Analysis of the *Arabidopsis* Cytosolic Ribosome Proteome Provides Detailed Insights into Its Components and Their Post-translational Modification*

Adam J. Carroll‡, Joshua L. Heazlewood§, Jun Ito¶, and A. Harvey Millar∥

Finding gene-specific peptides by mass spectrometry analysis to pinpoint gene loci responsible for particular protein products is a major challenge in proteomics especially in highly conserved gene families in higher eukaryotes. We used a combination of *in silico* approaches coupled to mass spectrometry analysis to advance the proteomics insight into *Arabidopsis* cytosolic ribosomal composition and its post-translational modifications. *In silico* digestion of all 409 ribosomal protein sequences in *Arabidopsis* defined the proportion of theoretical gene-specific peptides for each gene family and highlighted the need for low m/z cutoffs of MS ion selection for MS/MS to characterize low molecular weight, highly basic ribosomal proteins. We undertook an extensive MS/MS survey of the cytosolic ribosome using trypsin and, when required, chymotrypsin and pepsin. We then used custom software to extract and filter peptide match information from Mascot result files and implement high confidence criteria for calling gene-specific identifications based on the highest quality unambiguous spectra matching exclusively to certain *in silico* predicted gene- or gene family-specific peptides. This provided an in-depth analysis of the protein composition based on 1446 high quality MS/MS spectra matching to 795 peptide sequences from ribosomal proteins. These identified peptides from five gene families of *Arabidopsis* were confirmed by trypsin digestion of all 409 ribosomal protein sequences in silico and its post-translational modification. We undertook an extensive MS/MS survey of the cytosolic ribosome using trypsin and, when required, chymotrypsin and pepsin. We then used custom software to extract and filter peptide match information from Mascot result files and implement high confidence criteria for calling gene-specific identifications based on the highest quality unambiguous spectra matching exclusively to certain *in silico* predicted gene- or gene family-specific peptides. This provided an in-depth analysis of the protein composition based on 1446 high quality MS/MS spectra matching to 795 peptide sequences from ribosomal proteins. These identified peptides from five gene families of *Arabidopsis* were confirmed by trypsin digestion of all 409 ribosomal protein sequences in silico and its post-translational modification.

Ribosomes are large ribonucleoprotein complexes that catalyze the peptidytransferase reaction in polypeptide synthesis and are thus responsible for the translation of transcripts encoded in cellular genomes. These complexes play the most fundamental role of any protein complex in the generation of the cellular proteome as a whole. Ribosomes consist of two subunits, large and small, but the internal composition of these subunits and their macromolecular size varies between bacteria, animals, fungi, and plants. Both these subunits are composed on rRNA and protein (r-protein) components. Among eukaryotes the 80 S cytosolic ribosomes of the yeast (*Saccharomyces cerevisiae*), rat (*Rattus norvegicus*), and human (*Homo sapiens*) have been the most extensively investigated. These studies have revealed four distinct rRNAs, the 18 S rRNA of the 40 S subunit and 5, 5.8, and 23 S rRNAs of the 60 S subunit. Up to 79 distinct proteins are part of these subunits in eukaryotes, 32 small subunit and 47 large subunit proteins, compared with only 54 proteins in the bacterial ribosome subunits (1). In eukaryotes a separate bacterial-like ribosome is present in the mitochondria and is referred to as a 70 S ribosome containing a large 50 S and small 30 S subunit; its smaller overall size is reflected in altered rRNA sizes and different protein subunits (2). Detailed analyses of yeast, rat, and human cytosolic ribosomes by peptide mass spectrometry have provided insights into the composition and post-translational modification of ribosomal proteins (3–7). Methionine removal, acetylation, methylolation, and phosphorylation are known ribosomal protein modifications, and both phosphorylation and methylation have been linked to changes in ribosomal function or biosynthesis (8, 9).

Early studies of the cytosolic ribosome in plants revealed it was slightly smaller than that found in mammals, notably due to lower apparent mass of the 60 S subunit (10).Early biochemical analyses of the protein components of ribosomes have been undertaken using 2D gels in the monocotyledon-
ous plants wheat, barley, and maize and the dicotyledonous plants soybean, tomato, and tobacco (11–16). Counting of distinct protein spots on these gels suggested that the 40 S subunit contained up to 40 proteins, whereas the 60 S subunit contained up to 59 proteins; however, without genomic sequences it was not possible to systematically assign these proteins to genes and gene families that would be required to resolve the composition of the complex in plants in a gene-specific manner.

The sequencing of the Arabidopsis genome provided the first opportunity in plants to consider the number and arrangement of ribosomal protein-coding genes in plants. Using the strong sequence conservation of the eukaryotic r-proteins, Barakat et al. (17) undertook an analysis of expressed sequence tags, and the early annotation of the complete genomic sequence identified 249 genes including 22 apparent pseudogenes that encoded 80 different types of r-proteins (32 small subunit and 48 large subunit proteins). The extra family of proteins not conserved in mammals was the plant-specific P3 component known to be in the 60 S subunit. Analysis of the r-protein gene families reveals that each family consists, on average, of three members. The sequences within these families can have very high conservation, leading to many paralogs with 97–100% sequence identity at the amino acid level, whereas other r-protein families contain significant sequence divergence. Based on public EST data and hybridization data on microarrays, most gene family members are expressed and could be present in the ribosome structure at different points of development, in different cell types, and under different conditions (18). Variation in ribosome composition by incorporation of different paralogs could be an important component in the regulation of transcript translation. This underpins the importance of a thorough understanding of the actual composition of ribosome protein complexes themselves and not just the set of genes that could encode ribosomes.

Several studies have sought to investigate the ribosomal protein complement of experimental samples from Arabidopsis to answer these questions. Chang et al. (19) performed a study using ribosomes purified from Arabidopsis cell culture. They undertook a MALDI-TOF analysis of spots from 2D gel-separated protein samples and also a limited tandem MS analysis of 1D SDS-PAGE-separated protein bands. Giavalisco et al. (20) undertook a similar analysis of ribosome samples extracted from whole Arabidopsis leaves using 2D gel-separated samples and MALDI-TOF MS analysis.

Chang et al. (19) identified proteins from 70 of the 80 gene families by their MALDI-TOF analysis and identified members from a further four gene families through their MS/MS analysis of 1D gel bands. In contrast Giavalisco et al. (20) could only identify 63 of the 80 gene families by their MALDI-TOF analysis. Both these studies made claims of products from particular genes within a large number of the gene families. This was often based on the presence of single MS ion masses, but often not even these could be found, and the proteins could only redundantly be linked back to gene families of two to seven members.

We used a combination of approaches to advance the MS/MS-based insight into Arabidopsis ribosomal composition and its post-translational modifications. First, we used an in silico digestion of all ribosomal proteins to define targets for data acquisition and to drive a strategy of data collection to maximize recognition of ribosomal protein-derived peptides. Second, we undertook an extensive MS/MS survey of the ribosome using trypsin and, when required, chymotrypsin and pepsin. We then used custom software to extract and filter peptide match information from Mascot result files and implement high confidence criteria for calling gene-specific identifications based on the highest quality unambiguous spectra matching exclusively to certain in silico predicted gene-specific peptides. This provided a much richer and more detailed analysis of the protein composition and also identified peptides from five gene families of r-proteins not identified in previous studies. Further we acquired strong MS/MS data to support gene-specific protein identification in 32 cases not revealed in the previous study by Chang et al. (19).

In addition, we provide a wealth of information on co- and post-translational modification of r-proteins in Arabidopsis by initiator methionine removal, N-terminal acetylation, N-terminal methylation, lysine N-methylation, and phosphorylation.

**EXPERIMENTAL PROCEDURES**

*In Silico Digestion Analysis—*Predicted sequences for Arabidopsis ribosomal proteins were obtained from the file “ATH1_pep_20051108,” which was downloaded from The Arabidopsis Information Resource (TAIR) website (www.Arabidopsis.org) as part of the TAIR6 Arabidopsis Genome Release. A modified version of the published Perl script, Proteogest (21), and custom hypertext preprocessor (PHP) scripts were used to populate peptide database tables containing rows for every tryptic, chymotryptic, or peptic peptide sequence predicted from the set of 409 protein sequences annotated as ribosomal proteins or ribosomal protein-like proteins (allowing one missed cleavage) whereby each row represented an individual peptide-to-gene association. In addition to columns for peptide sequence and Arabidopsis Gene Index (AGI), the tables also included a column containing the name of the r-protein family to which each gene belonged according to the nomenclature of Barakat et al. (17). These tables were used to map Mascot-reported peptide match information onto r-proteins and r-protein families and were processed by additional PHP scripts to generate lists of peptide sequences that were only derived from single genes (gene-specific) and calculate the predicted m/z ratios of [M + 2H]^2+ ions of total and gene-specific peptides for r-protein peptide ion m/z distribution analysis.

**Ribosomal Isolation—**All steps were carried out on ice or at 4 °C. Approximately 100 g fresh weight of Arabidopsis cell culture was homogenized in 300 ml of extraction buffer (0.45 M mannitol, 30 mM Tris, 0.5% (w/v) polyvinylpyrrolidone 40, 100 mM KCl, 20 mM MgCl2, 0.5% (w/v) bovine serum albumin, 20 mM cysteine) in a Waring blender for 2 min, and the homogenate was filtered through four layers of muslin. The filtrate was centrifuged for 5 min at 1500 g, and the supernatant was again centrifuged at 16,000 × g for 15 min. The supernatant of this 16,000 × g sample was then centrifuged at 30,000 × g for 30 min. In 70-mL polycarbonate bottle assembles.
(Beckman part number 355655), 50-ml portions of the 30,000 × g supernatant were then layered over 20-ml cushions of 1.5 M sucrose dissolved in Ribosome Buffer (30 mM Tris, 100 mM KCl, 20 mM MgCl₂, 5 mM β-mercaptoethanol) and centrifuged at 180,000 × g for 14.5 h to form a crude ribosomal pellet. Each crude ribosomal pellet was resuspended in 1 ml of Ribosome Buffer and centrifuged at 20,800 × g for 30 min to pellet insoluble material. The supernatants were combined, brought to a volume of 50 ml with Ribosome Buffer, and ultracentrifuged through 1.5 M sucrose as above. The pure ribosomal pellet was resuspended to a protein concentration of 6.5 mg/ml in Ribosome Buffer, snap frozen in liquid nitrogen, and stored at −80 °C until use. This protocol yielded 2–3 mg of ribosomal protein from 100 g fresh weight of cells.

For the isolation of ribosomes from mitochondria, mitochondria were first purified from Arabidopsis by differential centrifugation and density gradient centrifugation essentially as described by Millar et al. (22). A concentrated suspension of mitochondria containing ~40 mg of mitochondrial protein was resuspended to a total volume of 50 ml in Ribosome Buffer containing 2% (w/v) Triton X-100. The suspension was incubated on ice for 30 min with occasional gentle mixing. The suspension was then clarified by centrifugation at 30,000 × g for 30 min, and the supernatant was layered over a 20-ml cushion of 1.5 M sucrose in Ribosome Buffer and centrifuged at 180,000 × g for 20 h. After removal of the supernatant, the ribosomal pellet was resuspended in a minimal volume of Ribosome Buffer, snap frozen in liquid nitrogen, and stored at −80 °C until use.

Arabidopsis Cytosolic Ribosome Dissociation—Ribosome dissociation was essentially carried out according to Lin and Key (23). Briefly 100 μg of purified cytosolic ribosomes were resuspended in 50 μl of either modified ribosome buffer (30 mM Tris, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, pH 7.5) or modified ribosome buffer containing 0.5 mM KCl. Resuspended samples were loaded onto a 10-ml linear sucrose gradient (15–30%) in either modified ribosome buffer or modified ribosome buffer containing 0.5 M KCl and subjected to ultracentrifugation at 260,800 × g at 4 °C for 4 h using a Beckman SW4 Ti rotor. Fractions of ~200 μl were collected directly into 96-well plates using a peristaltic pump. Absorbance readings (A₅₉₅ and A₆₅₀) were conducted on a POLARStar OPTIMA microplate reader (BMG Labtech) using 200 flashes/well.

**Gel Electrophoresis—**SDS-PAGE gels used were 4% acrylamide stacking gels above 12% (w/v) acrylamide, 0.1% (w/v) SDS or 5–16% (w/v) acrylamide, 0.1% (w/v) SDS in a large gel format (0.1 × 16 × 16 cm) and were run with a Tris-base running buffer system. Gel electrophoresis was performed at 25 mA/gel and completed in 3 h.

**Pro-Q Phosphoprotein Detection—**Approximately 25 μg of ribosomal proteins were solubilized in SDS-PAGE sample buffer and loaded onto a 12% polyacrylamide gel (14 cm × 16 cm × 0.75 mm) overlaid with a 4% stacking gel. The sample underwent electrophoresis for 4 h at 30 mA and upon completion was fixed in a solution consisting of 10% acetic acid and 50% methanol overnight. Following three 10-min washes in ddH₂O each, gels were stained with 100 ml of 0.1% (w/v) silver stain for 90 min and destained using three successive 30-min washes with 100 ml of destain solution containing 50 mM sodium acetate (pH 4.0) and 20% acetic acid. After two washes in ~100 ml of ddH₂O for 5 min at a time, fluorescent images of in-gel phosphorylated proteins were acquired using a Typhoon fluorescence scanner (GE Healthcare) with 532 nm excitation, a 580-nm band pass emission filter, and the photo multiplier tube set at 500 for optimal Pro-Q dye detection. ImageQuant™ software was used to view Pro-Q staining of the 1D PAGE gel. The gel was then stained with colloidal Coomassie overnight with gentle rocking. Following staining, the solution was discarded, and the gel was destained in 0.5% phosphoric acid for 4 h.

**Phosphopeptide Enrichment Using Titanium Dioxide—**TiO₂ tips (NuTip) were supplied by Glygen Inc., and phosphopeptide enrichment procedures were essentially those outlined by Larsen et al. (24) with some modifications. Briefly TiO₂ tips were conditioned prior to binding of phosphopeptides by aspiring/expelling 10 μl of 0.1% TFA solution in ddH₂O (pH ~1.9) five times using a pipette and expelling all of the TFA solution from the tip. Before binding, the pH of a 5-μl sample containing 10 μg of trypsin-digested ribosomal peptide lysate mixture was adjusted to pH ~1.9 by adding one part 1% TFA solution in ddH₂O to four parts peptide mixture. The peptide mixture was aspirated/expelled 50 times using a pipette to ensure maximal binding of phosphorylated peptides to TiO₂. The binding solution was then completely expelled from the tip. Bound samples were washed with 10 μl of 50% acetonitrile, 0.1% TFA aspirating/expelling the wash solution 10 times through the tip. This step was repeated two more times for a total of three wash steps. At the end of each wash step, all of the wash solution was expelled from the tip. Finally the bound phosphopeptides were eluted from the TiO₂ tip by aspirating/expelling 10 μl of 0.3 M NH₄OH solution in ddH₂O (pH 10.5) 10 times through the tip. Portions of this eluate were used for nano-ESI-MS/MS analysis.

**Q-TOF MS—**For gel-arrayed proteins, samples to be analyzed were cut from the gels and destained twice for 45 min in 10 mM NH₄CO₃ and 50% (v/v) acetonitrile. Samples were dehydrated at 50 °C in a dry block heater for 30 min and rehydrated with 15 μl of digestion solution, which for trypsin and chymotrypsin consisted of 25 mM NH₄CO₃ and 25 μg/ml protease in 0.01% (v/v) trifluoroacetic acid and for pepsin consisted of 25 μg/ml pepsin in 7% (v/v) formic acid, and incubated overnight at 37 °C. Peptides were extracted by adding 15 μl of acetonitrile and vigorous shaking for 15 min, removing liquid, and adding 15 μl of 50% (v/v) acetonitrile and 5% (w/v) formic acid to the gel plugs followed by another 15 min of shaking (this step was repeated); washes were pooled after each extraction step. Samples were loaded onto self-packed Microsorb (Varian Inc.) C₁₈ (5-μm, 100-Å) reverse phase columns (0.5 × 50 mm) using an Agilent Technologies 1100 series capillary liquid chromatography system and eluted into a QStar Pulsar / MS/MS system fitted with an IonSpray source (Applied Biosystems). Peptides were eluted from the C₁₈ reverse phase column at 8 μl/min using a 9-min acetonitrile gradient (5–60%) in 0.1% (v/v) formic acid. Ions were selected automatically for the N₂ collision cell utilizing the information-dependent acquisition capabilities of Analyst QS version 1.1 (Applied Biosystems) and the rolling collision energy feature for automated collision energy determination based on the m/z of the ions. The method used a 1-s TOF MS scan that automatically switched (using information-dependent acquisition) to a 2-s product ion scan (MS/MS) when a target ion reached an intensity of greater that 30 counts and its charge state was identified as 2+, 3+, or 4+. TOF MS scanning was undertaken on an m/z range of 200–900 m/z using a Q1 transmission window of 180 amu (100%). Product ion scans were undertaken at m/z ranges of 70–2000 m/z at low resolution utilizing Q2 transmission windows of 50 (33%), 190 (33%), and 650 amu (34%).

For ribosomal protein lysates, samples were digested with trypsin (1:10, w/v) in 10 mM NH₄CO₃ overnight at 37 °C. For initial analyses digested samples of ~1–10 μg were analyzed directly or after TiO₂ selection as above except bound peptides were eluted over a 30-min period (5–60% acetonitrile in 0.1% (v/v) formic acid) at 10 μl/min. The analysis method used was similar to that described above except a TOF MS scan range of 400–1200 m/z was used with a Q1 transmission window of 380 amu (100%). For more detailed analyses samples of ~0.2–1 μg were analyzed on a QStar Pulsar / MS/MS system (Applied Biosystems) fitted with a NanoES source (Protana Inc.) outfitted with a New Objective adapter (ADPT-PRO) to allow direct liquid chromatographic coupling to the source. Samples were loaded using an 1100 series capillary LC system (Agilent Technologies) onto a
capillary sample trap column containing a 100-μm × 2.5-cm C₁₈ insert (Upchurch Scientific) at 1 μl/min in 5% (v/v) acetonitrile and 0.1% (v/v) formic acid. Bound peptides were eluted at 500 nl/min into the mass spectrometer after automatically switching a retrofitted 10-port nanobore valve (Valco Inc.) on the QStar Pulsar by an 1100 series Nano Pump (Agilent Technologies). Peptides were eluted over a 30-min period and analyzed using parameters similar to those outlined above for LC-MS/MS analysis of lysates.

Resulting MS/MS-derived spectra were analyzed against an in-house Arabidopsis database comprising ATH1.pep (release 6) from TAIR and the mitochondrial and plastid protein sets (TAIR). This sequence database contained a total of 30,700 protein sequences (12,656,682 residues). Mascot Generic Format (.MGF) MS/MS peak lists were generated from raw Sciex Analyst format (.WIFF) data files with Mascot Daemon using the “mascot.dll 1.6b21 for Analyst QS 1.1” data import filter available from Matrix Science. The settings used for MS/MS peak list generation were: centroid survey scan ions (TOF MS) at a height percentage of 50% and a merge distance of 0.1 amu (for charge state determination); centroid MS/MS data at a height percentage of 50% and a merge distance of 2 amu; reject a CID if less than 10 peaks or if precursor mass less than 50 Da or greater than 10,000 Da; precursor mass tolerance for grouping, 1 Da; maximum number of cycles between groups, 10; minimum number of cycles per group, 1; default precursor charge states, 2+ and 3+. Searches were conducted using Mascot Search Engine version 2.1.04 (Matrix Science) with the following parameters: mass error tolerances of ±75 ppm for MS and ±0.6 Da for MS/MS; “Max missed cleavages” set to 2; variable modification, oxidation (Met); instrument set to ESI-Q-TOF.

Results were filtered using “Standard scoring.” “Maximum number of hits” was set to 200. “Significance threshold” was set at \( p < 0.05 \). For initial searching, “ions Score cut-off” was set at 0; however, peptide matches were later filtered to remove peptide matches with expect values above 0.05. To test the expected confidence levels of the analytical approach, false positive peptide identification rates under the matching criteria described above were estimated at 2.7% by searching against a randomized version of the sequence database used for real searches. More extensive protein modification options were used after initial searching was performed with putative modified peptide spectra manually interpreted to confirm modification.

Mascot peptide match reports thus produced were exported to comma-separated value text file format and parsed with a custom PHP script to collect and filter peptide match results, map peptide hits onto r-protein genes and r-protein families, produce customized tab-delimited text and HTML (HyperText Markup Language) reports quantitatively summarizing MS/MS evidence for the detection of specific r-protein genes and r-protein families, and automatically generate annotated MS/MS spectral diagrams in scalable vector graphics (.SVG) format for manual verification of peptide identifications in cases where a protein identification was supported by a single spectrum. The script included steps to filter out peptide matches with expect values above a threshold value of 0.05, select the highest scoring MS/MS match to each non-redundant peptide sequence (different charge or modification states on the same peptide sequence were regarded as the same peptide sequence), and use the tables of in silico predicted peptides described earlier under “In Silico Digestion Analysis” to automatically assign and count “total” and “gene-specific” peptide hits to r-protein genes and total peptide hits to r-protein families and sum the Mascot “ions Scores” for each of the resultant sets of unique top scoring hits to total and gene-specific peptides assigned to each r-protein and r-protein family. To avoid false positive claims of gene-specific r-protein identification based on ambiguous spectra, MS/MS spectra that gave significant matches to more than one peptide sequence were not allowed to contribute to gene-specific peptide counts and scores.

Supplemental Material—Size analyses of theoretical peptides derived from ribosomal proteins following cleavage by different enzymes are provided in supplemental Data 1. Generic format data files (.MGF) generated from the primary mass spectra based on the script explained above are provided as supplemental Data 2; these can be directly reanalyzed on line at Matrix Science by the reader. Detailed information of the mass spectra matching and individual peptide scores and sequence coverage information that form the basis of the identifications in Table I are provided in supplemental Data 3. Detailed information about the mass spectral matching and peptide scores for gene-specific matches to members of protein families among S, L, and P ribosomal proteins are in supplemental Data 4 together with annotated MS/MS spectra and MS/MS peak lists for gene-specific identifications and protein families supported by single peptide identifications and the protein families S29, L36a, S30, L39, and L29. Detailed information of the mass spectra matching and peptide scores for non-ribosomal proteins identified is supplied in supplemental Data 5. Detailed information on the spectra interpreted for different classes of post-translational modified peptides is presented in supplemental Data 6; annotated, centroided, and deisotoped spectra supporting these post-translational modification claims are provided in supplemental Data 7.

RESULTS

In Silico Digestion to Develop Strategies for Analysis of the Arabidopsis Ribosomal Proteome—To develop strategies to identify ribosomal proteins from the cytosolic ribosome with confidence based on peptide MS/MS, we undertook a wide survey of the predicted ribosomal protein sequences from Arabidopsis. The latest release of the Arabidopsis thaliana genome (TAIR6) contains 409 genes (excluding pseudogenes) annotated as encoding cytosolic, mitochondrial, or chloroplastic ribosomal proteins or ribosomal protein-like proteins. These include the protein sequences of nine genes that were not annotated at the time of the Barakat et al. (17) analysis of Arabidopsis ribosomal sequences that are new isoforms of S15, S25, P0, P1, L10a, L19, L24, and L41. The corresponding set of 409 predicted protein sequences were subjected to batch in silico digestion with either “trypsin,” “chymotrypsin,” or “pepsin” using an in-house modified version of the published script Proteogenest (21).

Processing of Proteogenest output using custom PHP scripts produced a table for each enzyme containing the entire list of peptide sequences theoretically produced and for each peptide the gene(s) and gene family(s) from which it could be derived and a list of those peptides that were predicted to come from only one of the 409 protein sequences (see supplemental Data 1 for lists of total and gene-specific peptide sequences and mz values for [M + 2H]⁺ ions). For example, in silico digestion using trypsin allowing one missed cleavage produced ~13,000 unique peptide sequences (Fig. 1). Approximately 9400 (72%) of these peptide sequences were predicted to come from only one of the 409 protein sequences (gene-specific peptides), whereas others could be derived from a number of different genes. Analysis of all these data showed that many low molecular mass peptides are produced by trypsinitization of ribosomal proteins due to the high arginine and lysine content of these relatively basic protein sequences.
Comparison with the data published by Chang et al. (19) and Giavalisco et al. (20) on the experimental identification of ribosomal proteins showed that six protein families were not represented in either of these earlier reports: the 40 S families S29 and S30 and the 60 S subunits L36a, L39, L29, and L41. These proteins are all small in size (< 12 kDa), highly basic in nature (pI 10.5–13.4; average, 11.7), and, from our in silico digestions, lead to especially small peptides for analysis when trypsinized.

From this analysis of both theoretical digestion and published data on the ribosome, a strategy was developed to maximize coverage of the ribosomal analysis. First, it would be necessary to obtain a large number of unique MS/MS spectra to provide coverage of the likely 80–249 ribosomal proteins expected in the cytosolic proteome. Given that 72% of tryptic peptides were likely to be gene-specific (Fig. 1), 500–1500 peptide identifications would be required to provide, on average, three to four specific peptides for each protein match depending on the number of specific proteins in the ribosome. Second, it would be important to use a low m/z cutoff for the selection of MS ions for MS/MS analysis; 200–900 was chosen on the basis of the abundance of peptides below 800 amu (which would be doubly charged MS ions below 400 m/z from electrospray ionization) especially in the families of proteins so far not identified in Arabidopsis samples. Third, it would be valuable to use a simple protein display technology such as 1D SDS-PAGE to ensure that proteins were not lost for analysis due to being too basic or not dissolving in IEF buffers for 2D gel analysis or being lightly stained and not analyzed due to lack of visible spots in 2D gel analysis of ribosomes. This was especially important because these technical reasons had been used to explain the lack of coverage of certain families in previous reports on the ribosome composition (19, 20). Fourth, given the similarity of peptide sequences within and between ribosomal protein families it would be necessary to undertake an extensive analysis of the gene specificity of any given peptide sequence to ensure that only real gene-specific matches were used as evidence of specific proteins.

Isolation, Display, Digestion, and MS Analysis of Ribosomal Proteins—The protein complement of ribosomes isolated from Arabidopsis cell culture was displayed by SDS-polyacrylamide gel electrophoresis. The gel separation of the successive steps in the purification of ribosomal proteins provides a view of the decreasing complexity of the protein sample and the refinement of the major bands of r-proteins. In Fig. 2, far right lane, the discrete banding pattern of the ribosomal samples is evident.

To assess whether our ribosomal preparation method produced intact 80 S ribosomes that could be dissociated into two distinct subunits by a buffer of high ionic strength, we compared the sucrose density gradient sedimentation profiles of ribosome samples suspended in 5 mM MgCl₂ ribosome buffer of either moderate (0.1 M KCl) or high ionic strength (0.5 M KCl). The sedimentation banding patterns of 80 S ribosomes and 60 and 40 S subunits was followed by UV absorb-
ance at 260 and 280 nm. This analysis revealed that at 0.1 M KCl, ribosomal subunits remained essentially completely associated into 80 S ribosomes, whereas at 0.5 M KCl, subunits were completely dissociated into 40 and 60 S subunits with virtually no 80 S band detected (Fig. 3). No other obvious UV-absorbing bands were detected.

For analysis of the proteins in the purified ribosomal samples, three lanes each containing 100 μg of ribosomal proteins were run in parallel, and the Coomassie-staining bands of the entire three-lane width were excised and digested in small pieces with trypsin in 12 separate wells of a 96-well plate. In total a set of 21 band regions were handled separately to provide 21 sets of peptides from 300 μg of ribosomal r-proteins (Fig. 4A). A further nine sets of bands for low molecular mass gel regions were cut from replicate gels and digested with either chymotrypsin or pepsin to increase the coverage of low molecular weight ribosomal proteins (Fig. 4B).

**Fig. 2.** Purification of 80 S cytosolic ribosomes from *A. thaliana* cell suspension culture. Cytosolic ribosomes were purified from whole cell homogenate by differential centrifugation and ultracentrifugation through two successive 1.5 M sucrose cushions (see “Experimental Procedures”). Protein samples of equivalent protein content were taken at each step in the purification and resolved by SDS-PAGE. The samples were loaded from left to right in order of increasing purity. The lanes represent (from left to right) filtered whole cell homogenate (“Whole Cell”), 1500 x g supernatant (“1500 x g”), 16,000 x g supernatant (“16000 x g”), 30,000 x g supernatant (“30000 x g”), crude ribosomal pellet from first ultracentrifugation through 1.5 M sucrose (“1st 1.5M Suc”), supernatant after clarification of crude ribosomal sample at 20,800 x g (“Clarified Pellet”), and pure ribosomal sample after second ultracentrifugation through 1.5 M sucrose and clarification at 20,800 x g (“2nd 1.5M Suc”).

**Fig. 3.** Dissociation of the *Arabidopsis* cytosolic ribosome with KCl. Sucrose gradient profile of intact *Arabidopsis* cytosolic ribosomes (0.1 M KCl or control) and dissociated cytosolic ribosomes (0.5 M KCl). Abs., absorbance.

**Fig. 4.** A, gel regions subjected to in-gel protease digestion and LC-MS/MS. Cytosolic ribosomal proteins were fractionated by one-dimensional SDS-PAGE, and 30 gel regions (indicated by numbers 1–30) were individually excised and subjected to in-gel protease digestion, and peptides were extracted and analyzed by LC-MS/MS (see “Experimental Procedures”). All gel regions were digested with trypsin (T), whereas certain regions were also digested separately with chymotrypsin (C) and pepsin (P) resulting in a total of 48 individual LC-MS/MS runs. B, Pro-Q staining of cytosolic ribosomal proteins. Ribosomal proteins were separated by SDS-PAGE and stained with Pro-Q phosphoprotein selective stain. Gel regions that are equivalent to those that yielded hits to L13, S6, P0, P1, P2, and P3 proteins (i.e. proteins for which phosphorylation sites were found by LC-MS/MS) are indicated.
4A). Each of the 48 (21 + 9 + 9 + 9) samples were separately analyzed by gradient elutions of peptides from a C18 column into an Applied Biosystems QStar Pulsar Q-TOF mass spectrometer at 8 \( \mu \text{mL/min} \) under the parameters outlined under “Experimental Procedures.”

**Analysis of Raw LC-ESI-MS/MS Data and Mapping of MS/MS Peptide Hit Data onto the Theoretical Arabidopsis Ribosomal Proteome**—The 48 raw data files generated by LC-ESI-MS/MS analysis of the in-gel digested 1D SDS-PAGE protein bands were divided into three groups based on the protease used, and each group of data files was used to generate a .MGF peak list file for analysis by Mascot as outlined under “Experimental Procedures” (supplemental Data 2). Peptide hits reported by Mascot were filtered to remove low confidence matches (expect value >0.05) and mapped, on the basis of peptide sequence, onto their associated r-protein genes, r-protein families, and gene specificity information using custom PHP scripts built around database tables of in silico predicted tryptic, chymotryptic, or peptic peptide sequences. These in silico digestion tables were built by using the data from Proteogeest (that were analyzed in Fig. 1). These data were automatically processed using the PHP script to generate peptide mapping report tables containing columns for peptide sequence, AGI, ribosomal protein family, and peptide MS/MS match data and a column indicating whether or not the peptide sequence was gene-specific (supplemental Data 3). In this way, each mapping between a peptide match and an r-protein gene was represented by a separate row. Hence gene-specific peptide matches only occurred in one row of the script-generated table, and nonspecific peptide matches occurred in multiple rows (one for each gene from which they could be derived).

Peptide hit information mapped onto genetic loci was then processed by the PHP script to automatically count the number of unique peptide sequences detected per gene in the ribosomal gene set, select the highest scoring MS/MS spectrum for each unique detected peptide sequence, and add together the peptide scores for all the best unique peptide hits for each r-protein family (compiled in Table I). The same approach was used to calculate the number of gene-specific peptides found and the sums of the peptide scores for non-redundant gene-specific peptides found for each gene (see supplemental Data 4). Systematic analysis of the sequences of in silico predicted gene-specific peptides revealed that some of these peptides differed from each other only through isobaric Leu/Ile or Gln/Lys substitutions that would not be resolvable in our MS/MS analysis and would result in these spectra giving rise to search engine matches against more than one gene-specific peptide sequence. This finding highlighted the importance of using only unambiguously matched MS/MS spectra to support gene-specific protein identification. Hence spectra that matched significantly to more than one peptide sequence were not allowed to contribute gene-specific peptide identification counts or scores. Fig. 5 provides a flow diagram overview of the whole data analysis procedure.

**Protein Family Identifications by MS/MS—**Our analysis mapped 1446 MS/MS spectra (1413 tryptic, 26 chymotryptic, and seven peptic) as positive matches to 795 (776 tryptic, 13 chymotryptic, and six peptic) non-redundant peptides from ribosomal proteins. We also detected a high scoring match (Mascot Ions Score = 61) to a tryptic peptide derived from the r-protein family L29 in a separate analysis of r-proteins enriched for phosphopeptides. Together these data confirmed the presence of peptides from 79 of the 80 protein families believed to make up the ribosome (Table I). In comparison, Chang et al. (19) and Giavalisco et al. (20) claimed data for 75 and 61 protein families, respectively. Only one protein family thus remains undetected in experimental ribosome preparations in Arabidopsis, L41. The four L41 proteins are extremely small (3–4 kDa) and extremely basic (pI 13.5).

The newly identified proteins from our analysis were from the 40 S S29 and S30 families and the 60 S L29, L36a, and L39 families. These small proteins were identified through a combination of spectral matches from trypsin, chymotrypsin, and pepsin digestions of low molecular mass protein bands (Table II). Inspection of the matched peptides showed that 64% of the matches were to MS ions smaller than 400 \( m/z \), justifying our strategy to include low \( m/z \) MS2 + and MS3 + ions in our analysis. All the presented spectra for these newly identified proteins were family-specific, but none were gene-specific. In the cases of the three members of S29, two members from L36a, and three members from S30, there are no specific peptides derived from theoretical digestes. Among the three members of L39, a very small number of gene-specific peptides can theoretically be produced, but none were found in this analysis.

**Gene-specific Identifications by MS/MS in the Ribosome—**One of the most taxing issues in analysis of the ribosomal dataset, but of high interest to researchers, is the question of heterogeneity in the proteins that make up the ribosome derived from the families of genes encoding proteins for each specific subunit. Table I shows that in every case in Arabidopsis ribosomal subunits are encoded by at least two and in some cases up to seven different genes with an average of three genes per subunit across the whole protein complex. Importantly our data for peptide identification are based on a large MS/MS dataset of over 1400 matching spectra, and they contain 267 unambiguous gene-specific peptide identifications (see supplemental Data 3). The claims of earlier studies of the ribosome were made on the basis of much smaller datasets, and although many specific member proteins were claimed in these publications, the strength of the evidence in any particular case was substantially weaker than that obtained in this study. Chang et al. (19) based their claims on \( \sim 850 \) peptide identifications mainly from MS ions in MALDI-TOF with only 172 claims based on MS/MS spectra of which only 66 were gene-specific MS/MS matches. The Giavalisco et al. (20) study was entirely based on an unknown number of
TABLE I

80 S ribosomal protein families of *A. thaliana*: detection by SDS-PAGE/LC-MS/MS

80 S cytosolic ribosomes were isolated from *A. thaliana* cell suspension culture, and ribosomal proteins were separated by SDS-PAGE. Gel regions stained positively with Coomassie stain were excised from the gel and in-gel digested with protease. Peptides were extracted, analyzed by LC-MS/MS, and identified by Mascot searching of MS/MS spectra. Peptide matches reported by Mascot were mapped onto bioinformatically predicted 80 S ribosomal protein families via relational tables of in silico predicted peptide sequences. For each family, the number of unique peptide matches (No. pep; if non-tryptic peptides were included in this peptide count, the protease(s) used to produce the peptides are shown by a "t" for trypsin, "c" for chymotrypsin, and/or "p" for pepsin) and the sum total of the associated Mascot peptide scores (including only the top scoring hit to each unique peptide; Total peptide score) were calculated. Gel regions that yielded peptides derived from each ribosomal protein family are shown with gel region numbers corresponding to those in Fig. 4. Ribosomal protein family nomenclature as described in Barakat et al. (17) is used. Y, claimed by publication; ND, not claimed by publication.

| Family | AGI numbers | No. members (TAIR6) | No. pep | Total peptide score | Gel region numbers | Chang et al. (19) | Giavalisco et al. (20) |
|--------|-------------|---------------------|---------|---------------------|--------------------|-------------------|-----------------------|
| Sa     | At1g72370, At3g04770 | 2 | 9 | 494 | 2–7 | Y | Y |
| S2     | At1g58684," At1g58983, At1g58380, At1g59359, At2g41840, At3g57490 | 6 | 18 | 920 | 4–6 | Y | Y |
| S3     | At2g31610, At3g53370, At5g35530 | 3 | 20 | 1007 | 2–7 | Y | Y |
| S3a    | At3g04840, At4g34670 | 2 | 21 | 1070 | 4–6 | Y | Y |
| S4     | At2g17360, At5g07090, At5g58420 | 3 | 16 | 702 | 6–7 | Y | Y |
| S5     | At2g32720, At3g11940 | 2 | 6 | 389 | 10–11 | Y | Y |
| S6     | At4g31700, At5g10360 | 2 | 19 | 1041 | 5–7 | — | — |
| S7     | At1g48830, At3g02560, At5g16130 | 3 | 12 | 566 | 9–12 | Y | Y |
| S8     | At5g20290, At5g59240 | 2 | 11 | 662 | 7 | Y | Y |
| S9     | At1g51750, At5g15200, At5g39850 | 3 | 13 | 566 | 10–12 | Y | Y |
| S10    | At4g25740, At4g15250, At5g52650 | 3 | 7 | 337 | 14–21 | Y | Y |
| S11    | At3g49890, At4g30800, At5g23740 | 3 | 8 | 400 | 14–16 | Y | Y |
| S12    | At1g15930, At4g23060 | 2 | 3 | 191 | 14–17 | Y | Y |
| S13    | At3g60770, At4g00100 | 2 | 8 | 400 | 15–17 | Y | Y |
| S14    | At2g36160, At3g11510, At3g52580 | 3 | 5 | 318 | 17–19 | Y | Y |
| S15    | At1g33850," At5g63070, At1g04270, At5g09500, At5g09510, At5g43640 | 7 | 4 | 202 | 15–16 | Y | Y |
| S15a   | At1g07770, At2g19720, At2g39590, At3g46040, At4g29430, At5g59850 | 6 | 11 | 495 | 21 | Y | Y |
| S16    | At2g09990, At3g04230, At5g18380 | 3 | 10 | 462 | 15–17 | Y | Y |
| S17    | At2g04390, At2g05220, At3g10610, At5g04800 | 4 | 5 | 222 | 16–19 | Y | Y |
| S18    | At1g22780, At1g34030, At4g09880 | 3 | 9 | 506 | 17–19 | Y | Y |
| S19    | At3g02080, At5g15250, At5g61170 | 3 | 15 | 719 | 19–21 | Y | Y |
| S20    | At3g45030, At3g47370, At5g62300 | 3 | 5 | 248 | 18–19 | Y | Y |
| S21    | At3g53890, At5g27770 | 2 | 2 | 239 | 24 | — | — |
| S22    | At3g09680, At5g02960 | 2 | 8 | 389 | 16–18 | Y | Y |
| S23    | At3g04920, At5g28060 | 2 | 9 | 443 | 17–19 | Y | Y |
| S24    | At4g34555," At2g16360, At2g21580, At4g39220 | 4 | 4 | 236 | 20–29 | Y | Y |
| S26    | At2g40510, At2g40590, At3g56340 | 3 | 1b | 57 | 18 | Y | Y |
| S27    | At2g45710, At3g61110, At5g47930 | 3 | 1b | 57 | 24,27 | Y | Y |
| S27a   | At1g23410, At2g47710, At3g62250 | 3 | 2 | 78 | 21–24 | Y | — |
| S28    | At3g10090, At5g03850, At5g64140 | 3 | 7 | 359 | 27–30 | Y | Y |
| P0     | At1g25260," At2g40010, At3g09200, At3g11250 | 4 | 18 | 885 | 3–8 | Y | Y |
| P1     | At3g49460," At5g24510," At1g01100, At4g00810, At5g47700 | 5 | 6 | 285 | 19–21 | Y | — |
| P2     | At2g27710, At2g27720, At3g28500, At3g44590, At5g40040 | 5 | 13 | 831 | 19–21 | Y | Y |
| P3     | At3g25890, At5g57290 | 2 | 2 | 105 | 18–19 | — | — |
| L3     | At1g43170, At1g61580 | 2 | 23 | 1136 | 1 | Y | Y |
| L4     | At3g09630, At5g02870 | 2 | 27 | 1486 | 1 | Y | Y |
| L5     | At3g25520, At5g39740 | 2 | 25 | 1148 | 3–4 | Y | Y |
| L6     | At1g18540, At1g74050, At1g74060 | 3 | 14 | 718 | 5–7 | Y | Y |
| L7     | At1g80750, At2g01250, At2g44120, At3g13580 | 4 | 30 | 1520 | 8 | Y | Y |
| L7a    | At2g47610, At3g62870 | 2 | 18 | 904 | 6–7 | Y | Y |
| L8     | At2g18020, At3g51190, At4g36130 | 3 | 17 | 787 | 5–7 | Y | Y |
| L9     | At1g33120, At1g33140, At4g10450 | 3 | 12 | 752 | 9–11 | Y | Y |
| L10    | At1g14320, At2g6910, At1g66580," | 3 | 11 | 486 | 9–11 | Y | Y |
| L10a   | At1g08360, At2g27530, At2g27535," At5g22440 | 4 | 18 | 955 | 8–10 | Y | Y |
| L11    | At2g42740, At3g58700, At4g18730, At5g45775 | 4 | 11 | 583 | 13–14 | Y | Y |
| L12    | At2g37190, At3g53430, At5g60670 | 3 | 7 | 452 | 13–16 | Y | Y |
MS ions derived by MALDI-TOF. Claiming gene-specific identifications within conserved protein families relies on the high confidence matching of data derived from often only a small number of individual peptides. In this context, MS/MS spectra are much more powerful discrimination tools than the matching of MS ion masses from MALDI-TOF spectra. We used a high confidence cutoff for claiming gene-specific identification. For genes with only a single detected gene-specific peptide, the peptide had to have a peptide score equal to or greater than 40 and an expect value below 0.05, and the MS/MS spectrum had to be manually verified. The annotated MS/MS spectra for these cases are supplied in supplemental Data 4. For genes with more than one detected gene-specific peptide, all peptides were required to have an expect value below 0.05 and an average peptide score of at least 30.

Our analysis of gene-specific peptides in silico revealed that within 10 gene families no theoretical gene-specific peptides exist: S18, S29, S30, L11, L21, L23, L36a, L38, L40, and L41. So for these proteins, no amount of spectral data will allow verification of heterogeneity of the protein population. For a further eight families the level of sequence identity is so high that few sequence-specific peptides exist, and hence it is rather unlikely they could be found (S4, S13, S17, S26, S27, L15, L35a, L37, and L29). But for the remaining 63 families enough theoretical gene-specific peptides exist to suggest that a high coverage MS/MS dataset could be able to define with clarity whether gene-specific data exist to justify the specific presence of gene products from different chromosomal loci.

Several errors in the assignments made by Chang et al. (19)
were apparent from our further analysis. In the family P3, Chang et al. (19) claimed specific identification of At4g25890, but the peptides they claimed as matches were actually not from this gene product but from the other member of the family, At5g57290. For this protein we also made the specific identification of a locus-specific peptide. Chang et al. (19) claimed a specific gene match to two L19 proteins from a family of four members, At1g02780 and At4g02230. However, on closer inspection the peptide claimed for At1g02780 (GPGGDVAPVAAPAPAATPAATPAPTAAVPK) is not present in the sequence of this protein. Our data strongly point to gene-specific data only for At4g02230. In the two-member L15 family, Chang et al. (19) claimed gene-specific identification of At4g16720 based on a single peptide, but on closer inspection this peptide is indistinguishable from the peptide from the other member of this family as it is a Leu/Ile difference that cannot be determined by MS. The two members of L15 are indistinguishable in our analysis; there is only one theoretical peptide that could distinguish this pair, but it remained undetected by MS even though we had MS/MS data on 12 peptides to this family. A similar story is evident in the S5 family of two members. Chang et al. (19) claimed a specific peptide to
Ribosomal proteins were separated by SDS-PAGE, and gel regions predicted to possibly contain the low molecular weight, previously undetected ribosomal protein families S29, L36a, L39, L29, and L41 families were in-gel digested in parallel with trypsin (T), chymotrypsin (C), and pepsin (P). Extracted peptides were analyzed by LC-MS/MS and identified by Mascot searching of MS/MS spectra. Identified peptides matching to each family are shown. Annotated MS/MS spectra for all tryptic peptides in this table are provided in supplemental Data 4E. obs, observed.

| Ribosomal protein family | Enzyme | m/z   | \(M_1\) (obs) | \(M^*\) | Delta | Miss | Mascot Ions Score | Expect value |
|--------------------------|--------|-------|----------------|--------|-------|------|------------------|-------------|
| S29 (At3g43980, At3g44010, At4g33865) | KYGPGSR | T     | 382.69         | 763.37 | 2     | −0.02 | 1                | 36          | 0.019       |
|                          | GHSNVWNSHPKKYGGPGSR | P     | 530.99         | 2119.94 | 4     | −0.13 | 1                | 32          | 0.015       |
| L36a (At3g23390, At4g14320) | DSSLAAQK | T     | 395.19         | 788.37 | 2     | −0.04 | 0                | 38          | 0.0088      |
|                          | GKSLSLAAQK | T     | 487.75         | 973.48 | 2     | −0.04 | 0                | 56          | 0.00014     |
|                          | HFSQRPIK | T     | 338.18         | 1011.51 | 3     | −0.05 | 0                | 32          | 0.029       |
|                          | VNLPK   | T     | 285.67         | 568.32 | 2     | −0.04 | 0                | 24          | 0.039       |
| S30 (At2g19750, At4g29390, At5g56670) | FVTAVGFGK | T     | 512.78         | 1023.55 | 2     | −0.03 | 0                | 61          | 3.20e−05    |
| L29 (At3g06680, At3g06700) | NHTAHNQSAK | T     | 554.27         | 1106.52 | 2     | 0    | 0                | 61          | 0.0001      |
| L39 (At2g23210, At3g02190, At4g31985) | TDNTR   | T     | 360.18         | 718.35 | 2     | −0.01 | 0                | 41          | 0.0064      |
|                          | PSHKSF  | P     | 351.66         | 701.31 | 2     | −0.04 | 0                | 41          | 0.0039      |
|                          | PSHKSF  | C     | 351.67         | 701.31 | 2     | −0.03 | 0                | 39          | 0.0097      |

* This peptide was identified in a tryptic digest of ribosomal proteins previously enriched for phosphopeptides.
family that evolved to direct specific members to be part of the mitochondrial ribosome and retain others in the cytosolic ribosome (25). Sequence comparisons showed two types of S15a sequences, type I (At1g07770, At2g35590, At3g46040, and At5g59850) and type II (At2g19720 and At4g29430). The type II proteins were shown to be translocated to the mitochondria, whereas the type I proteins remained in the cytosol (25). Chang et al. (19) claimed peptide-specific evidence for both type I and type II proteins in the cytosolic ribosome. In our analysis we were unable to find any evidence for peptides even redundantly matching the type I proteins (At2g19720 and At4g29430) in our cytosolic ribosome samples; all 11 peptides found were specific for type I protein members (Tables I and III). In parallel, we partially enriched ribosomes from mitochondrial samples and analyzed the analogous band on 1D gels that matched band 21 where S15a peptides were found. Here we found strong evidence of type II sequences, one peptide that redundantly matched to At2g19720 and At4g29430 and two peptides specific for type II protein members (At4g29430) in our cytosolic ribosome samples; thus we propose that the presence of type II peptides in the study of Chang et al. (19) is likely due to mitochondrial ribosome contamination of their samples rather than the real presence of type II proteins in the cytosolic ribosome.

Non-ribosomal Proteins Bound to Ribosomes—A set of five proteins that do not represent members of the 80 established families of ribosomal proteins were identified in this study (supplemental Data 5). These were two guanine nucleotide-binding family proteins (At1g16800 and At1g48630), several of which have been reported previously in ribosomal extracts (19, 20); a eukaryotic translation initiation factor (At3g56620) that is likely bound to active ribosomes; and a ferritin-like protein (At3g56090).

Covalent Modifications of Arabidopsis Cytosolic Ribosomal Proteins—Covalent protein modifications such as acetylation, methylation, and phosphorylation have emerged as potentially being important factors contributing to ribosomal heterogeneity in both eukaryotes (4, 19, 26, 27) and prokaryotes (28, 29). In the ribosomes of higher plants, studying the role of covalent r-protein modification in ribosomal heterogeneity is complicated by the frequent expression of relatively high (compared with other classes of organism) numbers of different, yet often highly conserved isoforms of each ribosomal subunit. This situation adds another layer of complexity to the MS-based analysis of covalent r-protein modification in higher plants because very often the mass difference between two closely related peptides from two closely related r-protein isoforms is exactly or almost equal to the mass difference expected from a covalent modification. Hence in many cases, parent ion mass, as obtained by simple MS, cannot distinguish the modified peptide of one family member from a closely related, unmodified peptide derived from another family member. As we demonstrate here, this problem can be largely overcome through the use of LC-ESI-Q-TOF-MS/MS, which allows isobaric peptides of very similar structure to be distinguished on the basis of their MS/MS fragmentation patterns.

To identify sites of covalent modification on ribosomal proteins, candidate MS/MS spectra potentially corresponding to covalently modified peptides were identified in the complete MS/MS dataset through a series of Mascot searches considering various combinations of acetylation; formylation; mono-, di-, and trimethylation; and phosphorylation as variable modifications. MS/MS spectra for which the top Mascot match (across all the Mascot searches) was a covalently modified peptide with a peptide score greater than 30 were manually inspected to verify the identity of the peptide and determine the exact position(s) of the modified residue(s). This manual inspection process allowed near isobaric modifications such as trimethylation/acyetylation and dimethylation/formylation (which each differ by 0.0364 amu) to be distinguished on the basis that for low m/z fragment ions carrying the modification...
the mass accuracy of the TOF analyzer was sufficient to resolve the two near isobaric modifications. This approach revealed strong MS/MS evidence for 30 unique covalently modified peptides corresponding to a total of 41 covalent modification events. Detected modification types included removal of N-terminal methionine (15 cases), N-terminal acetylation (12 cases), N-terminal dimethylation (one case), N-methylation of lysine side chains (three cases), and phosphorylation (nine cases). Of the 30 covalently modified peptides detected, 13 (43%) could be assigned to a specific genetic locus, whereas the remaining 57% could only be redundantly assigned to two to four members of a particular r-protein family due to sequence conservation between family members. Overall the results of this analysis (Table IV and supplemental Data 6) suggest extensive covalent modification of the ribosome with at least 23 (29%) of the 80 ribosomal protein families exhibiting at least one covalent modification.

**N-terminal Modifications: Initiator Methionine Removal and Acetylation** — Under the stringent MS/MS matching criteria, a total of 23 unique N-terminal peptides representing 18 r-protein families were detected (Table IV). Of these 23 N-terminal peptides, 20 (87%) were modified by either initiator methionine removal, N-terminal acetylation, or both. The remaining three (13%) peptides, specific to the r-proteins S10C (At5g52650), P2A (At2g27710), and P2B (At2g27720), retained an unmodified N-terminal methionine residue. Five (22%) of the N-terminal peptides detected, representing r-protein families L18, L21, L32, L36, and S27, had parent ion masses and CID fragmentation patterns consistent with initiator methionine removal without further modification. In these peptides, the N-terminal residue was either glycine, proline, alanine, or valine, consistent with previous observations that initiator methionine removal occurs preferentially when the residue immediately adjacent to the initiator methionine has a relatively small radius of gyration (30). Ten (43%) of the detected N-terminal peptides, representing r-protein families L10a, L28, S15, S18, S20, S3, S5, and Sa, were detected as ions with parent masses and CID fragmentation patterns consistent with initiator methionine removal and acetylation or near the N-terminal residue. In these peptides the N-terminal residue was invariably alanine (80% of cases) or serine (20% of cases). Two peptides, which were specific to the r-proteins S21B (At3g53890) and S7A (At1g48830), were detected with MS/MS signals within 23 ppm of the expected m/z of the acetylated b2 ion were observed. Interestingly in these three peptides (one representing the r-protein family L10a and two representing L28) the residue in the second position was capable of being acetylated, and hence acetylation of this second residue cannot be ruled out.

**Unique Case: N-terminal Proline Dimethylation of Cytosolic R-protein L12** — Several Mascot searches considering various combinations and types of acetylation; formylation; and mono-, di-, and trimethylation as variable modifications matched several spectra to a modified form of the N-terminal peptide PPKLDPSQIVDYVR, which matches redundantly to all three members of the L12 r-protein family (At2g37190, At3g53430, and At5g60670) (Table IV). The actual modifications reported by Mascot varied from acetyl (N-terminal) + formyl (Lys) to dimethyl (N-terminal) + trimethyl (Lys) depending on which combination of variable modifications was included in the search. The high peptide scores (often >50), the lack of any other significantly scoring assignments to this spectrum, and the fact that the spectra matched by Mascot to this modified peptide were detected in SDS-PAGE gel bands that gave rise to other L12-matching spectra suggested that this was a real peptide match, although the exact nature of the modification(s) was not clear from Mascot results except that they introduced a total nominal mass increase of 70 Da. Upon manual inspection of the most highly scoring spectrum, several lines of evidence suggested that the peptide was in fact dimethylated at the N-terminal proline residue and N-trimethylated at the e-amino group of the Lys-3 residue. First, in the raw spectrum, signals corresponding to singly charged, unmodified y1–y6 and y10 ions were clearly present suggesting that the modification(s) adding 70 Da to the parent ion mass was located within the first five residues. This conclusion was supported by strongly doubly charged b3–b5, b8, and b11–b14 series ions that were all ~35 m/z (~70 Da) higher than expected for the unmodified peptide. This confirmed that the modifications were localized to the first three residues, Pro-Pro-Lys-. Although a number of combinations of modifications that could occur at these residues could explain a nominal mass increase of 70 Da (e.g. formyl + acetyl, acetyl + dimethyl, formyl + trimethyl, or dimethyl + trimethyl), these modifications could be divided into three different accurate mass groups resolvable on the TOF mass analyzer: +C$_3$H$_6$O$_2$N$_2$ (+70.0055 Da), +C$_4$H$_8$O$_2$ (+70.0419 Da), and +C$_5$H$_{10}$O$_2$ (+70.0783 Da). Calculation of the observed accurate mass shifts of the putatively modified doubly charged b3, b4, and b5 ions relative to the theoretical unmodified masses revealed that the observed signals for these ions in the raw spectrum corresponded to respective mass shifts of 70.0688, 70.0692, and 70.0782 Da. That is, the observed masses of the modified b3, b4, and b5 ions were, respectively, 44, 36, and 58 ppm closer to the masses expected for a +C$_5$H$_{10}$O$_2$ modification (e.g. N-terminal dimethylation + lysine N-trimethylation) than to the next most isobaric modification, +C$_4$H$_8$O$_2$ (e.g. N-terminal

---

Cytosolic Ribosome of Arabidopsis

---

**Molecular & Cellular Proteomics 7.2 359**
**TABLE IV**

Covalent modifications of Arabidopsis 80 S ribosomal proteins as identified by LC-MS/MS

MS/MS spectra derived from covalently modified peptides carrying internal or N-terminal modifications were identified in raw LC-MS/MS data files through a combination of Mascot searching and manual interpretation of MS/MS spectra as outlined under “Results.” For peptides carrying internal modifications, the modified residue, as determined by manual spectral interpretation, is underlined. M(ox), oxidized methionine.

| Modification         | RP family | Gene product(s)                  | Peptide sequence | Mass error (Da) | Mascot Ions Score | Expect value | Diagnostic MS/MS fragments detected |
|----------------------|-----------|----------------------------------|-----------------|-----------------|-------------------|--------------|--------------------------------------|
| **Internal modifications** |           |                                  |                 |                 |                   |              |                                      |
| Trimethylation (lysine) | L10a      | At1g08360                        | M(ox)GLENM(ox)DVESLK | -0.05           | 59                | 0.00069      | y1–y12, b2–b7, b9; accurate mass shift (y2, 1+): 42.0462 Da (Δ = -0.0008 Da) |
|                      | L10a      | At2g27530                        | M(ox)GLSNM(ox)DVESLK | -0.05           | 58                | 0.001        | y1–y12, b2–b8, b11; accurate mass shift (y3, 2+): 42.0328 Da (Δ = -0.0142 Da) |
|                      | L12       | At2g37190, At3g53430, At5g60670  | Dimethyl-PPKLDPSQIVDVYVR | -0.09           | 62                | 0.0013       | y1, y4, y10, b3, b5, b7–b14; accurate mass shift (b3, 2+): 70.0656 Da (Δ = -0.0127 Da) |
| Monomethylation (lysine) | L36a      | At3g23390, At4g14320             | QSGYGGGQTPVYVR  | -0.07           | 70                | 6.80e–05     | y1–y3, y5–y7, y9–y13, b2–b8, b13; accurate mass shift (y6, 2+): 13.98 Da (Δ = -0.0353 Da) |
| Phosphorylation       | S6        | At4g31700                        | SRLSSAAKPSVTA   | -0.04           | 49                | 0.0018       | Intact: b2–b3, b5–b9, b14; neutral loss of 98 Da: parent, b4, b5 |
|                      | S6        | At5g10360                        | SRLSSAPAKPVA    | -0.05           | 33                | 0.069        | Intact: b2–b4, b6, b8–b9, b11–b12; neutral loss of 98 Da: parent, b6, b8, b10 |
|                      | P0        | At3g11250                        | KEESDEEDYEGGFGLFDEE | -0.06           | 27                | 0.1          | Intact: b2–b3, b5, b8, y1–y5, y7, y10; neutral loss of 98 Da: b4, b5, b7, b8, b10 |
|                      | P1        | At1g01100, At4g00810, At5g47700 | KDEPAEESDGLGFGLFD | -0.05           | 91                | 8.30e–08     | Intact: b2–b17, y2–y6, y10–y12, y16; neutral loss of 98 Da: b9, b10, b12, b14, b16 |
|                      | P1        | At1g01100, At4g00810, At5g47700 | DEPAEESDGLGFGLFD | -0.06           | 28                | 0.12         | Intact: b2–b10, b12, y1–y8, y14; neutral loss of 98 Da: b11, b12, b14 |
|                      | P2        | At2g27720, At2g27710, At3g44590 | EESDDDM(Ox)GFSLE | -0.04           | 25                | 0.084        | Intact: b1–b2, b6–b9, b12, y1–y2, y4–y10; neutral loss of 98 Da: b4, b6–b10 |
|                      | P3        | At4g25890, At5g57290             | KEESDEEDYEGGFGLFDEE | -0.05           | 38                | 0.011        | Intact: b2–b3, b5–b8, b10–b13, b15, y2–y3, y5–y6, y8, y10, y12; neutral loss of 98 Da: b4–b10 |
|                      | L13       | At3g49010                        | AGDSPEELANATQVQGDYLPVR | -0.08           | 62                | 0.00012      | Intact: b3, b6, b8, b9–b12, b14, b18, y1–y12, y16; neutral loss of 98 Da: b4, b5, b8, b9 |
| **N-terminal modifications** |           |                                  |                 |                 |                   |              |                                      |
| Initiator methionine removal | L18      | At3g05590, At5g27850             | GIDLJAGK        | -0.05           | 53                | 0.00065      | b2–b4, y1–y8 |
|                      | L21       | At1g09590, At1g09690, At1g57660, At1g57860 | PAGHGVR      | -0.03           | 49                | 0.0015       | b2–b4, b6, y1–y6 |
| Modification | RP family<sup>a</sup> | Gene product(s) | Peptide sequence<sup>b</sup> | Mass error | Mascot Ions Score | Expect value | Diagnostic MS/MS fragments detected |
|-------------|---------------------|-----------------|----------------------------|------------|-----------------|-------------|----------------------------------|
| L32         | At4g18100, At5g46430 | AVPLLTK         | -0.04                      | 33         | 0.031           | b2–b4, y1–y5 |
| L36         | At5g02450           | VATGLFVGGLNK    | -0.05                      | 77         | 3.30e–06        | b2–b7, b11, y1–y10 |
| S27         | At2g45710, At3g61110 | VLQNDIILNPPAELEKR | -0.03                     | 57         | 0.00023        | b2–b8, y1–y16 |
| L10a        | At1g08360, At2g27530, At5g22440 | Acetyl-SKLQSEAVR | -0.03                     | 47         | 0.0034          | b1–b3, y1, y3–y8; accurate mass shift (b2): 42.0107 Da (Δ 0.0001 Da) |
| L36         | At5g02450           | VATGLFVGGLNK    | -0.05                      | 77         | 3.30e–06        | b2–b7, b11, y1–y10 |
| S27         | At2g45710, At3g61110 | VLQNDIILNPPAELEKR | -0.03                     | 57         | 0.00023        | b2–b8, y1–y16 |
| L28         | At2g19730, At4g29410 | Acetyl-ATVPQLIWENK | -0.07                     | 36         | 0.031           | b2, b3, b7, b10, y1–y7, y9–y10; accurate mass shift (b2): 42.0069 Da (Δ −0.0037 Da) |
| L28         | At4g29410           | Acetyl-ATVPQLIWENKR | -0.06                     | 51         | 0.00094        | b2, b3, b5, y2–y3, y5–y8, y11; accurate mass shift (b2): 42.0089 Da (Δ −0.0017 Da) |
| S15         | At1g04270, At5g09510 | Acetyl-ADVEPEVAAGVPK | -0.06                     | 67         | 2.20e–05        | b1–b4, b6–b8, b10, y1–y2, y4–y11; accurate mass shift (b2): 42.0087 Da (Δ −0.0019 Da) |
| S18         | At1g22780, At1g34030, At4g09800 | Acetyl-SLVANEEFQHILR | -0.08                     | 88         | 1.70e–07        | b1–b8, b11, y1–y12; accurate mass shift (b2): 42.0066 Da (Δ −0.004 Da) |
| S20         | At3g47370           | Acetyl-AYPEMKT | -0.05                     | 45         | 0.0028          | b1–b3, y1–y8; accurate mass shift (b2): 42.0023 Da (Δ −0.0083 Da) |
| S20         | At3g45030, At5g62300 | Acetyl-ATAYQPMKPGK | -0.06                     | 48         | 0.0016          | b1–b4, b6–b7, y1–y10; accurate mass shift (b2): 42.0059 Da (Δ −0.0047 Da) |
| S3          | At2g31610, At5g35530 | Acetyl-ATQISKK  | -0.03                     | 37         | 0.036           | b3, b5, y2–y6; accurate mass shift (b3): 42.0084 Da (Δ −0.0022 Da) |
| S5          | At2g37270, At3g11940 | Acetyl-AAASAEIAEQQLTNEVK | -0.09                     | 73         | 5.40e–06        | b2–b10, y1–y13; accurate mass shift (b2): 42.0065 Da (Δ −0.0041 Da) |
| S5          | At3g11940           | Acetyl-ATAADVDAEQQLTNEVK | -0.08                     | 123        | 4.40e–11        | b2–b13, y1–y13, y15; accurate mass shift (b2): 42.0049 Da (Δ −0.0057 Da) |
| S6          | At1g72370           | Acetyl-ATNGSASSAQLSQK 0.94<sup>c</sup> | 63 | 0.00046 | b1–b4, b7, b10, y1–y11, y13; accurate mass shift (b1): 42.0076 Da (Δ −0.003 Da) |
| Acetylation of N-terminal methionine | S21 | At3g53890 | Acetyl-M(ox)ENDAGQVTELIPR | -0.04 | 135 | 2.50e–12 | b1, b2, b4–b8, b11, y1–y14; accurate mass shift (b1): 42.0143 Da (Δ 0.0037 Da) |
| S7          | At1g48830           | Acetyl-M(ox)FSAQHK | -0.04                     | 37         | 0.0088          | b1–b4, y1–y6; accurate mass shift (b1): 42.0113 Da (Δ 0.0007 Da) |
| Dimethylation of N-terminal proline | L12 | At2g37190, At3g53430, At5g60670 | Dimethyl-PPKLDPQSIVDVYVR | -0.09 | 62 | 0.0013 | y1, y4, y10, b3, b5, b7–b14; accurate mass shift (b3, 2+): 70.0656 Da (Δ −0.0127 Da) |

<sup>a</sup> Naming of ribosomal protein families uses nomenclature of Barakat et al. (17).

<sup>b</sup> See Supplemental Data 6 for MS/MS match details including notes on spectral interpretation.

<sup>c</sup> The large delta mass on this peptide was due to selection of the second isotopic peak for MS/MS by the instrument during information-dependent acquisition.
acetylation + lysine N-dimethylation). This observation strongly suggested that the modifications responsible for the 70-Da mass increase within the first three N-terminal residues of L12 were caused entirely by the addition of alkyl groups, most probably five methyl groups. Given that the proline N terminus is theoretically capable of accommodating up to two methyl groups and the \( \text{NH}_2 \) -amino group of lysine is capable of accommodating up to three, we propose that the most likely explanation for the observed data is that the N-terminal region of the Arabidopsis cytosolic r-protein L12 is \( \text{N}^\text{Me} \), \( \text{N}^\text{Me} \), \( \text{N}^\text{Me} \)-trimethylated at the \( \text{NH}_2 \)-amino group and \( \text{N}^\text{Me} \), \( \text{N}^\text{Me} \), \( \text{N}^\text{Me} \), \( \text{N}^\text{Me} \)-trimethylated at the \( \text{NH}_2 \)-amino group of the Lys-3 residue (see annotated spectra in Fig. 6B).

\[ \text{N-Methylation of Lysine—Mascot searches considering mono-, di-, and trimethylation of lysine as variable modifications matched 27 MS/MS spectra to a total of 20 unique lysine-methylated r-protein-derived peptides with a parent ion delta mass <0.1 Da and peptide score >30. However, on closer inspection it was found that the spectra matched to nine of the 20 peptides actually matched more strongly with unmodified, isobaric peptides of similar sequence from closely related family members. Manual interpretation of the highest scoring spectra for the 11 remaining peptides revealed strong \( \text{y} \) series evidence for lysine methylation in three peptides (Table IV). Two of these manually confirmed peptide matches, MGLENMDVESLKMe3K \( (M_r = 1566.76) \) and MGLSNMDVEALKMe3K \( (M_r = 1508.71) \) where KMe3 represents trimethylated lysine, matched specifically to two members of the L10a family (At1g08360/RPL10aA and At2g27530/RPL10aB, respectively) providing strong evidence for site-specific \( \text{N}^\text{Me} \)-trimethylation of Lys-90, which resides in a conserved region of At1g08360 (RPL10aA) and At2g27530 (RPL10aB). The third manually confirmed hit to a lysine-methylated peptide was to QSGYGGQTKMePVFHK \( (M_r = 1546.79) \), which matches redundantly to the two completely homologous isoforms of the L36a r-protein family, At3g23390 and At4g14320. The annotated spectrum for this peptide is shown in Fig. 6C.

\[ \text{Phosphorylation—Previous analyses of the eukaryotic ribosome from a variety of species have characterized a range of} \]
phosphoproteins from this complex. To obtain a more complete picture of the extent of protein phosphorylation of the Arabidopsis cytosolic ribosome, the in-gel phosphoprotein stain Pro-Q Diamond (Invitrogen) was used to obtain an in-gel phosphorylation profile (Fig. 4B). Significant staining was observed with the phosphostain in several regions of the 1D PAGE and cross-referenced with the Coomassie-stained version of the gel. Ribosomal proteins identified through mass spectrometry in these regions were targeted as putative phosphorylation targets (Fig. 4B). Data from samples previously analyzed from these regions were reinterrogated with Mascot (Matrix Science) using the phosphorylation modification feature. Significant matches were only found for two peptides using this process, and both were derived from the two isoforms of RPS6 (Table IV). In an attempt to identify further phosphoproteins from the samples and further validate the phosphostaining, a phosphopeptide enrichment procedure using TiO2 was used on trypsin-digested ribosome lysates. A total of 10 μg of total ribosome protein was used for phosphoprotein enrichment with the enriched fraction analyzed by nano-LC-MS/MS. Resultant data were queried using Mascot (Matrix Science) with the phosphorylation modification enabled. Several significant matches were obtained to phosphopeptides derived from ribosomal proteins (Table IV), and all except the S6 peptides were shown to be highly enriched by the TiO2 procedure (data not shown). These included peptides derived from the entire family of acidic ribosomal proteins, P0, P1, P2, and P3, and a specific isoform of RPL13. The MS/MS spectra for all phosphopeptides were manually interrogated to ensure high levels of confidence for the match (supplemental Data 6), and the annotated MS/MS diagrams and MS/MS peak lists are provided in supplemental Data 7. The P3 spectrum is shown as an example in Fig. 6D.

**DISCUSSION**

We provide direct mass spectral evidence for gene-specific identification of 31 small subunit proteins, 46 large subunit proteins, and seven P-proteins in the Arabidopsis cytosolic ribosome. This total of 87 specific identifications is based on the tight specifications of peptide number and MS/MS quality outlined under “Results.” In comparison, Chang et al. (19) have claimed 77 gene-specific identifications but have used a much looser specification set. Overall we agree on a set of 56; we do not have significant data on the 26 extra identifications claimed by Chang et al. (19) but do have data on single significant peptides (expect value <0.05) of lower peptide score (peptide score <40) for three of the 26 (supplemental Data 3). In addition we have high quality MS/MS spectra to establish the presence of an additional 32 specific proteins in cytosolic ribosomes. In addition to heterogeneity derived from different ribosomal genes for the same subunit, covalent modification introduces further potential differences in ribosomal protein structure and function. Covalent modifications of ribosomal proteins have been reported in bacteria (28, 29, 31–35), fungi (3, 4, 8, 36–38), mammals (5–7, 39), and plants (19). However, in almost all of these studies, limitations of the techniques used, such as peptide mass fingerprinting, top-down MS analysis of intact r-proteins, and acid hydrolysis to remove rRNA, have precluded detection of acid/base-labile, low stoichiometry, or difficult-to-detect modifications or the determination of precise details about the exact position(s) and structure(s) of the modified residue(s). By taking a targeted proteomics approach that combines chemically mild sample preparation conditions, electrophoretic prefractionation of the ribosomal protein mixture, and phosphoprotein/phosphopeptide enrichment techniques with a highly sensitive and mass-accurate LC-MS/MS analytical technique, we detected an unprecedented number and structural diversity of residue-specific covalent modification sites in the cytosolic ribosome of A. thaliana.

**Conservation of Covalent Modifications between Eukaryotic Ribosomes**—Comparisons of the covalent r-protein modifications identified by this study in Arabidopsis with previous reports from other organisms indicate a considerable degree of conservation across eukaryotes in terms of the types of modifications present and the orthology of the modified subunits (Table V). For example, in S. cerevisiae, mass spectrometry of intact proteins of the large ribosomal subunit has suggested methylation of the yeast r-protein families L1 (L10a in Arabidopsis), L3, L12, L23, L42 (L36a in Arabidopsis), and L43 (L37a in Arabidopsis), although the sites of methylation could not be determined by this approach (4). We have identified methylated residues in members of three orthologous Arabidopsis r-protein families, L10a, L12, and L36a, with these accounting for all of our lysine-methylated peptide matches. Although we did not find evidence for methylation of Arabidopsis orthologues of yeast L3, L23, or L43 (Arabidopsis L3, L23, and L37a, respectively) it is possible that these r-proteins are also methylated and that the peptides carrying these modifications escaped detection in our study. Indeed the fact that we did not detect any cases of arginine methylation, as has been reported for yeast (8, 40) and mammalian (41) ribosomes, is not surprising due to the frequent occurrence of this modification within arginine- and glycine-rich protein regions (42) that, upon trypsinization, would be expected to have given rise to very small peptides giving low m/z peptide ions below the minimum of our parent ion scanning range. Patterns of initiator methionine removal, N-terminal acetylation, and phosphorylation also appear to be largely conserved between plants, animals, and fungi (Table V). Overall the high degree of conservation of covalent modifications between diverse eukaryotes suggests that many of these modifications are fundamentally important for ribosomal function.

**N-terminal Methylation of rpl12**—We have provided strong MS/MS evidence that sites within the first three N-terminal residues of L12 (Pro-Pro-Lys-) are covalently modified by the addition of a total of five methyl groups in A. thaliana cell suspension culture, most probably through N,N-dimethylation.
**Table V**

Post-translational modifications of 80 S ribosomal proteins: evidence for conservation across plants, yeast, and mammals

Post-translational modifications of *Arabidopsis* 80 S ribosomal proteins (identified in this study) are organized by the currently accepted *Arabidopsis* ribosomal protein family nomenclature (17), which is essentially identical to the mammalian nomenclature but slightly different from the nomenclature currently in use for yeast (61). Modifications previously reported for homologous 80 S ribosomal protein families in yeast and mammalian ribosomes are given in the columns labeled " Yeast" and " Mammals," respectively. In cases where the yeast protein family homologous to the named *Arabidopsis* ribosomal protein family is annotated under a different name in yeast, the yeast ribosomal protein family name is given in parentheses in the relevant cell. Modifications that have been assigned to a particular residue on the basis of "bottom-up" MS/MS evidence, mutational studies, or other residue-specific data are highlighted in bold text, and the modified residues are indicated next to the modification as the three-letter code of the amino acid and the residue number (counting from the initiator methionine as residue 1) or " N-term" for modifications of the N terminus. Many of the modifications listed here for yeast and mammalian ribosomal proteins were inferred from top-down mass analyses of intact ribosomal proteins or other "whole-protein" data and hence cannot be unambiguously assigned to specific residues. In these cases, the inferred modifications are shown in normal text (not bold), and a question mark (?) is given in place of a residue identifier. For each listed modification, the literature reference(s) reporting the modification(s) are indicated as superscript single letter codes (explained below). In cases where multiple possible modifications for a protein family were given in a single literature reference, those possible modifications are separated by a forward slash (/).

| Family | AGI numbers | Arabidopsis (this study) | Yeast | Mammals | Conservation |
|--------|-------------|--------------------------|-------|---------|-------------|
| Sa     | A1g72370, A1g504770 | −Met + Acetyl: N-term (rpS0) | −Met + Acetyl: ?a | −Met + Acetyl: N-termb | ++ + |
| S2     | A1g58684, A1g58983, A1g58380, A1g59359, A2g41840, A3g57490 | −Met + Acetyl: ?a asymmetric dimethyl-Arg: ?c | −Met + Acetyl: ?d | −Met + Acetyl: N-termb | ++ |
| S3     | A2g31610, A3g35870, A5g35530 | −Met + Acetyl: N-term | −Metb | −Metb | − |
| S3a    | A3g04840, A4g34670 | (rpS1) | −Met + Acetyl: ?a | −Metb | − |
| S4     | A2g17360, A5g07090, A5g58420 | −Met + Acetyl: N-term | −Met | −Metb | − |
| S5     | A2g32720, A3g11940 | + Phospho: Ser-240 | + Phospho: serines in C-terminal regionf | + Phospho: + + + |
| S6     | A4g31700, A5g10360 | + Acetyl: N-term | + Acetyl: ?f | Acetyl: ?f | + |
| S7     | A1g48830, A3g25260, A5g16130 | + Acetyl: N-term | −Met | −Met | − |
| S8     | A5g20290, A5g59240, A5g53850 | −Metb | −Metb | −Metb | − |
| S9     | A5g15750, A5g15200, A5g39850 | Met retained | Met | Met | − |
| S10    | A4g25740, A5g41520, A5g52650 | −Methyl: ? | −Met + Acetyl: ?f | −Met + Acetyl: N-termf | − |
| S11    | A3g48930, A4g30800, A5g23740 | −Met + Acetyl: ?f | −Met + Acetyl: ?f | −Met + Acetyl: N-termf | − |
| S11a   | A1g15930, A2g32060 | −Met | −Met | −Met | − |
| S12    | A3g60770, A4g00100 | −Met | −Met | −Met | − |
| S13    | A2g36160, A3g11510, A3g52580 | −Met + Acetyl: ?f | −Met + Acetyl: ?f | −Met + Acetyl: N-termf | − |
| S14    | A1g3850, A5g53600, A1g04270, A5g09490, A5g09500, A5g09510, A5g43640 | −Met + Acetyl: N-term | −Met + Acetyl: ?f | −Met + Acetyl: N-termf | − |
| S14a   | A1g07770, A2g19720, A2g39590, A3g46040, A4g29430, A5g59850 | (rpS22) | −Met | −Met | − |
| S15    | A2g09990, A3g04230, A5g18380 | −Met + Acetyl: ?f | −Met + Acetyl: ?f | −Met + Acetyl: N-termf | − |
| S17    | A2g04390, A2g02520, A3g10610, A5g04800 | −Met + Acetyl: N-term | −Met + Acetyl: ?f | −Met + Acetyl: N-termf | − |
| S18    | A1g22780, A1g34030, A4g09800 | −Met + Acetyl: N-term | −Met + Acetyl: ?f | −Met + Acetyl: N-termf | − |
| Family | AGI numbers | Arabidopsis (this study) | Yeast | Mammals | Conservation |
|--------|-------------|--------------------------|-------|---------|--------------|
| S19    | At3g02080, At5g15520, At5g61170 | Met<sup>a</sup> | Met<sup>a</sup> | Met<sup>a</sup> | + |
| S20    | At3g45030, At3g47370, At5g62300 | −Met + Acetyl: N-term | −Met + Acetyl: ?<sup>a</sup> | −Met + Acetyl: N-term<sup>b</sup> | ++ + |
| S21    | At3g53890, At5g27700 | +Acetyl: N-term | +Acetyl: ?<sup>a</sup> | +Acetyl: N-term<sup>b</sup> | ++ + |
| S22    | At3g09680, At5g09260 | −Met + Acetyl: ?<sup>a</sup> | −Met + Acetyl: ?<sup>a</sup> | −Met<sup>b</sup> | ? |
| S23    | At3g04920, At5g28060 | −Met + Acetyl: ?<sup>a</sup> | −Met + Acetyl: ?<sup>a</sup> | −Met<sup>b</sup> | ? |
| S24    | At4g34555, At2g16360, At2g21580, At4g39200 | −Met + possible 2 Methyl: ?<sup>a</sup> | −Met + 2 Methyl: ?<sup>a</sup> | −Met | + |
| S25    | At3g040510, At2g40590, At3g56340 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| S27    | At2g45710, At3g11100, At5g47930 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| S27a   | At1g23410, At2g47110, At3g62250 | (rpS31) | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| S28    | At3g10090, At5g03850, At5g64140 | +Acetyl: ?<sup>a</sup> | +Acetyl: ?<sup>a</sup> | +Acetyl: N-term<sup>b</sup> | + |
| S29    | At3g43980, At3g44100, At4g33865 | Met<sup>a</sup> | Met<sup>a</sup> | Met<sup>a</sup> | + |
| S30    | At2g19750, At4g29390, At5g56670 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| P0     | At1g25260, At2g40010, At3g09200, At3g11250 | Phospho: Ser-305 | Phospho: Ser-305 | Phospho: Ser-305 | ++ + |
| P1     | At3g49460, At5g24510, At1g01100, At4g00810, At5g47700 | Phospho: Ser-102/103 | Phospho: Ser-96<sup>i</sup> | Phospho: Ser-96<sup>i</sup> | ++ |
| P2     | At2g27710, At2g27720, At3g28500, At3g44590, At5g40040 | Met retained + Phospho: Ser-105 | Met retained<sup>a</sup> + Phospho: Ser-96 (P2α)<sup>i</sup> | Met retained<sup>a</sup> + Phospho: Ser-96 (P2β)<sup>i</sup> | +++++ |
| P3     | At4g25890, At5g57290 | +Phospho: Ser-107 | +Phospho: Ser-100 (P2β)<sup>i</sup> | Met retained<sup>a</sup> + Phospho: Ser-100 (P2β)<sup>i</sup> | +++++ |
| L3     | At1g43170, At1g61580 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| L4     | At3g09630, At5g02870 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| L5     | At3g25520, At5g39740 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| L6     | At1g18540, At1g74050, At1g74060 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| L7     | At1g80750, At2g01250, At2g44120, At3g13580 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| L7a    | At2g47610, At3g62870 | (rpL8) − Met<sup>m</sup> | (rpL8) − Met<sup>m</sup> | (rpL8) − Met<sup>m</sup> | + |
| L8     | At2g18020, At3g51190, At4g36130 | Unmodified<sup>m</sup> | Unmodified<sup>m</sup> | Unmodified<sup>m</sup> | + |
| L9     | At1g33120, At1g33140, At4g10450 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| L10    | At1g14320, At1g26910, At1g66580 | Met retained<sup>a</sup> | Met retained<sup>a</sup> | Met retained<sup>a</sup> | + |
| L10a   | At1g08360, At2g27530, At2g27535, At5g22440 | (rpL1) − Met + Acetyl: ?<sup>a</sup> | (rpL1) − Met + Acetyl: ?<sup>a</sup> | (rpL1) − Met + Acetyl: ?<sup>a</sup> | + |
| L11    | At2g2740, At3g58700, At4g18730, At5g45775 | +Trimethyl: Lys-90 − Met + Acetyl: N-term | +Acetyl: N-term<sup>a</sup> | +Acetyl: N-term<sup>a</sup> | + |
| L12    | At2g37190, At3g53430, At5g60670 | +Trimethyl: Lys-3 + Dimethyl: N-term | 6 Methyl: ?/3 Methyl: ? + Acetyl: ?<sup>m</sup> | Possible Methyl: ?<sup>i</sup> | + |
| L13    | At3g48960, At3g49010, At5g23900 | Phospho: Ser-137 | Phospho: Ser-137 | Phospho: Ser-137 | ++ |
| L13a   | At3g07110, At3g24830, At4g13170, At5g48780 | (rpL16) − Met + Acetyl: ?<sup>a</sup>m − Met<sup>m</sup> | −Met<sup>i</sup> | −Met<sup>i</sup> | − |
| L14    | At2g20450, At4g27900 | −Met<sup>a</sup>m | −Met<sup>a</sup>m | −Met<sup>a</sup>m | − |
| L15    | At4g16720, At4g17390 | −Met<sup>a</sup>m | −Met<sup>a</sup>m | −Met<sup>a</sup>m | − |
| L17    | At1g27400, At1g67430 | −Met<sup>a</sup>m | −Met<sup>a</sup>m | −Met<sup>a</sup>m | − |
| L18    | At2g47570, At3g05590, At5g27850 | −Met | −Met | −Met | + + |
| L18a   | At1g29965, At3g24480, At3g14600 | (rpL20) | Met retained<sup>i</sup> | Met retained<sup>i</sup> | ? |
| Family | AGI numbers | Arabidopsis (this study) | Yeast | Mammals | Conservation |
|--------|-------------|--------------------------|-------|---------|--------------|
| L19    | At4g16030, At1g02780, At3g16780, At4g02230 | −Met<sup>a</sup> | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L21    | At1g57860, At1g09590, At1g09690, At1g57660 | −Met | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L22    | At1g02830, At5g27770, At3g05560 | −Met<sup>a</sup> | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L23    | At1g04480, At2g33370, At3g04400 | −Met +7 Methyl: ?/−Met +4 Methyl: ? + Acetyl: ?<sup>a</sup> +2 Methyl: Lys<sup>a</sup> | Possible −Met + Acetyl: ?<sup>a</sup> | ? |
| L23a   | At2g39460, At3g55280 | (rpl25) − Met<sup>a</sup> | Possible − Met + Acetyl: ?<sup>a</sup> | ? |
| L24    | At2g44860, At2g36620, At3g53020 | Unmodified<sup>a</sup> | Met retained, unmodified<sup>a</sup> | + |
| L26    | At3g49910, At5g67510 | −Met/Met retained<sup>a</sup> − Met<sup>a</sup> | Met retained, unmodified<sup>a</sup> | + − |
| L27    | At2g32220, At3g22230, At4g15000 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L27a   | At1g12960, At1g23290, At1g70600 | (rpl28) − Met<sup>a</sup> | Unidentified mod of +16 Da<sup>a</sup> | ? |
| L28    | At2g19730, At4g29410 | −Met + Acetyl: N-term<sup>a</sup> | −Met + Acetyl: N-term<sup>a</sup> | + |
| L29    | At3g06680, At3g06700 | −Met<sup>a</sup> | −Met +1 Methyl: Lys-4<sup>a</sup> | − |
| L30    | At1g36240, At1g77940, At3g18740 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L31    | At1g19740, At4g26230, At5g56710 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L32    | At4g18100, At5g46430 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L34    | At1g26880, At1g68620, At3g28900 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L35    | At2g93930, At3g9500, At3g59170, At5g02610 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L35a   | At1g07070, At1g41880, At1g74270, At3g55750 | (rpl33) − Met + Acetyl: ?<sup>a</sup> − Met<sup>a</sup> | ? |
| L36    | At2g37600, At3g53740, At5g02450 | −Met<sup>a</sup> | −Met + Acetyl: ?/−Met<sup>a</sup> − Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L36a   | At3g23390, At4g14320 | + Methyl: Lys-55<sup>a</sup> | (rpl42) Met retained +2 Methyl: ? <sup>a</sup> − Met +2 Methyl: ?<sup>a</sup> | −Met<sup>a</sup> | ++ + + |
| L37    | At1g15250, At1g52300, At3g16080 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L37a   | At3g10950, At3g60245 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L38    | At2g43460, At3g59540 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L39    | At2g25210, At3g02190, At4g31985 | −Met<sup>a</sup> | −Met<sup>a</sup> | ? |
| L40    | At3g52590, At2g36170 | Fragment, 77−128<sup>a</sup> | −Met (other mods possible)<sup>a</sup> | ? |
| L41    | At2g40205, At3g08520, At3g11120, At3g56020 | Unmodified<sup>a</sup> | Unmodified<sup>a</sup> | + |
| RACK1  | At1g18080, At1g48630, At3g18130 | ? |

93 cases of conservation/121 opportunities for comparison = 77% conservation

<sup>a</sup> Arnold et al. (3).
<sup>b</sup> Louie et al. (5).
<sup>c</sup> Bachand and Silver (8).
<sup>d</sup> Swiercz et al. (41).
<sup>e</sup> Yu et al. (7).
<sup>f</sup> Kruse et al. (64).
<sup>g</sup> Krieg et al. (54).
<sup>h</sup> Meng et al. (47).
<sup>i</sup> Rodriguez-Gabriel et al. (62).
<sup>j</sup> Odintsova et al. (6).
<sup>k</sup> Hasler et al. (63).
<sup>l</sup> Zambrano et al. (58).
<sup>m</sup> Lee et al. (4).
<sup>n</sup> Chern et al. (40).
<sup>o</sup> Porras-Yakushi et al. (37).
<sup>p</sup> Mazumder et al. (9).
<sup>q</sup> Porras-Yakushi et al. (38).
of the N-terminal proline residue and trimethylation of the ε-amino group of lysine 3. Interestingly L12 is the only Arabidopsis cytosolic ribosomal protein predicted to have the Pro-Pro-Lys- N-terminal sequence after methionine removal, and this sequence is conserved in almost every eukaryotic L12 protein sequence available. At the time of this publication, we could find in the literature relatively few reported examples of proteins modified by N-terminal proline methylation: the histone H2B from the starfish Asterias rubens that was confirmed by NMR to have an N,N-dimethylproline structure at its N terminus (43), the orthologous histone H2B of Drosophila (44), the cytochrome c557 of the protozoan Crithidia oncopelti (45, 46), and the cytosolic ribosomal protein S25 of S. cerevisiae (47). Remarkably all of these proteins were shown to have the same Pro-Pro-Lys- N-terminal structure as we have shown here for ribosomal protein L12 in Arabidopsis. This finding provides further support for previous suggestions that the N-terminal Pro-Pro-Lys motif may be recognized by an as yet unknown N-terminal N-methyltransferase enzyme that is widely distributed in nature (48) and, to the best of our knowledge, provides the first evidence that such an enzyme occurs in plants.

Evidence has very recently emerged that the N-terminal region of the S. cerevisiae L12 orthologue is also pentamethylated with the detection, by LC-MS/MS, of a pentamethylated form of the N-terminal peptide, PPKFDPNEVKYLYLR, in a proteolytic digest of HPLC-purified rpL12ab (37). The authors of that study interpreted their MS/MS spectrum of this peptide to suggest the presence of a mixture consisting of two iso- meric forms: one with dimethylation at Lys-3 and trimethylation at Lys-10 and another with trimethylation at Lys-3 and dimethylation at Lys-10. Given the similarity of our findings with those of Porras-Yakushi et al. (37) and the presence of the putative motif for N-terminal proline dimethylation, Pro-Pro-Lys-, in the yeast rpL12 protein, we decided to examine the presented yeast MS/MS spectrum for evidence of combined N-terminal proline dimethylation and Lys-3 trimethylation. Indeed we found strong evidence to suggest that a significant proportion of the yeast peptide is modified in the same way as the equivalent peptide from Arabidopsis (see, in particular, the high intensity doubly charged b5 fragment at m/z 328.19 exhibiting the proline effect enhanced by the adjacent Asp residue, the doubly charged b3 ion at m/z 197.15, the singly charged y7 ion at m/z 954.53, and the singly charged y10 ion at m/z 1294.63 also exhibiting the Asp/Pro effect-induced enhancement of intensity). Although Porras-Yakushi et al. (37) identified MS/MS signals consistent with their proposed peptide structures (which include di- or trimethylation at Lys-10), the relative intensities of the above signals in the presented MS/MS spectrum suggest that at least a significant proportion of the peptide population responsible for the spectrum consists of a form that is not methylated at Lys-10 but carries the five methyl groups within the first three residues. Porras-Yakushi et al. (37) observed that deletion of the S. cerevisiae gene ydr198c, encoding a putative suppressor of variegation, Enhancer-of-ze ste, Trithorax domain-containing lysine N-methyltransferase, led to the loss of three methyl groups and the detection of only two methyl groups on the N-terminal peptide of rpL12ab. Interestingly examination of the presented MS/MS spectrum for this dimethylated peptide derived from the Δydr198c mutant strain revealed the disappearance of the putatively pentamethylated b5 ion (present at m/z 328.19 in the wild-type spectrum) and the appearance of a peak with similar relative intensity at m/z 307.15. The m/z difference between these putative doubly charged b5 ion signals (21.034 amu) is consistent with a loss of 42.07 Da (equivalent to three methyl groups) from the first five residues of the peptide. Similarly the putative pentamethylated, doubly charged b3 ion (observed at m/z 197.15 in the wild-type spectrum) is apparently missing in the Δydr198c spectrum, whereas the relative intensity of a putative dimethylated, singly charged b3 ion is greatly increased at m/z 351.23. In addition to the 42.06-Da mass shift, the shift in charge state from z = 2 to z = 1, observed for the b3 signal, is consistent with the loss of constitutive positive charge associated with the quaternary ammonium group of the trimethyllysine residue. Together these observations strongly suggest that deletion of ydr198c gene from S. cerevisiae has no effect on the dimethylation of the N-terminal proline residue but leads to a loss of trimethylation from Lys-3 in rpL12ab. It should be noted that our interpretations of MS/MS data presented by Porras-Yakushi et al. (37) are not mutually exclusive with those of the authors and that the existence of multiple isomeric methylation states for rpL12ab in S. cerevisiae is an interesting possibility worthy of further investigation. We could find no convincing evidence for such a phenomenon in Arabidopsis rpL12 or any other Arabidopsis ribosomal protein for which we detected lysine-methylated peptides while manually interpreting the MS/MS spectra supporting our claims of covalent protein modification. The A. thaliana genome contains at least 29 actively transcribed genes encoding suppressor of variegation, Enhancer-of-ze ste, Trithorax domain-containing proteins (49). However, a role for any of these proteins in the methylation of Arabidopsis ribosomal proteins has yet to be established.

Phosphoproteins of the Arabidopsis Cytosolic Ribosome— The cytosolic ribosome represents a well characterized system for the study of protein phosphorylation. The most widely studied components, RPS6 and the acidic P-proteins (P0–3), actively modulate ribosomal function through their phosphorylation and are involved in protein synthesis (50, 51). The RPS6 protein is a part of the small head region of the 40 S subunit, which is a central location for interactions with mRNA, tRNA, and factors that drive initiation of translation (52). The C terminus of RPS6 is sequentially phosphorylated on five serine residues in response to a variety of stimuli and developmental cues (53, 54). This C-terminal region is highly conserved across eukaryotic lineages, and in plants its mul-
tiphosphorylation sites have been well characterized in maize and *Arabidopsis* (26, 27). Initial characterization of this region in mammals has demonstrated an ambiguity in phosphorylation site determination (54). Previous analysis of this region in *Arabidopsis* has only utilized MALDI-TOF MS analysis, limiting the determination of the specific site(s) of phosphorylation (19). Confirmation of single phosphorylation sites for the two *Arabidopsis* isoforms of RPS6 were confirmed for the first time here using MS/MS interrogation at Ser-240 for RPS6A and RPS6B.

The acidic P-proteins form a multimeric complex that comprises the lateral stalk of the 60 S ribosomal subunit (55). This complex is highly conserved across eukaryotes and is directly involved in translational regulation through interactions with associated factors. In addition to P-proteins P0, P1, and P2, plants possess an additional P1/P2 type protein termed P3 (56). The P-proteins were initially identified as phosphoproteins in the yeast ribosome (57), and their phosphorylation states likely play a direct role in protein synthesis and translational responses to external stimuli (26, 58). Phosphorylation sites in the yeast P-proteins occur in the C-terminal region of the proteins (58), a feature that has now been confirmed for all members of the *Arabidopsis* family of ribosomal P-proteins. Although recent technical reports using phosphoproteomics technologies have reported the presence of several of these phosphopeptides from *Arabidopsis* (RPP1A and RPP2B (59) and RPPOB (60)), this study has demonstrated that the entire acidic P-protein family in *Arabidopsis* (P0, P1, P2, and P3) all contain phosphorylated C-terminal regions.

**Conclusions**—The complexity of the plant cytosolic ribosome in terms of its size, number of protein constituents, and the evolutionary history, divergence, and multiplication of the genes encoding its subunits makes it a major challenge for biochemical analysis. Within this complexity are undoubtedly key elements in understanding ribosome structure-function relationships and the regulation of translation in plants that remain undiscovered. Analysis of peptides derived from proteolysis of ribosomes provides both protein composition and post-translational modification data. Although the biological roles of most of these covalent modifications are presently unknown, the specific knowledge generated by this proteomics characterization study will enable the immediate design of targeted genetics experiments to explore the roles of these covalent modifications and the mechanisms through which they occur.

*S* This work was supported in part by grants from the Australian Research Council (ARC) through the Centres of Excellence Program and the Discovery Program. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by a Ph.D. scholarship from the Grains Research and Development Corp.

§ An ARC Australian postdoctoral fellow.

**REFERENCES**

1. Doudna, J. A., and Rath, V. L. (2002) Structure and function of the eukaryotic ribosome: the next frontier. *Cell* **109**, 153–156

2. O’Brien, T. W. (2003) Properties of human mitochondrial ribosomes. *IUBMB Life* **55**, 505–513

3. Arnold, R. J., Polevoda, B., Reilly, J. P., and Sherman, F. (1999) The action of N-terminal acetyltransferases on yeast ribosomal proteins. *J. Biol. Chem.* **274**, 37035–37040

4. Lee, S. W., Berger, S. J., Martinovic, S., Pasa-Tolic, L., Anderson, G. A., Shen, Y., Zhao, R., and Smith, R. D. (2002) Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 5942–5947

5. Louie, D. F., Resing, K. A., Lewis, T. S., and Ahn, N. G. (1996) Mass spectrometric analysis of 40 S ribosomal proteins from Rat-1 fibroblasts. *J. Biol. Chem.* **271**, 28189–28198

6. Odintsova, T. I., Muller, E. C., Ivanov, A. V., Egorov, T. A., Biener, R., Vladimirov, S. N., Kostka, S., Otto, A., Wittmann-Liebold, B., and Karpova, G. G. (2003) Characterization and analysis of posttranslational modifications of the human large cytoplasmic ribosomal subunit proteins by mass spectrometry and Edman sequencing. *J. Proteome Res.* **2**, 249–258

7. Yu, Y., Ji, H., Doudna, J. A., and Leary, J. A. (2005) Mass spectrometric analysis of the human 40S ribosomal subunit: native and HCV IRES-bound complexes. *Protein Sci.* **14**, 1438–1446

8. Bachand, F., and Silver, P. A. (2004) PRMT3 is a ribosomal protein methytransferase that affects the cellular levels of ribosomal subunits. *EMBO J.* **23**, 2641–2650

9. Mazumder, B., Sampath, P., Seshadri, V., Maitra, R. K., DiCorleto, P. E., and Fox, P. L. (2003) Regulated release of L13a from the 60S ribosomal subunit as a mechanism of transcript-specific translational control. *Cell* **115**, 187–198

10. Cammarano, P., Pons, S., Romeo, A., Galdieri, M., and Gualerzi, C. (1972) Characterization of unfolded and compact ribosomal subunits from plants and their relationship to those of lower and higher animals: evidence for physicochemical heterogeneity among eucaryotic ribosomes. *Biochim. Biophys. Acta* **281**, 571–596

11. Bailey-Serres, J., and Freeling, M. (1990) Hypoxic stress-induced changes in ribosomes of maize seedling roots. *Plant Physiol.* **94**, 1237–1243

12. Capel, M. S., and Bourque, D. P. (1982) Characterization of *Nicotiana tabacum* chloroplast and cytoplasmic ribosomal proteins. *J. Biol. Chem.* **257**, 7746–7755

13. Gannt, J., and Key, J. (1983) Auxin-induced changes in the level of translatable ribosomal protein messenger ribonucleic acids in soybean hypocotyl. *Biochemistry* **22**, 4131–4139

14. Koyama, K., Wada, A., Maki, Y., and Tanaka, A. (1996) Changes in the protein composition of cytoplasmic ribosomes during the greening of etiolated barley leaves. *Physiol. Plant* **96**, 85–90

15. Sikorski, M. M., Przybyl, D., Legocki, A. B., and Nierhaus, K. H. (1983) Group fractionation of wheat germ ribosomal proteins. *Plant Sci. Lett.* **30**, 303–320

16. Scharf, K.-D., and Nover, L. (1987) Control of ribosome biosynthesis in plant cell cultures under heat shock conditions. II Ribosomal Proteins. *Biochim. Biophys. Acta* **896**, 44–57

17. Barakat, A., Szczek-Miranda, K., Chang, I. F., Guoyt, R., Blanc, G., Cooke, R., Delseny, M., and Bailey-Serres, J. (2001) The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiol.* **127**, 398–415

18. Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Dernar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J. U. (2005) A gene expression map of *Arabidopsis* thaliana development. Nat. Genet. **37**, 501–506

19. Chang, I. F., Szczek-Miranda, K., Pan, S., and Bailey-Serres, J. (2005) Proteomic characterization of evolutionarily conserved and variable proteins of *Arabidopsis* cytosolic ribosomes. *Plant Physiol.* **137**, 848–862

20. Giavalisco, P., Wilson, D., Kreitler, T., Lehraich, H., Kloes, J., Goborn, J., and Fucini, P. (2005) High heterogeneity within the ribosomal proteins of the *Arabidopsis thaliana* 80S ribosome. *Plant Mol. Biol.* **57**, 577–591

368 *Molecular & Cellular Proteomics 7.2*
21. Cagney, G., Amiri, S., Premawaradeniya, T., Lindo, M., and Emili, A. (2003) In silico proteome analysis to facilitate proteomics experiments using mass spectrometry. *Proteome Sci.* 1, 5

22. Millar, A. H., Liddell, A., and Leaver, C. J. (2001) Isolation and subfractionation of mitochondria from plants. *Methods Cell Biol.* 65, 53–74

23. Lin, C., and Key, J. (1971) Dissociation of N2 gas-induced monomeric ribosomes and functioning of the derived subunits in protein synthesis in pea. *Plant Physiol.* 48, 547–552

24. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jorgensen, T. D. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* 4, 873–886

25. Adams, K. L., Daley, D. O., Whelan, J., and Palmer, J. D. (2002) Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell* 14, 931–943

26. Szick-Miranda, K., and Bailey-Serres, J. (2004) Regulated heterogeneity in 12-kDa P-protein phosphorylation and composition of ribosomes in maize (Zea mays L.). *J. Biol. Chem.* 276, 10921–10928

27. Williams, A. J., Wemer-Fraczek, J., Chang, I. F., and Bailey-Serres, J. (2003) Regulated phosphorylation of 40S ribosomal protein S6 in root tips of maize. *Plant Physiol.* 132, 2086–2097

28. Ramagopal, S., and Subramanian, A. R. (1974) Alteration in the acetylation level of ribosomal protein L12 during growth cycle of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 71, 2136–2140

29. Strader, M. B., Verberkmoes, N. C., Tabb, D. L., Connelly, H. M., Barton, J. W., Bruce, B. D., Peletier, D. A., Davison, B. H., Hetlich, R. L., Lanner, F. W., and Hurst, G. B. (2004) Characterization of the 70S ribosome from *Rhodopseudomonas palustris* using an integrated “top-down” and “bot- tom-up” mass spectrometric approach. *J. Proteome Res.* 3, 965–978

30. Dalboge, H., Bayne, S., and Pedersen, J. (1990) In vivo processing of N-terminal methionine in E. coli. *FEBS Lett.* 266, 1–3

31. Arnold, R. J., and Reilly, J. P. (1999) Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry, *Anal. Biochem.* 269, 105–112

32. Isono, K., and Isono, S. (1980) Ribosomal protein modification in *Escherichia coli*. II. Studies of a mutant lacking the N-terminal acetylation of protein S18. *Mol. Gen. Genet.* 177, 645–651

33. Cumberledge, A. G., and Isono, K. (1979) Ribosomal protein modification in *Escherichia coli*. I. A mutant lacking the N-terminal acetylation of protein S5 exhibits thermostability. *J. Mol. Biol.* 131, 169–189

34. Kowalak, J. A., and Walsh, K. A. (1996) Identification of the 40 S ribosomal proteins and their posttranslational modifications by mass spectrometry. *Anal. Biochem.* 243, 49–59

35. Kim, S., Lew, B., and Chang, F. N. (1977) Enzymatic methyl esterification of methionine at proline, lysine, and arginine residues during heat shock. *J. Biol. Chem.* 252, 4866–4892

36. Stock, A., Clarke, S., Clarke, C., and Stock, J. (1987) N-terminal methylation of proteins: structure, function and specificity. *FEBS Lett.* 220, 8–14

37. Hênaff, M., Aeschbach, D., Zeller, H., and Hennet, T. (2001) The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res.* 29, 4319–4333

38. Montoya-García, L., Murcia-Otero, V., Aguilar, R., and Sanchez de Jimenez, E. (2002) Regulation of acidic ribosomal protein expression and phosphorylation in maize. *Biochemistry* 41, 10166–10172

39. Spahn, C. M., Beckmann, R., Eswar, N., Penczek, P. A., Sali, A., Blobel, G., and Frank, J. (2001) Structure of the 80S ribosome from Saccharomyces cerevisiae—tRNA-ribosome and subunit-subunit interactions. *Cell* 107, 373–386

40. Szick, K., Springer, M., and Bailey-Serres, J. (1998) Evolutionary analyses of the 12-kDa acidic ribosomal P-proteins reveal a distinct protein of higher plant ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2378–2383

41. Zinker, S., and Warner, J. R. (1976) The ribosomal proteins of Saccharomyces cerevisiae. Phosphorylated and exchangeable proteins. *J. Biol. Chem.* 251, 1799–1807

42. Zambrano, R., Briones, E., Remacha, M., and Ballesta, J. P. (1997) Phosphorylation of the acidic ribosomal P proteins in Saccharomyces cerevisiae: a reappraisal. *Biochemistry* 36, 14439–14446

43. Wolsch, F., Winkoepfl, S., and Weckwerth, W. (2005) Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC). *Proteomics* 5, 4389–4397

44. Laugesen, S., Messinese, E., Hem, S., Pichereaux, C., Grat, S., Ranjeva, R., Rossignol, M., and Bono, J. J. (2006) Phosphoproteins analysis in plants: a proteomic approach. *Phytochemistry* 67, 2208–2214

45. Mager, W. H., Planta, R. J., Ballesta, J. G., Lee, J. C., Mizuta, K., Suzuki, K., Warner, J. R., and Woolford, J. (1997) A new nomenclature for the cytoplasmic ribosomal proteins of Saccharomyces cerevisiae. *Nucleic Acids Res.* 25, 4872–4875

46. Rodriguez-Galicia, M. A., Remacha, M., and Ballesta, J. P. (1998) Phosphorylation of ribosomal protein P0 is not essential for ribosomal function but can affect translation. *Biochemistry* 37, 16620–16626

47. Hasler, P., Bro, N., Weissbach, H., Parnassa, A. P., and Elkon, K. B. (1991) Ribosomal proteins P0, P1, and P2 are phosphorylated by casein kinase II at their conserved carboxyl termini. *J. Biol. Chem.* 266, 13815–13820

48. Kruisz, J., Johnson, S. P., and Warner, J. R. (1985) Phosphorylation of the yeast equivalent of ribosomal protein S6 is not essential for growth. *Proc. Natl. Acad. Sci. U. S. A.* 82, 7515–7519

**Molecular & Cellular Proteomics 7.2** 369