Direct analysis of protein complexes using mass spectrometry

Andrew J. Link[1,2], Jimmy Eng[1], David M. Schieltz[1], Edwin Carmack[1], Gregory J. Mize[3], David R. Morris[3], Barbara M. Garvik[4], and John R. Yates, III[1]*

*[1]Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195. [2]Current address: Millennium Predictive Medicine, Inc., One Kendall Square, Building 700, Cambridge, MA 02139. [3]Department of Biochemistry, University of Washington, Seattle, WA 98195. [4]Fred Hutchinson Cancer Research Institute, Seattle, WA 98109. *Corresponding author (e-mail: jyates@u.washington.edu).*

Received 21 September 1998; accepted 23 April 1999

We describe a rapid, sensitive process for comprehensively identifying proteins in macromolecular complexes that uses multidimensional liquid chromatography (LC) and tandem mass spectrometry (MS/MS) to separate and fragment peptides. The SEQUEST algorithm, relying upon translated genomic sequences, infers amino acid sequences from the fragment ions. The method was applied to the *Saccharomyces cerevisiae* ribosome leading to the identification of a novel protein component of the yeast and human 40S subunit. By offering the ability to identify >100 proteins in a single run, this process enables components in even the largest macromolecular complexes to be analyzed comprehensively.

Keywords: protein identification, mass spectrometry, multidimensional chromatography, ribosome, yeast genome

Most cellular processes are performed and regulated by proteins acting in macromolecular complexes. Many of these complexes are composed of large numbers of unique proteins. For example, RNA polymerase II transcription complexes in eukaryotic cells probably contain at least 50 different proteins[1], and the eukaryotic ribosome consists of at least 78–80 unique proteins[2,3]. Conventional genetic and biochemical approaches for identifying protein–protein interactions generally focus on one target gene or protein at a time[4]. Even when macromolecular complexes are isolated intact from the cell, the protein components are separated and identified individually[5]. Methods to identify complex mixtures of proteins without the need to purify each component to homogeneity not only would improve the efficiency of protein identification, but should also increase the sensitivity of detection.

We report the development of a process for the direct analysis of large protein complexes (DALPC) that is capable of comprehensively identifying individual proteins in even the most complex macromolecular complex in the cell without first purifying each protein component to homogeneity. The DALPC process is shown schematically in Figure 1. It couples multidimensional chromatography and tandem mass spectrometry with automated comparison of tandem mass spectra with translated genomic sequences.

Although proteins in simple mixtures can be identified by combining liquid chromatography and tandem mass spectrometry, complex mixtures overwhelm the resolution capability of any single-dimensional (1D) chromatography system[6–8]. Orthogonal two-dimensional (2D) separation methods dramatically improve the resolution of highly complex mixtures of proteins and peptides compared with any 1D separation[9–13]. The DALPC process uses the independent physical properties of charge and hydrophobicity to resolve complex peptide mixtures before mass spectrometry. A denatured and reduced protein complex is first digested to generate a mixture of peptide fragments (Fig. 1). The acidified complex peptide mixture is applied to a strong cation exchange (SCX) chromatography column, and a discrete fraction of the absorbed peptides are displaced onto a reversed-phase (RP) chromatography column using a salt step gradient. Peptides are retained on the RP column, but contaminating salts and buffers are washed away and diverted to waste. The peptides are then eluted from the RP column into the mass spectrometer using a gradient of increasing acetonitrile. Finally, the RP column is reequilibrated in preparation for absorbing another fraction of peptides from the SCX column. An iterative process of increasing salt concentration is then used to displace additional fractions of peptides from the SCX column onto the RP column. Each simplified fraction is eluted from the RP column into the mass spectrometer.

Using the SEQUEST algorithm, acquired fragmentation spectra of peptides are correlated with predicted amino acid sequences in translated genomic databases[6,14]. The DALPC process relies on the predictive powers of whole genome sequences to identify actual proteins in complexes. This approach avoids the manual interpretation of amino acid sequences from each tandem mass spectrum and provides a rapid method to process the spectrometry data. The resulting list of peptide sequences identifies the proteins in the starting complex. To demonstrate these concepts, the *Saccharomyces cerevisiae* ribosome was analyzed using DALPC.

The eukaryotic 80S ribosome, comprising a small 40S subunit and a large 60S subunit, is one of the largest macromolecular complexes in the cell. For *S. cerevisiae*, the number and identities of proteins in each subunit have been determined primarily by isolating the ribosomal particles, electrophoretically separating the proteins using 2D gels, and sequencing the proteins by Edman degradation. Several groups have estimated that the yeast cytoplasmic 80S ribosome complex contains 78 different proteins present at unimolar amounts (32 proteins in the 40S and 46 in the 60S subunits)[3,15,16]. In yeast, 137 genes encode the 78 proteins with 59 of the proteins encoded by duplicate genes[3]. For duplicated genes that have been transcriptionally analyzed, both genes are actively expressed but usually at different levels[17]. Although a small percentage of yeast genes contain introns (~3.7% of the annotated open reading frames, ORFs), 66% of the ribosomal genes contain a single intron in the 5′ region of the gene.

Results

Validating DALPC using the 80S ribosome. We isolated the yeast 80S ribosomes and proteolytically digested 120 μg of the purified

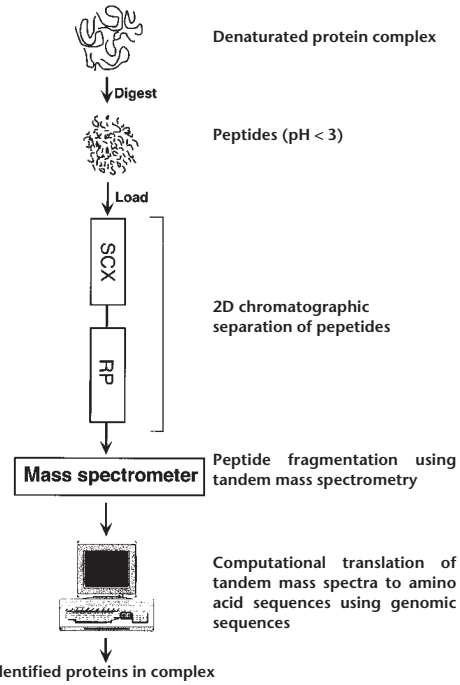


Figure 1. Direct analysis of large protein complexes (DALPC). In the flow diagram, the rectangles represent a strong cation exchange (SCX) and a reversed-phase (RP) liquid chromatography column. Typically, a denatured protein complex is digested with trypsin. The acidified peptide mixture is loaded onto the SCX column. A discrete fraction of peptides is displaced from the SCX column to the RP column. This fraction is eluted from the RP column into the mass spectrometer. This iterative process is repeated, obtaining the fragmentation patterns of peptides in the original peptide mixture. The program SEQUEST is used to correlate the tandem mass spectra of fragmented peptides to amino acid sequences using nucleotide databases[6]. The filtered outputs from the program are used to identify the proteins in the original protein complex.

proteins (~1.5 μg per protein) under highly denaturing conditions to produce a mixture of tryptic peptides. As a benchmark, the digested proteins were first applied to a 1D RP column coupled to an electrospray ionization–tandem mass spectrometer to identify the proteins in the original mixture (Fig. 2A and B). The experiment identified only 56 ribosomal proteins, 22 fewer than the predicted number. Many of the protein identifications were based on single-peptide fragmentation spectra and required manual verification. Examination of the mass spectra of ionized peptides before parent ion selection and fragmentation showed that multiple peptide ions were coeluting and escaping identification.

Automated 2D liquid chromatography was employed to improve the separation of peptides. Starting with the same amount of the peptide mixture used in the single-column experiment, 12 iterative cycles of DALPC were performed (i.e., 12 independent fractions from the SCX column were eluted to the RP column and analyzed). Compared with the 1D chromatography experiment, the number of unique peptides that correlated with the translated yeast genome dramatically increased (Fig. 2A). Of the unique peptides identified, 84% were detected in only one of the 12 cycles, suggesting most of the peptides were displaced from the SCX column in discrete salt displacement cycles. In ~24 h, the single DALPC experiment identified 95 unique proteins in the mixture (Fig. 2C). Most of these identifications were based on two or more unique peptides per protein, which greatly increased the certainty of the identifications.

Further analysis of the proteins identified from the ribosome complex showed that 90 of the 95 polypeptides were from distinct ribosomal genes (Fig. 3C). These represent 75 of the 78 predicted ribosomal proteins. Even though some of the duplicate genes expressed proteins with >98% amino acid sequence identity, uniquely identified peptides from both proteins verified that the duplicate genes were each expressed and that both proteins cosedimented in the ribosome complexes. Overall, 58% of the 571 predicted tryptic peptides from the ribosome complex with masses in the scan range of the mass spectrometry were successfully identified during the DALPC run. Three predicted ribosomal proteins, RPL41A/B, RPL40A/B, and RPL29, which are assumed present in the purified complex, were not identified. RPL41A/B is a 3.3 kDa protein that produces tryptic peptides less than two amino acids long. The other two proteins, RPL40A/B (6.0 kDa) and RPL29 (6.7 kDa), generate short tryptic peptides with an average length of three to four amino acids, only four of which are in the mass spectrometer's range. A mass spectrometric analysis of the trypsin-digested 80S complex designed only to perform tandem mass spectrometry on the parent masses of the four missing tryptic peptides failed to identify the peptides. We suspect the missing proteins did not cosediment with the 80S complex in this preparation. DALPC did successfully identify the small ribosomal proteins RPS29A/B (6.2 kDa) and RPL39 (5.7 kDa).

To confirm that the DALPC process successfully identified the major proteins in the 80S ribosome preparation, we separated 120 μg of the purified ribosomal proteins on a 2D gel and observed 55–60 intense silver-stained spots (Fig. 2D). Individual spots were excised from the gel and proteolytically digested. Recovered peptides were analyzed using microcapillary HPLC–electrospray ionization–tandem mass spectrometry[18]. From the 56 spots analyzed, we identified 64 proteins on the 2D gel. All 64 proteins were also identified using the single DALPC experiment (Fig. 3C). RPL29, RPL40A/B, RPL41A/B were not identified on the 2D gel. We suspect the limited fractionation range of the 2D gel and unequal solubilization of ribosomal proteins limited the number of proteins that could be identified. Various 2D electrophoretic methods used to define the protein composition of the ribosome often have given ambiguous and incomplete results[15]. The DALPC system provided a more comprehensive analysis of the complex in a fraction of the time it would take to purify each component to homogeneity and subsequently identify each protein.

Next, we tested the ability of DALPC to identify the ribosomal proteins in the background of a total yeast cell extract. At the same time, we wanted to investigate the potential for analyzing proteins in mixtures more complex than the purified 80S ribosome. Starting with 375 μg of trypsin-digested total cellular protein purified from yeast growing in rich media (1.4×10^8 cell equivalents), 21 iterative cycles of DALPC were performed with the chromatography and mass spectrometry conditions employed in the analysis of the 80S complex. The increase in iterative cycles was used to elute peptides from the SCX column in smaller, less complex increments. We identified 749 unique peptides in the mixture leading to the identification of 189 unique proteins (Fig. 3). Without any prior enrichment, the process was able to detect 71 of 78 predicted ribosomal proteins. This total is only four fewer than that achieved when starting with the purified 80S ribosome. The seven ribosomal proteins that were not identified were small ribosomal proteins that generate only two to three tryptic peptides in the scan range of the mass spectrometer. Given the number of cell equivalents applied and mass spectrometer detection limit of 1,000 fmol, we estimate that this DALPC experiment was starting to identify proteins present at levels of 4,300 molecules/cell. However, the signal from the more abundant proteins still obscures many proteins present at or below this cellular abundance. Increasing the peak capacity of the 2D liquid chromatography by increasing the fractionation steps is one way to identify less abundant proteins. Nonetheless, the results demonstrate that DALPC has

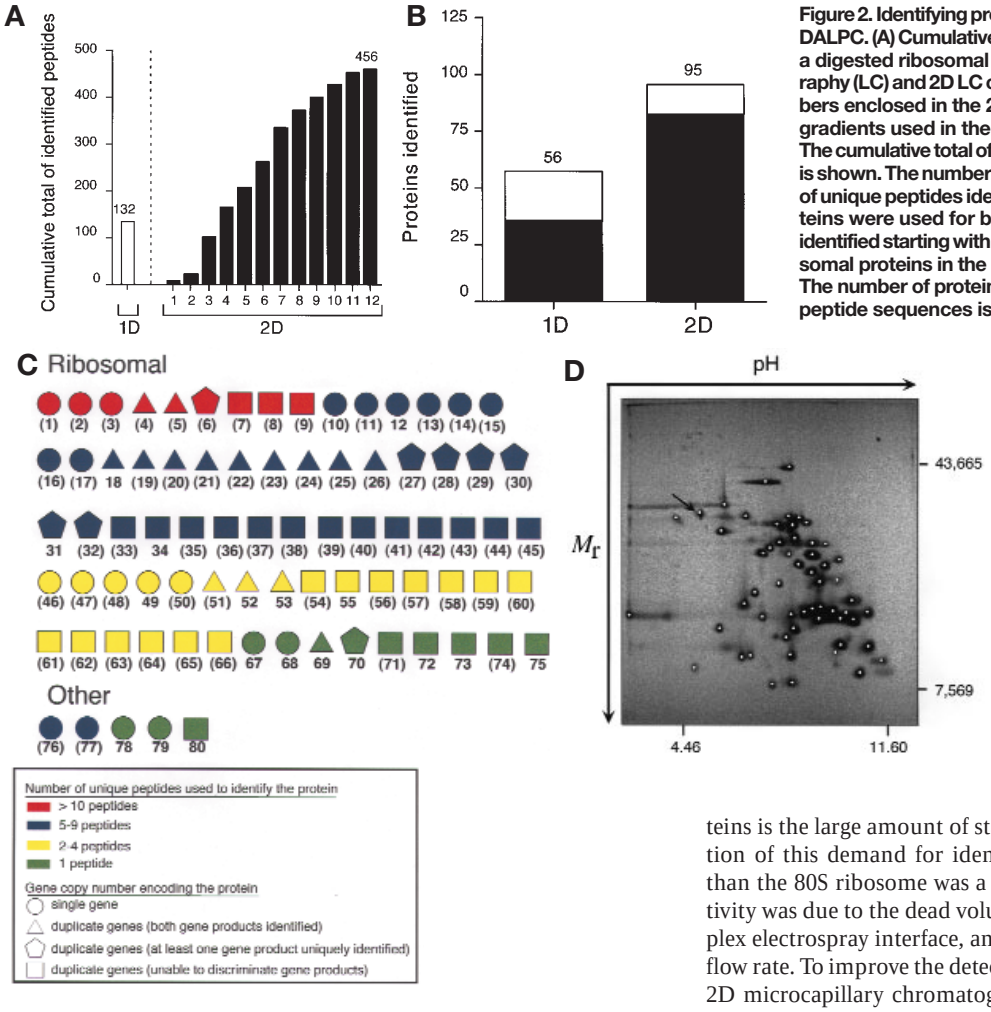


Figure 2. Identifying proteins in the yeast ribosome complex using DALPC. (A) Cumulative number of unique peptides identified from a digested ribosomal 80S complex using 1D liquid chromatography (LC) and 2D LC coupled to a mass spectrometer. The numbers enclosed in the 2D bracket are the salt displacement step gradients used in the experiment (see Experimental protocol). The cumulative total of peptides identified after each DALPC cycle is shown. The numbers above the columns are the total numbers of unique peptides identified. Identical amounts of digested proteins were used for both experiments. (B) Number of proteins identified starting with identical amounts of trypsin-digested ribosomal proteins in the 1D and 2D chromatography experiments. The number of proteins identified based on two or more unique peptide sequences is shown in black. The numbers above the columns are the total numbers of unique proteins identified. (C) Schematic of the proteins identified from the purified ribosome 80S complex using DALPC. The numbers below the icons identify the protein. Parentheses around a number indicate that the protein was identified on the 2D gel in Figure 4D. The protein names have been put into a web site[32]. (D) A 120 μg fraction of the purified ribosomal complex displayed on a silver-stained 2D gel. A white dot indicates the spot was sequenced using methods already described[18]. The arrow points to the 2D gel protein identified as YMR116p (BEL1). The largest and smallest calculated pI and M_r (Da) values of the identified 2D gel proteins are shown.

the capacity to directly identify proteins even more complex then the core 80S ribosome.

Increasing sensitivity using integrated DALPC. A serious limitation of two-column DALPC as an analytical tool for identifying proteins is the large amount of starting material required. The implication of this demand for identifying proteins in complexes scarcer than the 80S ribosome was a serious concern. The decreased sensitivity was due to the dead volumes in the chromatography, the complex electrospray interface, and the relatively high chromatographic flow rate. To improve the detection limit, we developed an integrated 2D microcapillary chromatography version of the DALPC process (Fig. 4A). Instead of two individual columns and multiple switching valves, we used a novel biphasic microcapillary column with sequential strong cation exchange and RP particles. To ionize eluted peptides, the column was coupled to the mass spectrometer using a simplified microelectrospray ionization interface[19]. By replacing the 1 mm i.d. columns with a 100 μm microcapillary, the mobile-phase flow rate was reduced from 50 μl/min to 0.3 μl/min. Compared with a 1,000 fmol detection limit using a 1 mm column, the microcapillary device has a 10 fmol detection limit for identifying a single trypsin-digested protein[19].

To demonstrate the concepts of the integrated DALPC process, we analyzed 0.2 μg of tryptic peptides from the yeast ribosomal 40S subunit. As a caveat, the amount of 40S peptides loaded onto the column may not reflect the amount of starting subunit proteins. Sample losses during digestion and desalting may be significant. After loading the peptide mixture onto the column, acetonitrile gradients were run through the biphasic column. The repeated RP elution gradients did not affect the retention of peptides in the SCX phase (Fig. 4B). Salt step gradients followed by acetonitrile gradients were required to detect peptides from the biphasic column. As was

Figure 3. Identifying ribosomal proteins in a total yeast extract using DALPC. A 375 μg aliquot of an acidified total yeast protein digest was analyzed using the DALPC conditions described in Experimental protocol, except 21 salt step gradients were used. The meanings of the icons are explained in Figure 2C and the protein names have been put into a web site[32]. The identified proteins have been grouped into two classes, ribosomal and other. The SCX step gradients were: 0%, 0–10%, 10–12.5%, 12.5–15%, 15–17.5%, 17.5–20%, 20–22.5%, 22.5–25%, 25–27.5%, 27.5–30%, 30–32.5%, 32.5–35%, 35–37.5%, 37.5–40%, 40–50%, 50–60%, 60–70%, 70–80%, 80–90%, 90–100% of SCX-B, and 100% SCX-C.

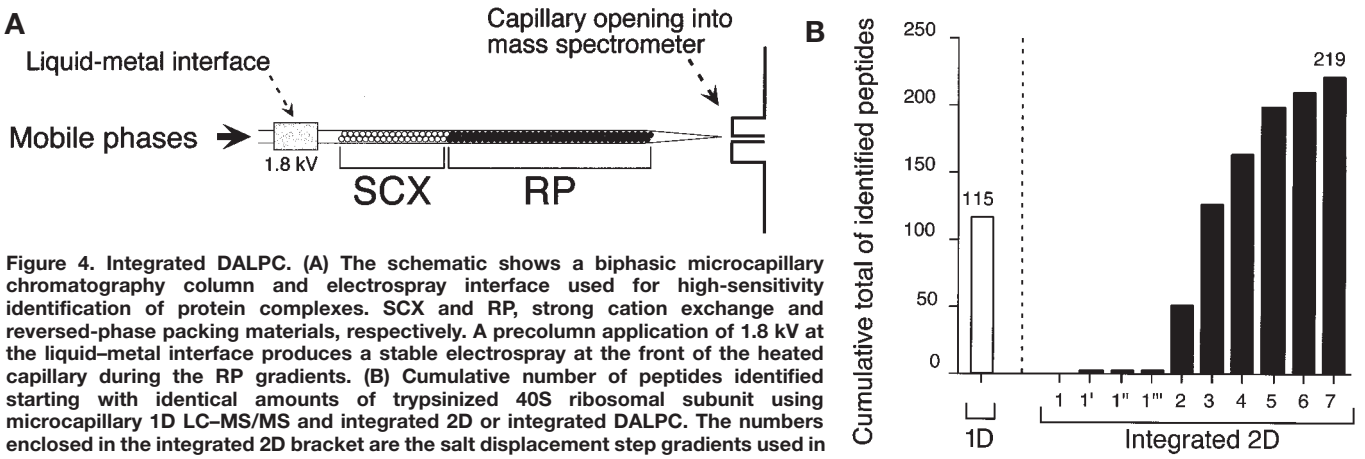


Figure 4. Integrated DALPC. (A) The schematic shows a biphasic microcapillary chromatography column and electrospray interface used for high-sensitivity identification of protein complexes. SCX and RP, strong cation exchange and reversed-phase packing materials, respectively. A precolumn application of 1.8 kV at the liquid–metal interface produces a stable electrospray at the front of the heated capillary during the RP gradients. (B) Cumulative number of peptides identified starting with identical amounts of trypsinized 40S ribosomal subunit using microcapillary 1D LC–MS/MS and integrated 2D or integrated DALPC. The numbers enclosed in the integrated 2D bracket are the salt displacement step gradients used in the experiment (see Experimental protocol). The numbers above the columns are the final numbers of unique peptides identified.

observed earlier, the number of unique peptides that could be identified increased significantly using the integrated DALPC process compared with a 1D RP microcapillary column experiment (Fig. 4B). The process identified 52 unique polypeptides, which included 30 of the 32 predicted 40S ribosomal proteins (Fig. 5). With a low nanogram detection limit, the ability of integrated DALPC to identify proteins in complex mixtures is equivalent to or better than the detection of proteins on silver-stained polyacrylamide gels.

Localizing a novel ribosomal component to the 40S subunit. We consistently found one protein cosedimenting in the complex that had not previously been identified as part of the ribosome: YMR116C (BEL1) (Fig. 2C, icon 76). YMR116p was found to be associated with the ribosome complex purified from the diploid strain 12441 (*MAT*a/*MAT*α *his7/his7 ura3/ura3*) and the haploid strain AAY1048 (*MAT*a *his3 leu2 lys2 ura3*) growing in either rich or minimal media. The protein was also identified on the 2D gel of the 80S complex (Fig. 2D). Comparison of the intensity of the 2D gel spot identified as YMR116p with those of other ribosomal proteins indicated YMR116C was present at equimolar concentration to other ribosomal proteins. Sequence analysis of YMR116C showed the gene has a single intron structure similar to that of the majority of yeast ribosomal genes[20], suggesting that YMR116p is an integral ribosomal protein in yeast. Discrepancies in the composition of the ribosome using earlier electrophoretic methods and the relative acidity of the protein probably caused YMR116p to be missed in previous studies. The pI of YMR116p is 6.04 compared with an average pI of 10.03 (standard deviation = 1.89) for the other 78 ribosomal proteins. We deleted YMR116C from the yeast genome. The deletion strain was viable but had a reduced growth rate and an abnormally large cell size compared with the wild-type haploid strain.

We fractionated the yeast 40S and 60S subunits from a total yeast lysate through a 1 M KCl sucrose gradient and separately pooled the fractions for each subunit (Fig. 5). After trypsin digestion, we analyzed each subunit using the integrated DALPC system. The acquired tandem mass spectra from the experiments was first correlated with the sequence from YMR116p to see if the protein sedimented in either fraction. YMR116p was found in the 40S but not the 60S subunit (Fig. 5). The data were then correlated with the entire yeast ORF database to verify the purity of the 40S–60S fractions. These results indicated that YMR116p was tightly associated with the 40S ribosomal subunit.

In a similarity search of protein databases, YMR116p showed significant sequence similarity throughout its sequence to the human protein RACK1 (54% amino acid identity). RACK1 has been previously identified as a putative intracellular receptor for activated protein kinase C (PKC)[21]. We wanted to determine whether RACK1 is associated with the human ribosome. A HeLa cell protein lysate was fractionated on a sucrose gradient. Fractions containing ribosomal subunits, 80S monosomes, and polysomes were collected (Fig. 6). To look for RACK1 sedimenting in the fractions, we rapidly screened an equivalent amount of each fraction using microcapillary HPLC–electrospray ionization–tandem mass spectrometry. RACK1 was absent in the fraction containing the 60S subunit alone, but present in fractions containing the 40S subunit including the polysome fractions (Fig. 6). The data strongly suggested that RACK1 is the human homolog of YMR116p and is a component of the human 40S ribosomal subunit. Like most other ribosomal proteins, the function of RACK1 is unknown. However, the indication that the ribosome-associated RACK1 protein is an intracellular receptor for activated protein kinase C (PKC) suggests a role of translation in PKC-mediated signal transduction.

Discussion

Direct analysis of protein complexes by mass spectrometry is a compelling approach to comprehensively identifying protein components. It provides a list of actual proteins present in a purified complex instead of a descriptive visualization of the components that must be individually identified later. Typically, 1D or 2D gel electrophoresis, a time- and labor-intensive process with limited molecular mass or pI ranges, is used to resolve complicated protein mixtures into individual bands or spots. Peptides from digested proteins must be recovered from the stained gel or an electroblot of the gel. Automation of this process requires expensive robotics to cut out and process the spots. Directly identifying proteins from complexes bypasses the potential limitations of gel electrophoresis, including protein insolubility, limited fractionation ranges, and limited recoveries of material. In addition, DALPC provides a highly automatable system and rapid process for repeated analysis of protein complexes.

There is a strong incentive to extract quantitative information from mass spectrometry data to directly measure changing stoichiometric ratios of proteins in complexes isolated under different biological circumstances and purification conditions. In general, the use of mass spectrometry to quantify amounts of individual proteins in a complex mixture is problematic. Incomplete proteolytic digestion, unequal loss of peptides, and differences in peptide ionization efficiencies currently make quantitative analysis unreliable. Experiments intended to produce quantitative measurements will need to incorporate appropriate controls and internal standards. We and others are working on this problem.

Broader applications of the DALPC process will depend on defining the dynamic range and the maximum number of proteins that the process can identify. Measurement of the dynamic range will be

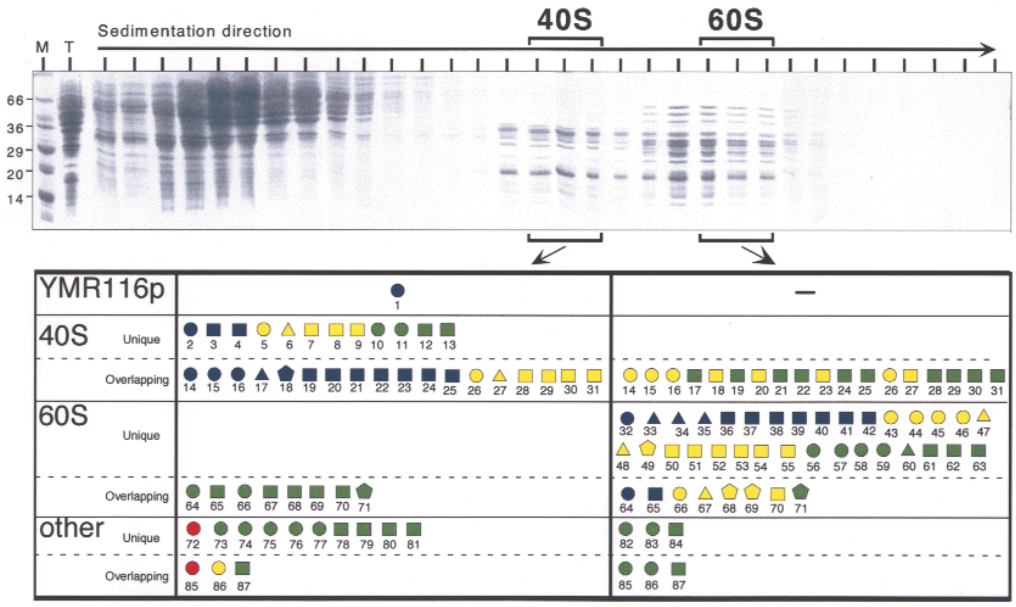


Figure 5. Localizing YMR116p to the yeast 40S ribosomal subunit. A Coomassie brilliant blue–stained 15% SDS–PAGE gel of sucrose-gradient fractions of a total yeast lysate. 40S and 60S are pooled ribosomal subunit fractions. (M, protein standards; T, protein lysate layered on the top of the sucrose gradient). Aliquots of 0.2 and 0.25 μg from the digested and desalted 40S and 60S subunit fractions, respectively, were analyzed using the integrated DALPC conditions described in Experimental protocol. The table shows the proteins identified in the fractions. The identified proteins have been divided into four classes (YMR116p, 40S, 60S, and other), and each class has been subdivided into proteins unique to a fraction or proteins identified in both fractions. The meaning of the icons is explained in Figure 2C, and the protein names have been put into a web site[32].

dependent on the complexity of the protein mixture and the resolving power of the liquid chromatography. Changing the number of salt displacement cycles, the slope of the RP gradients, and the chromatography–electrospray interface makes DALPC a process tunable to the complexity of the starting protein complex. Ultimately, the ability to identify rare proteins will be limited only by the detection limit of the mass spectrometer and not by limits imposed by protein visualization and sample manipulation techniques.

Note added in proof

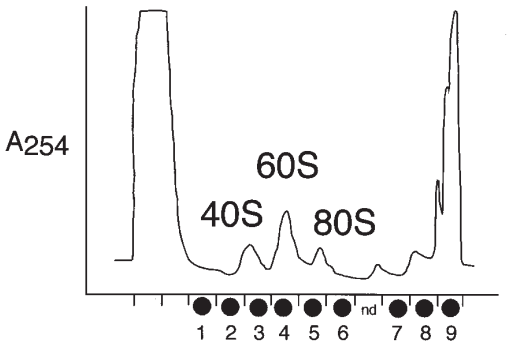
Chantrel *et al.*[22] have recently shown that by immunoblotting ribosomal polysome profiles, the yeast protein ASC1 or BEL1 is associated with the yeast 40S ribosomal subunit.

Experimental protocol

Purifying yeast 80S ribosomes. Yeast strain 12441 (*MAT*a/*MAT*α *ura3 his3*) was grown in YPD to a cell density approximately 3×10^7 cells/ml. The 80S ribosomes were purified using discontinuous sucrose gradients essentially as described[23]. The pelleted ribosome complexes were resolubilized and denatured in 8 M urea, 1 M NaCl, 100 mM β-mercaptoethanol (β-ME), 100 mM Tris-HCl pH 7.5. Ribosomal RNA was extracted using acetic acid and the proteins lyophilized essentially as described[24].

Purifying yeast 40S–60S ribosomal subunits. Yeast strain BJ5622 (ref. 25) was grown in YPD to a cell density of approximately 5×10^7 cells/ml. The subunits were purified from a total cell lysate using a 10–40% sucrose gradient essentially as described[26]. The fractionated proteins were precipitated[27] and resuspended in 8 M urea, 200 mM NH_4HCO_3, 20 mM $CaCl_2$. Proteins were quantitated using a Bradford assay. The subunit fractions were diluted fourfold with dH_2O, modified trypsin (Boehringer-Mannheim, Germany) was added to a final substrate-to-enzyme ratio of 50:1, and the reaction was incubated at 37°C for 15 h. The peptide mixture was desalted on an RP column, lyophilized, and resuspended in 0.5% acetic acid. The desalted peptide mixtures were quantitated using amino acid composition.

Preparing total yeast extracts. Yeast strain 12441 was grown in YPD to a cell density of 1.5×10^7 cells/ml. Cells were disrupted using glass beads in the



	1	2	3	4	5	6	7	8	9
RACK1	-	-	+	-	-	-	+	-	+
40S	-	-	5	-	-	-	5	4	17
60S	-	-	-	16	15	7	14	15	24
Other	11	25	8	12	9	9	27	14	8

Figure 6. Localizing RACK1 to the human 40S ribosomal subunit. Ultraviolet trace of the HeLa cell extract fractionated on a sucrose gradient. Dark balls indicate the fractions analyzed by microcapillary HPLC–electrospray ionization–tandem mass spectrometry. nd, Not done. A 20% aliquot of each fraction was analyzed by microcapillary 1D LC–MS/MS. The column was eluted with a gradient of 0–60% acetonitrile in 0.5% acetic acid over 90 min at 0.5 μl/min. Using the SEQUEST program, the acquired spectra were first correlated with the RACK1 protein sequence and then with Frederick Biomedical Supercomputing Center nonredundant protein database (release 980530) (Frederick, MD). For each fraction, the table shows whether RACK1 was identified and the total number of human 40S, 60S, and nonribosomal (other) proteins identified. The names of the identified proteins have been put into a web site[32]. In fractions 3, 7, and 9, RACK1 was identified based on acquired tandem mass spectra correlating with the RACK1 peptide sequence IIVDELKQEVISTSSK with an *m/z* 896 for the $(M+2H)^{2+}$ ion. An ion with the same *m/z* and similar retention time was detected in fraction 8 but was below the ion abundance threshold for being fragmented. No ion signal above background noise with a similar *m/z* and elution time was detected in fractions 1, 2, 4, 5, and 6.

presence of protease inhibitors[28]. Urea, NaCl, and β-ME were added to final concentrations of 8 M urea, 1 M NaCl, and 100 mM β-ME, and the lysate centrifuged at 30,000 *g* for 30 min. Nucleic acids were extracted using acetic acid essentially as described, and the proteins were lyophilized[24].

Purifying human ribosomal complexes. The human ribosomal subunits were purified by sucrose gradient centrifugation essentially as described[29]. A cytosolic extract from HeLa cells (approximately 7×10^7 cells) was prepared and layered onto the top of 15–50% sucrose gradient. After centrifugation at 36,000 r.p.m. in a Beckman SW-40 rotor (Beckman Instruments, Fullerton, CA) for 8 h, the gradient was fractionated into 12 fractions of 1-ml each using an Isco Density Gradient Fractionator (model 185; Isco, Lincoln, NE), while monitoring absorbance at 254 nm. The fractions were processed as described for the yeast 40S–60S protein complexes.

Digestion of 80S and total yeast extracts. Lyophilized proteins were resolubilized in 8 M urea, 200 mM NH_4HCO_3, and 20 mM $CaCl_2$ and quantified using a Bradford assay. Endoprotease Lys-C (Boehringer-Mannheim) was added to a final substrate-to-enzyme ratio of 100:1, and the reaction was incubated at 37°C for 15 h. The Lys-C digestion was diluted fourfold with dH_20, and modified trypsin (Boehringer-Mannheim) was added to a final substrate-to-enzyme ratio of 50:1. The trypsin digestion was incubated at 37°C for 15 h. The peptide mixture was desalted on an RP column, lyophilized, and resuspended in 5 mM K_2HPO_4, 5% acetonitrile (pH 3).

2D gels of 80S ribosomes. Lyophilized proteins from the 80S ribosome were resolubilized in 9 M urea, 4% 3-[(3-Cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 2% 3-10 Serva ampholytes, 1% dithiothreitol (DTT), and resolved using nonequilibrium 2D gel electrophoresis[18].

Mass spectrometry. All analyses were performed on Finnigan LCQ ion trap mass spectrometer (Finnigan Corp., San Jose, CA) that was run and operated as described[18].

1D and 2D LC–MS/MS. An Integral chromatography workstation (PE Biosystems, Foster City, CA) was configured either with a 1.0 × 250 mm Vydac C_{18} (218TP) column (The Separations Group, Hesperia, CA) or with a 1.0 × 150 mm PolySULFOETHYL Aspartamide (PolyLC, Columbia, MD) and 1.0 × 250 mm Vydac C_{18} (218TP) columns. The workstation was directly coupled to an LCQ ion trap mass spectrometer equipped with an electrospray ion source. The electrospray needle was operated with a voltage differential of 5.5 kV, and the heated desolvation capillary was held at 250°C.

Microcapillary 1D LC–MS/MS. A Finnigan LCQ ion trap mass spectrometer equipped with a fritless electrospray interface and an RP microcapillary–HPLC column (100 μm i.d. packed with 8 cm of 5 μm C_{18} RP particles (218TP $C_{18,}$ Vydac) was run and operated essentially as described[19]. For the microcapillary 1D LC–MS/MS experiment, a 0.2 μg aliquot of the digested and desalted 40S ribosomal peptide mixture was analyzed[19]. The RP column was eluted into the mass spectrometer with a linear gradient of 0 to 60% RP-B over 30 min at 300 nl/min. For the integrated DALPC experiment, 0.2 μg aliquot of the digested and desalted ribosomal 40S subunit was loaded onto the biphasic microcapillary. Peptides fractions were displaced from the SCX to the RP particles using the following salt step gradients: (1) 0% (1') 0% (1''), 0% (1''') 0% (2) 0–10% (3) 10–20% (4) 20–30% (5) 30–40% (6) 40–100% of SCX-B', and (7) 100% SCX-C'. Peptides were eluted from the RP particles into the mass spectrometer using a linear gradient of 0–60% RP-B over 30 min at 300 nl/min. Mobile-phase buffers were, for RP-A buffer, 0.5% acetic acid, 5% acetonitrile; for RP-B, 0.5% acetic acid, 80% acetonitrile; for SCX-B', 0.5% acetic acid, 5% acetonitrile, 250 mM KCl; for SCX-C', 0.5% acetic acid, 5% acetonitrile, 500 mM KCl.

Biphasic microcapillary column. To construct the biphasic column, a fused-silica capillary (100 μm i.d. × 365 μm o.d.) was pulled with a CO_2-based laser puller to make a fritless column as described[19]. The column was packed first with 8 cm of 5 μm C_{18} RP particles (218TP $C_{18,}$ Vydac) and then with 4 cm of 5 μm strong cation exchange particles (PolySULFOETHYL Aspartamide; PolyLC). For the 1D LC–MS/MS experiment in Figure 2, 120 μg of the 80S peptide mixture were loaded directly onto an RP column equilibrated in RP-A. The RP column was then eluted into the mass spectrometer with a linear gradient of 5–60% RP-B over 60 min at 50 μl/min. For the 2D LC–MS/MS experiment in figure 2, the peptide mixture was loaded onto an SCX column equilibrated in SCX-A. Peptide fractions were displaced from the SCX to the RP column using the following salt step gradients: (1) 0% (2) 0–10% (3) 10–15% (4) 15–20% (5) 20–25% (6) 25–30% (7) 30–35% (8) 35–40% (9) 40–55% (10) 55–70% (11) 70–100% of SCX-B, and (12) 100% SCX-C. Peptides were eluted from the RP column into the mass spectrometer using a linear gradient of 5–60% RP-B over 60 min at 50 μl/min. Mobile-phase buffers were, for SCX-A, 5 mM K_2HPO_4, 5% acetonitrile (pH 3); for SCX-B, 250 mM KCl, 5% acetonitrile (pH 3); for SCX-C, 500 mM KCl, 5% acetonitrile (pH 3); for RP-A, 0.5% acetic acid; for RP-B, 0.5% acetic acid, 80% acetonitrile.

Analysis of MS/MS data. For automated spectrum and data analysis, each raw tandem spectrum was processed as described here. Spectra derived from single or multiply charged parent ions were first identified. For spectra from a multiply charged peptide, an independent search was performed on both the +2 and +3 mass of the parent ion. Except where noted, processed tandem mass spectra were correlated with the standard *S. cerevisiae* ORFs[30] using the program SEQUEST running on a DEC Alpha workstation[6]. All searches were performed without considering the protease used because we observed that many proteins in the mixtures do not digest to completion. For multiply charged peptides, the following criteria were used to determine whether to select the +2 or +3 charge state: (1) A particular charge state was chosen if the cross-correlation score was ≥1 U more than that of the other charge state. (2) A score was assigned to each charge state (+5 for tryptic start, +5 for tryptic end, +2 if the cross-correlation score is greater than the other charge state, +2 if the preliminary score ranking is <50), and the charge state with the highest score was chosen. (3) If the charge state scores were equal, then the charge state with the highest cross-correlation score was chosen. (4) If both charge states had the same cross-correlation score, the +2 charge state was chosen. The selected charge of the peptide was used in the final protein identification analysis and the SEQUEST output from the other charge state discarded. The correlation results were then filtered using the value of the cross-correlation score and the matched sequence for each spectrum. For singly charged peptides, spectra with a cross-correlation score to a tryptic peptide ≥1.5 were retained. For multiply charged peptides, spectra with a cross-correlation to a tryptic peptide ≥2 were retained. All spectra with cross-correlation scores not meeting these criteria were eliminated from further consideration.

For the protein identifications, the filtered results were sorted to show unique peptide sequences that were derived from the same annotated ORFs in the genome. Protein identifications based on mass spectra correlating to one or more unique tryptic peptides were considered valid identifications. Single peptides that alone identify a protein were manually validated after meeting the following criteria. First, the SEQUEST cross-correlation score must be >1.5 for a +1 tryptic peptide or >2 for a +2 or +3 tryptic peptide. Second, the MS/MS spectrum must be of good quality with fragment ions clearly above baseline noise. Third, there must be some continuity to the b or y ion series. Fourth, the y ions that correspond to a proline residue should be intense ions. Fifth, unidentified, intense fragment ions either corresponds to +2 fragment ions or the loss of one or two amino acids from one of the ends of the peptide. After going through this process we are fairly confident of a protein identification. If there is any doubt, we err on the side of caution. Typically, 35–50% of the protein identifications based on a single peptide sequence are accepted.

Deletion of YMR116C. A deletion of YMR116C in the yeast diploid strain 12441 was created using the method of Güldener *et al*[31]. The DNA primers `TTTCCAAAAAATCCTTATAACACACTAAAGTAAATAAAGTGAAAAG-CATAGGCCACTAGTGGATCTG` and `TAACTAGAAGATACATAAAAGAA-CAAATGAACTTTATACATATTCCAGCTGAAGCTTCGTACGC` were used to amplify by PCR the loxP-kanMX-loxP YMR116C disruption cassette, and the PCR product was transformed into the yeast strain 12441. The deletion was verified using PCR and DNA primers flanking YMR116C (`GGGCTATTC-CTTTAATTGTGG` and `TTATAGTACCTAATGATAACC`). The disruption cassette was excised and the diploid was sporulated.

Acknowledgments

We thank D. Tabb, L. Hayes, G. Kleeman, E. Malone, and M. Olson for critical reading of the manuscript and T. Gatlin for assistance with the figures. Partially supported by NIH grant CA39053 (G.T.M and D.R.M). Supported by National Center for Research Resources Yeast Center grant RR11823, NIH grant GM52095 and NSF's Science and Technology Center grant BIR 8809710 to J.R.Y., and NIH postdoctoral fellowship grant T32 HG00035-03 to A.J.L.

1. Pugh, B.F. Mechanisms of transcription complex assembly. *Curr. Opin. Cell Biol.* **8,** 303–311 (1996).
2. Wool, I.G., Chan, Y.-L. & Gluck, A. Structure and evolution of mammalian ribosomal proteins. *Biochem. Cell Biol.* **73**, 933–947 (1995).
3. Mager, W.H. *et al*. A new nomenclature for the cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Nucleic. Acids Res.* **25**, 4872–4875 (1997).
4. Phizicky, E.M. & Fields, S. Protein–protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59**, 94–123 (1995).
5. Neubauer, G. *et al.* Identification of the proteins of the yeast U1 small nuclear

ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci. USA* **94**, 385–390 (1997).
6. Eng, J.K., McCormack, A.L. & Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Amer. Soc. Mass Spectrom.*. **5**, 976–989 (1994).
7. Link, A.J., Carmack, E. & Yates, J.R. A strategy for the identification of proteins localized to subcellular spaces: applications to *E. coli* periplasmic proteins. *Int. J. Mass Spectrom. Ion Proc.* **160**, 303–316 (1997).
8. McCormack, A.L. *et al*. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **69**, 767–776 (1997).
9. Giddings, J.C. Concepts and comparisons in multidimensional separation. *J. High Resol. Chromatogr. Commun.* **10**, 319–323 (1987).
10. Lundell, N. & Markides, K. Two-dimensional liquid chromatography of peptides: an optimization of strategy. *Chromatographia* **34**, 369–375 (1992).
11. O'Farrell, P.H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007–4021 (1975).
12. Takahashi, N., Ishioka, N., Takahashi, Y. & Putnam, F.W. Automated tandem high-performance liquid chromatographic system for separation of extremely complex peptide mixtures. *J. Chromatogr.* **326**, 407-418 (1985).
13. Opiteck, G.J., Lewis, K.C. & Jorgenson, J.W. Comprehensive on-line LC/LC/MS of proteins. *Anal. Chem.* **69**, 1518–1524 (1997).
14. Yates, J.R., Eng, J.K., McCormack, A.L. & Schieltz, D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67**, 1426–1436 (1995).
15. Otaka, E. & Osawa, S. Yeast ribosomal proteins: V. Correlation of several nomenclatures and proposal of a standard nomenclature. *Mol. Gen. Genet.* **181**, 176–182 (1981).
16. Mager, W.H. & Planta, R.J. Coordinate expression of ribosomal protein genes in yeast as a function of cellular growth rate. *Mol. Cell. Biol.* **104**, 181–187 (1991).
17. Warner, J. Synthesis of ribosomes in *Saccharomyces cerevisiae*. *Microbiol. Rev.* **53**, 256–271 (1989).
18. Link, A.J., Hays, L.G., Carmack, E.B. & Yates, J.R. Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143. *Electrophoresis* **18**, 1314–1334 (1997).
19. Gatlin, C.L., Kleemann, G.R., Hays, L.G., Link, A.J. & Yates, J.R. Protein identification at the low femtomole level from silver stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal. Biochem.* **263**, 93–101 (1998).
20. Cherry, J.M. *et al*. Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* **387**, 67–73 (1997).
21. Ron, D. *et al*. Cloning of an intracellular receptor for protein kinase C: a homolog of the β subunit of G proteins. *Proc. Natl. Acad. Sci. USA* **91**, 839–843 (1994).
22. Chantrel, Y., Gaisne, M., Lions, C. & Verdiere, J. The transcriptional regulator Hap1p (Cyp1p) is essential for anaerobic or heme-deficient growth of *Saccharomyces cerevisiae*: genetic and molecular characterization of an extragenic suppressor that encodes a WD repeat protein. *Genetics* **148**, 559–569 (1998).
23. Battaner, E. & Vazquez, D. Preparation of active 60S and 40S subunits from yeast ribosomes. *Methods Enzymol.* **20**, 446–449 (1971).
24. Warner, J.R. & Gorenstein, C. in *Methods in cell biology*. (ed. Prescott, D.M.) 45–60 (Academic, New York; 1978).
25. Jones, E.W. Tackling the protease problem in *Saccharomyces cerevisiae*. *Methods. Enzymol.* **194**, 428–453 (1991).
26. Raue, H.A., Mager, W.H. & Planta, R.J. Structural and functional analysis of yeast ribosomal proteins. *Methods Enzymol.* **194**, 453–477 (1991).
27. Wessel, D. & Flügge, U.I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984).
28. Drubin, D.G., Miller, K.G. & Botstein, D. Yeast actin-binding proteins: evidence for a role in morphogenesis. *J. Cell Biol.* **107**, 2551–2561 (1988).
29. Ruan, H., Brown, C.Y. & Morris, D.R. in *mRNA formation and function*. (ed. Richter, J.D.) 305–321 (Academic, New York; 1997).
30. *Saccharomyces* genome database at http://genome-www.stanford.edu/Saccharomyces/.
31. Güldener, U., Heck, S., Fiedler, T., Beinhauer, J. & Hegemann, J.H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* **24**, 2519–2524 (1996).
32. Direct Analysis of Large Protein Complexes (DALPC) at http://thompson.mbt.washington.edu/dalpc/.