proteins by western blotting². Transcription buffer consisted of 20 mM HEPES buffer (pH 7.6), 100 mM potassium acetate, 5 mM magnesium acetate and 1 mM EDTA. After incubating the TBP(I143N) nuclear extract with promoter templates

for 60 min, we washed the templates four times with transcription buffer containing 0.05% Nonidet P-40 and 2.5 mM dithiothreitol and once with transcription buffer containing 0.003% Nonidet P-40. We resuspended templates in 1 ml *Pst*I buffer (100 mM NaCl, 50 mM Tris-HCl (pH 7.9) and 10 mM MgCl₂) with 645 Units *Pst*I (Boehringer Mannheim) and incubated them for 30 min at 22 °C with agitation. We concentrated the beads with a magnet and recovered the supernatants. For western-blot analysis, proteins were electrophoresed on a 4–12% Bis-Tris gel (Invitrogen), transferred to Immobilon membranes (Millipore) and probed with antibodies to known components of the core Pol II PIC. Proteins were detected by ECL (Amersham). We quantified proteins by densitometry with IQMACv1.2 software (Molecular Dynamics). The relative levels of subunits were determined from standard curves that were generated with increasing amounts of nuclear extract.

ICAT labeling of DNA affinity-purified proteins and preparation of peptides for μ LC-MS/MS. We concentrated protein samples in Microcon 10 devices (Amicon) and exchanged buffer by diluting the samples tenfold with 20 mM Tris-HCl (pH 8.3), 50 mM NaCl and 1 mM EDTA. After concentrating the samples to 50 μ l, we added SDS to 0.3% and heated the samples at 100 °C for 5 min. Proteins (approximately 130 μ g per sample) were reduced with 5 mM TBP at 37 °C for 30 min and then diluted with 250 μ l 20 mM Tris-HCl (pH 8.3), 1 mM EDTA and 7.2 M urea. We added isotopically heavy (+TBP) or normal (–TBP) ICAT reagent to 1.75 mM and incubated samples for 90 min at 22 °C. We quenched reactions by adding 10 mM dithiothreitol for 20 min at 37 °C, combined samples and digested proteins by adding 4 μ g endoproteinase Lys-C (Boehringer Mannheim) at 37 °C for 3 h. SDS and urea concentrations were reduced to 0.01% and 1.2 M, respectively, by adding 20 mM Tris-HCl (pH 8.3) and 1 mM EDTA, and samples were digested with trypsin (Promega, sequencing grade modified, 1:25 w/w) overnight at 37 °C. We diluted the sample with an equal volume of Buffer A (5 mM KH₂PO₄ (pH 3), 25% CH₃CN) and adjusted the pH to 3 with 10% trifluoroacetic acid. Peptides were fractionated by SCX HPLC (2.1 \times 200 mm PolySULFOETHYL A; PolyLC) by running the following gradient: 0–15% Buffer B (5 mM KH₂PO₄ (pH 3), 600 mM KCl, 25% CH₃CN) over 30 min, 15–60% Buffer B in 20 min and 60–100% Buffer B in 15 min at 0.2 ml min⁻¹. We collected 30 fractions of 0.4 ml. Labeled peptides were purified over monomeric avidin cartridges (ABI) and washed with 2 \times phosphate-buffered saline (PBS; pH 7.2), 1 \times PBS (pH 7.2) and 50 mM NH₄HCO₃ (pH 8.3), 20% CH₃OH. Peptides were eluted with 0.4% trifluoroacetic acid in 30% CH₃CN, dried under reduced pressure and resuspended in 10% CH₃CN, 0.1% trifluoroacetic acid. Peptides from avidin flowthrough fractions were captured on C18 cartridges (Vydac), washed with 0.5% CH₃CN, 0.1% trifluoroacetic acid and eluted with 60% CH₃CN, 0.1% trifluoroacetic acid. Samples were dried and resuspended in 0.5% CH₃CN, 0.1% trifluoroacetic acid.

Purification of protein A-tagged STE12. We prepared 240 μ l IgG Sepharose beads (Pharmacia) by washing them four times with 2.5 ml LB containing 10% dimethylsulfoxide and 1% Triton X-100 (LBDT) with 500 mM NaCl and then once with 2.5 ml LBDT containing 150 mM NaCl. Next, we incubated beads in 0.5 M acetic acid (pH 3.2) for 5 min and then washed them three times with 2.5 ml LBDT containing 150 mM NaCl. Beads were equilibrated in LBDT containing 150 mM NaCl for 30 min before use. We incubated 120 mg of extract from control and α -factor-treated cells with 120 μ l IgG Sepharose beads at 4 °C for 15 h. We transferred beads to columns and washed them extensively with wash buffer (WB; 20 mM Tris (pH 7.6), 150 mM NaCl, 1 mM EDTA and 0.1 mM MgCl₂) containing 0.1% Nonidet P-40 and protease inhibitors and then with WB containing 0.005% Nonidet P-40. Proteins were first eluted with 1 ml WB containing 500 mM NaCl and 0.005% Nonidet P-40. After washing the beads with water, a second elution was carried out by incubating the beads in 250 μ l 0.5 M acetic acid (pH 3.2) for 10 min. We repeated this elution twice and pooled the eluates.

ICAT labeling of IgG purified proteins and preparation of peptides for μ LC-MS/MS. We concentrated the 0.5 M fractions in Microcon 10 devices (Amicon) and exchanged buffer by diluting the samples tenfold with 20 mM Tris-HCl (pH 8.3), 30 mM NaCl, 1 mM EDTA. After concentrating the samples to 25 μ l, we added SDS to 0.25% and heated the samples at 100 °C for 5 min. The acetic acid-eluted fractions were dried down, resuspended in 25 μ l 20 mM Tris-HCl (pH 8.3), 1 mM EDTA and 0.25% SDS and heated at 100 °C

for 5 min. Proteins (approximately 20–40 µg per sample) were reduced with 5 mM TBP at 37 °C for 30 min and then diluted with 100 µl 20 mM Tris-HCl (pH 8.3), 1 mM EDTA, 7.5 M urea. We added isotopically heavy (α -factor) or normal (untreated) ICAT reagent to 1.5 mM and incubated samples for 90 min at 22 °C. We quenched reactions by adding 10 mM dithiothreitol for 20 min at 37 °C, combined samples and digested proteins by adding 1 µg endoproteinase Lys-C (Boehringer Mannheim) at 37 °C for 3 h. SDS and urea concentrations were reduced to 0.01% and 1.2 M, respectively, by adding 20 mM Tris-HCl (pH 8.3), 1 mM EDTA, and samples were digested with trypsin (Promega, sequencing grade modified, 1:20 w/w) overnight at 37 °C. The samples were diluted with an equal volume of Buffer A, and the pH was adjusted to 3 with 10% trifluoroacetic acid. Peptides were loaded onto SCX cartridges (ABI) and successively eluted with 0.75 ml Buffer A containing 40 mM, 200 mM, 350 mM and 600 mM KCl. Labeled peptides were purified over monomeric avidin cartridges (ABI) and prepared for MS analysis as described above.

μ LC-MS/MS and data analysis. We pressure-loaded 50–100% of avidin-purified fractions onto 10 cm \times 75 µm fused silica microcapillary reversed phase columns (5 µm Magic C18 beads, Michrom Bioresources) equilibrated with 10% CH₃OH, 0.4% acetic acid, 0.005% heptafluorobutyric acid (HFBA). Peptides were resolved by running 80-min gradients from 10–40% Buffer C (100% CH₃OH, 0.4% acetic acid, 0.005% HFBA) at 0.3 µl min⁻¹ and analyzed by μ LC-MS/MS using an LCQ ion trap mass spectrometer (ThermoFinnigan) as described¹. We also analyzed 33% of each of the avidin flowthrough fractions from the DNA affinity experiment by automated μ LC-MS/MS³⁰. We identified peptides by searching MS/MS spectra against a yeast protein database using SEQUEST as described¹⁰. We quantified and analyzed data essentially as described¹⁰ using XPRESS and INTERACT computer programs, respectively. We normalized the relative ratios to adjust for differences in labeling efficiency by identifying a representative mean ratio. This was done by calculating the mean ratio from an approximately normal distribution identified from the histogram of all ratios. For the DNA affinity experiment, this distribution included 160 ratios ranging from 1.0:0.5 to 1.0:1.2. Fifty-five ratios that were greater than 1.0:1.2 lay outside this distribution. All peptide identifications and quantifications were confirmed by manual inspection of the data.

URL. Saccharomyces Genome Database is available at <http://genome-www.stanford.edu/Saccharomyces/>.

Note: Supplementary information is available on the Nature Genetics website.

Acknowledgments

We thank S. Hahn for the gift of rTBP and antibodies to TBP, TFIIB and SRB4, A. Nesvizhskii for help with data analysis and J. Aitchison, M. Wright and B. Wollscheid for comments on the manuscript. This work was supported by grants from the US National Cancer Institute and US National Institutes of Health Research Resource Center, by federal funds from the National Heart, Lung, and Blood Institute of the National Institutes of Health and by a postdoctoral fellowship from the National Institutes of Health to J.A.R. Partial funding for this work came through a gift from Merck and Co. to the Institute for Systems Biology.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 7 October 2002; accepted 19 January 2003.

- Gygi, S.P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
- Ranish, J.A., Yudkovsky, N. & Hahn, S. Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev.* **13**, 49–63 (1999).
- Hartwell, L.H., Hopfield, J.J., Leibler, S. & Murray, A.W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
- Fields, S. & Song, O.-k. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246 (1989).
- Uetz, P. et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Gavin, A.C. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Ho, Y. et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
- Zhou, H., Ranish, J.A., Watts, J.D. & Aebersold, R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat. Biotechnol.* **20**, 512–515 (2002).
- Cagney, G. & Emili, A. *De novo* peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat. Biotechnol.* **20**, 163–170 (2002).
- Han, D.K., Eng, J., Zhou, H. & Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951 (2001).
- Eng, J.K., McCormack, A.L. & Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in protein databases. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
- Lee, T.I. & Young, R.A. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137 (2000).
- Li, X.Y., Bhaumik, S.R. & Green, M.R. Distinct classes of yeast promoters revealed by differential TAF recruitment. *Science* **288**, 1242–1244 (2000).
- Kuras, L., Kosa, P., Mencia, M. & Struhl, K. TAF-containing and TAF-independent forms of transcriptionally active TBP *in vivo*. *Science* **288**, 1244–1248 (2000).
- Fry, C.J. & Peterson, C.L. Chromatin remodeling enzymes: who's on first? *Curr. Biol.* **11**, R185–R197 (2001).
- Reddy, P. & Hahn, S. Dominant negative mutations in yeast TFIID define a bipartite DNA-binding region. *Cell* **65**, 349–357 (1991).
- Costanzo, M.C. et al. YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res.* **29**, 75–79 (2001).
- Zhang, X., Jin, Q.K., Carr, S.A. & Annan, R.S. N-terminal peptide labeling strategy for incorporation of isotopic tags: a method for the determination of site-specific absolute phosphorylation stoichiometry. *Rapid Commun. Mass Spectrom.* **16**, 2325–2332 (2002).
- Yudkovsky, N., Ranish, J.A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225–229 (2000).
- Auble, D.T. et al. Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. *Genes Dev.* **8**, 1920–1934 (1994).
- Sprague, G.F. & Thorner, J.W. Pheromone response and signal transduction during the mating process of *Saccharomyces cerevisiae*. in *The Molecular and Cellular Biology of the Yeast Saccharomyces* Vol. 2 (eds. Jones, E.W., Pringle, J.R. & Broach, J.R.) 657–744 (Cold Spring Harbor Laboratory Press, Plainview, 1992).
- Tedford, K., Kim, S., Sa, D., Stevens, K. & Tyers, M. Regulation of the mating pheromone and invasive growth responses in yeast by two MAP kinase substrates. *Curr. Biol.* **7**, 228–238 (1997).
- Olson, M.V. Genome structure and organization in *Saccharomyces cerevisiae*. in *The Molecular and Cellular Biology of the Yeast Saccharomyces* Vol. 1 (eds. Broach, J.R., Pringle, J. & Jones, E.) 1–39 (Cold Spring Harbor Laboratory Press, Plainview, 1991).
- Roberts, C.J. et al. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287**, 873–880 (2000).
- Griffin, T.J. et al. Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer. *Anal. Chem.* **73**, 978–986 (2001).
- Griffin, T.J. et al. Toward a high-throughput approach to quantitative proteomic analysis: expression-dependent protein identification by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **12**, 1238–1246 (2001).
- Shio, Y. et al. Quantitative proteomic analysis of Myc oncoprotein function. *EMBO J.* **21**, 5088–5096 (2002).
- Kang, J.J., Auble, D.T., Ranish, J.A. & Hahn, S. Analysis of the yeast transcription factor TFIIA: distinct functional regions and a polymerase II-specific role in basal and activated transcription. *Mol. Cell. Biol.* **15**, 1234–1243 (1995).
- Aitchison, J.D., Rout, M.P., Marelli, M., Blobel, G. & Wozniak, R.W. Two novel related yeast nucleoporins Nup170p and Nup157p: complementation with the vertebrate homologue Nup155p and functional interactions with the yeast nuclear pore-membrane protein Pom152p. *J. Cell Biol.* **131**, 1133–1148 (1995).
- Yi, E.C. et al. Approaching complete peroxisome characterization by gas-phase fractionation. *Electrophoresis* **23**, 3205–3216 (2002).