

# Gene Expression Analyzed by High-resolution State Array Analysis and Quantitative Proteomics

RESPONSE OF YEAST TO MATING PHEROMONE\*<sup>[S]</sup>

Vivian L. MacKay<sup>‡</sup>, Xiaohong Li<sup>§</sup>, Mark R. Flory<sup>¶</sup>, Eileen Turcott<sup>‡</sup>, G. Lynn Law<sup>‡</sup>, Kyle A. Serikawa<sup>‡</sup>, X. L. Xu<sup>§</sup>, Hookeun Lee<sup>¶</sup>, David R. Goodlett<sup>¶</sup>, Ruedi Aebersold<sup>¶</sup>, Lue Ping Zhao<sup>§</sup>, and David R. Morris<sup>‡</sup>

The transcriptome provides the database from which a cell assembles its collection of proteins. Translation of individual mRNA species into their encoded proteins is regulated, producing discrepancies between mRNA and protein levels. Using a new modeling approach to data analysis, a striking diversity is revealed in association of the transcriptome with the translational machinery. Each mRNA has its own pattern of ribosome loading, a circumstance that provides an extraordinary dynamic range of regulation, above and beyond actual transcript levels. Using this approach together with quantitative proteomics, we explored the immediate changes in gene expression in response to activation of a mitogen-activated protein kinase pathway in yeast by mating pheromone. Interestingly, in 26% of those transcripts where the predicted protein synthesis rate changed by at least 3-fold, more than half of these changes resulted from altered translational efficiencies. These observations underscore that analysis of transcript level, albeit extremely important, is insufficient by itself to describe completely the phenotypes of cells under different conditions. *Molecular & Cellular Proteomics* 3:478–489, 2004.

Genome-wide analysis of gene expression generally involves quantitative surveys of the transcriptome by such powerful technologies as interrogation of microarrays or serial analysis of gene expression (SAGE), which provide predictors of physiological state or cellular phenotype. Indeed, genome-scale transcript analysis has produced clear successes; e.g. discovery of co-regulated networks of transcripts in yeast (1–5), intracellular locations of mRNAs (6, 7), and clinically valuable phenotypes of cancer cells (8).

Despite the powerful information obtained, documentation

of a transcriptome provides only the inventory that is available for a cell to draw upon for translation under certain physiological conditions. Proteins are the effectors of cell phenotype, and their levels and activities do not necessarily correlate with mRNA levels (9–12). One source of this lack of correlation is discrepancies in protein half-lives. A second arises from the fact that the synthesis of individual protein species is regulated, not only by transcript level, but by *cis* elements that confer unique translational properties on individual mRNA molecules (13). In light of this latter point, it would be valuable to supplement mRNA expression patterns with estimates of translation efficiencies of individual transcripts.

Translation occurs on ribosomes, and variable numbers of ribosomes are loaded onto actively translated mRNAs, forming polysomes of various sizes. The density of ribosome packing on transcripts is proportional to the rate of synthesis of the protein products (14, 15). Several groups have carried out polysome fractionation prior to transcript analysis (16–20), but in these approaches fractions from sucrose-gradient centrifugation were combined into pools before transcript analysis, thereby losing the rich information associated with the mRNA distributions across these fractions. In the present study, we undertook a “high resolution” analysis of transcript distributions across these polysome profiles, defining ribosome-loading parameters for each detectable mRNA species. Selective, quantitatively significant changes in translation of the yeast transcriptome were found in response to the mating pheromone  $\alpha$ -factor. Groups of functionally related genes were found to be co-regulated by a combination of altered transcript levels and translational efficiencies; these observations were supplemented by proteome analysis.

## EXPERIMENTAL PROCEDURES

**Yeast Cultures**—All experiments used either strain BY2125 (*MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100 ssd1-d*; W303 background; kindly provided by L. Breeden, Fred Hutchinson Cancer Research Center, Seattle, WA) or a derivative VM1601 with the additional mutations  $\Delta his3::TRP1 \Delta upf3::LEU2$ . For the genome-wide experiments, BY2125 cells were grown at 25 °C in rich-glucose medium (1% yeast extract, 2% peptone, 2% glucose) to mid-log phase ( $\sim 1 \times 10^7$  cells/ml). For pheromone induction, cells were treated with 10

From the <sup>‡</sup>Department of Biochemistry, University of Washington, Seattle, WA 98195; <sup>§</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; and <sup>¶</sup>Institute for Systems Biology, Seattle, WA 98103

Received, December 4, 2003, and in revised form, January 31, 2004

Published, MCP Papers in Press, February 6, 2004, DOI 10.1074/mcp.M300129-MCP200

$\mu\text{g/ml}$  ( $7 \mu\text{M}$ )  $\alpha$ -factor (United Biochemical Research, Inc., Seattle, WA) for 30 min ( $<30\%$  of a cell cycle under these conditions).

**Polysome Fractionation and RNA Isolation**—Polysomes were prepared and fractionated by modification of a previously described method (21, 22) that employed detergent extraction of cell lysates and centrifugation through sucrose gradients in high salt to disrupt inactive 80S monomers (see supplemental materials).

After centrifugation (see Fig. 2a), fractions from identical gradients were pooled on ice and adjusted to  $0.5\%$  SDS after addition of three artificial mRNAs to monitor recovery and integrity after purification (Universal ScoreCard control poly(A)<sup>+</sup>-RNAs; Amersham Biosciences, Sunnyvale, CA). These “utility control” RNAs were added at  $0.075$ ,  $0.75$ , or  $15 \text{ pg}$  per pooled fraction before precipitation of RNA at  $-70^\circ\text{C}$  with  $2.5$  volumes  $100\%$  ethanol. RNA samples from each of the  $25$  pooled fractions were purified using Qiagen RNeasy mid-kits, and the eluted RNA was precipitated with  $1/10$  volume of  $10 \text{ M}$  LiCl, resuspended in  $25 \mu\text{l}$  of RNase-free  $\text{H}_2\text{O}$ , quantitated by absorbance at  $260 \text{ nm}$ , and stored at  $-70^\circ\text{C}$ .

For total unfractionated RNA,  $600 \text{ ml}$  of cells from the same culture as the polysome preparation were pelleted, washed with ice-cold RNase-free water, quick frozen, then subsequently resuspended in lysis buffer, processed to the clarified lysate stage, and ethanol-precipitated, as above. The resulting pellet was resuspended in water, and total RNA was isolated using Qiagen midi-columns followed by LiCl precipitation. The RNA samples were pooled in a final volume of  $1 \text{ ml}$  of RNase-free  $\text{H}_2\text{O}$  and the  $A_{260}$  was measured.

Specific transcripts were quantitated with quantitative real-time PCR (Q-PCR)<sup>1</sup> using an iCycler (Bio-Rad, Hercules, CA) and SYBR-green detection of products (according to Bio-Rad’s specifications). Primers for the reactions were designed to amplify regions of  $75$ – $200$  nucleotides near the  $3'$  end of the coding region. Q-PCR data were normalized with one of the artificial utility control RNAs added before purification of RNA in the gradient fractions (see above).

**Microarray Hybridization and Quantitation**—For target generation,  $80 \mu\text{g}$  of total RNA from the peak fraction in the polysome region of the gradient and the equivalent volume from all other fractions was converted to fluorescently labeled cDNA using Cy3-labeled dCTP (Amersham Pharmacia Biotech, Uppsala, Sweden) and primed with oligo-dT<sub>25</sub> with a G/C/A  $3'$  anchor as described previously (20). For each reaction,  $31 \mu\text{l}$  of the “Test spike mix” of Universal ScoreCard control RNAs was added to monitor labeling and hybridization efficiencies, as well as to generate a standard curve determining the linear range of the signal intensity. Similarly,  $2 \text{ mg}$  of unfractionated total RNA was converted to Cy5-labeled cDNA target after addition of the “Reference spike mix” of Universal ScoreCard controls. The yield of the Cy3-labeled target for the peak fraction was  $52 \text{ pmol}$  of product; for each fraction, the entire Cy3-labeled target was mixed with  $31 \text{ pmol}$  of Cy5-labeled target from unfractionated RNA, then taken to dryness and resuspended in  $80 \mu\text{l}$  of hybridization buffer ( $50\%$  formamide,  $5\times \text{SSC}$ ,  $5\times \text{Denhardt's solution}$ ,  $0.1\%$  SDS).

Custom yeast high-density microarrays spotted with PCR-amplified genomic DNAs corresponding to open reading frames (ORFs), with negative control artificial cDNAs, and with artificial cDNAs for the Universal ScoreCard control RNAs were produced by the Center for Expression Array Analysis at the University of Washington (ra.microslu.washington.edu). Each slide represents two arrays of  $\sim 6000$  ORFs comprising nearly the entire yeast genome.

The hybridization mixes of Cy3- and Cy5-labeled targets were heated at  $96^\circ\text{C}$  for  $3 \text{ min}$ , cooled on ice for  $30 \text{ s}$ , and spun briefly. For

each fraction,  $40 \mu\text{l}$  was applied to each of two washed slides, followed by hybridization and washing as described previously (20).

**Normalization and Analysis of Microarray Data**—Initial processing of the raw data included correction based on negative control spots, calibration with internal standards, RNA recovery correction by utility controls, as well as filtering those genes of low abundance or with inconsistent Cy3:Cy5 ratios, as detailed below:

1. Correction by negative controls. All images with intensity values below those for the negative control cDNA spots were transformed to be equivalent to the negative controls.
2. Calibration. To correct for possible array-specific effects, standard curves were generated for each array from  $12$  replicates of each of  $10$  standards in the calibration controls added at different concentrations, as provided by Amersham Bioscience. By linear interpolation of signal intensity, we calculated each image’s expression level relative to the absolute amounts of the artificial calibration controls on the arrays. In the plus  $\alpha$ -factor dataset, two arrays from fraction  $4$  of the sucrose gradient were deleted because of poor quality.
3. Recovery correction. After calibration, the expression levels of the utility controls (added in known amounts to each gradient fraction before RNA purification) were used to determine mRNA recovery and to adjust and standardize the Cy3-labeled target population used on each array.
4. Filtering of genes with low expression values. The same amount of Cy5-labeled target made from unfractionated RNA was used for each hybridization, and the image intensity reflects the relative abundance of individual transcripts. Using a paired  $t$  test with a significance level of  $5\%$ , we filtered out those mRNAs when a majority of the  $100$  Cy5 intensity measurements were lower than the low boundary of the linear range of the Cy5 calibration curve.
5. Filtering of inconsistent ratios. For every gradient fraction, there are four replicate arrays and hence four repeated measurements of the Cy3:Cy5 ratios. When Cy5 values are low, ratios for specific mRNAs could become unstable and vary radically, particularly for the first few fractions where both RNA concentrations and recoveries are lower than for later fractions. Using the assumptions that repeatedly measured ratios are largely consistent and do not vary substantially between two adjacent fractions, we adopted two filtering strategies: 1) when one of the four measurements is a significant outlier, it is filtered out in computing the averaged signal, and 2) when two out of four ratios are extremely high for fraction  $1$  compared with the other two ratios, the average of the two low ratio values is used. Ratios in the first fraction are expected to be generally low, as it is unlikely that authentic transcripts would sediment in this fraction under these centrifugation conditions. Also, the Cy5 expression values for fractions  $2$  and  $5$  were low compared with the other Cy5 signals across the gradient; we therefore used the averaged Cy5 values from the two neighboring fractions for these calculations.

**Assumptions and Derivation of the Multiple Peak Model (MPM)**—Transcripts with varying numbers of ribosomes, ranging from  $0$  up to an estimated  $25$  ribosomes, have different quantized weights, due to the large size of a ribosome relative to an mRNA. Using sucrose-gradient centrifugation and fractionation, we can separate this disparate pool of ribosome-bound mRNAs by weight, then identify and quantify each mRNA species by microarray technology. The individual mRNA profiles across the gradient provide the weight distribution for that molecule and hence insight on its translational properties.

The mathematical objective here is to model the weight distributions of all molecules. Let  $Y_{jk}$  denote the abundance measurement for

<sup>1</sup> The abbreviations used are: Q-PCR, quantitative real-time PCR; ORF, open reading frame; ICAT, isotope-coded affinity tag; MS, mass spectrometry; uORF, upstream ORF; MPM, multiple peak model; CAI, codon adaptation index; TSAA, translation state array analysis.

the  $j$ th transcript in the  $k$ th fraction. Let  $f_k = k$  denote the fraction number, from 1 to 25. The fraction number is indicative of the weight of the mRNA-ribosome complex, which is denoted by  $w_k$ . Conceptually, experimental data of relevance may be represented as in supplemental Table 1S. The abundance measurement  $Y_{jk}$  is the averaged Cy3:Cy5 ratio for the four replicate arrays. Data have been normalized following the protocol described above. Because the modeling process is the same for all individual transcripts, we drop the subscript  $j$  for simplicity hereafter, unless noted otherwise.

To facilitate modeling, we make the following assumptions. First, for each mRNA, the mixture of mRNA complexes has at most four different particles: mRNP particles, monosomes, polysomes, and an uncharacterized, rapidly sedimenting component. The second assumption is that the weight of a specific particle distributes as a Gaussian distribution, centering on the actual weight of the particle population. This distribution results, at least in part, from diffusion theory. Let  $R$  take value 1, 2, 3, and 4 and denote four different populations of particles. Under this assumption, the distribution of abundance given the particle population  $R$  may be written by

$$Y|R \approx N(\mu_R, \sigma_R^2),$$

where  $\mu_R$  is the mean and  $\sigma_R^2$  measures the spread of the peak.

Under these two assumptions, one can derive models for the observed expression levels in the  $k$ th fraction interval via the following steps. First, the abundance in the  $k$ th fraction is proportional to the total number of particles in the interval,

$$Y_k \propto \sum_i I(w_k \leq D_i < w_{k+1}) \approx NE[I(w_k \leq D < w_{k+1})],$$

where the summation is over all particles, indicator  $I(w_k \leq D_i < w_{k+1})$  for the particle in the  $k$ th fraction,  $N$  represents the total number of particles, and  $E[I(w_k \leq D_i < w_{k+1})]$  represents the expectation. Now the expectation may be computed as

$$\int_{w_k}^{w_{k+1}} g(D) dD = \int_{w_k}^{w_{k+1}} \sum_{R=1}^4 g(D|R) g(R) dD,$$

where  $g(D|R)$  represents a Gaussian distribution with mean  $\mu_R$  and with the variance  $\sigma_R^2$ . Resulting from the above derivation is the model for multiple peaks:

$$Y_k \propto N \sum_{R=1}^4 g(R) \int_{w_k}^{w_{k+1}} g(D|R) dD = \sum_{R=1}^4 \beta_R \left[ \Phi\left(\frac{w_{k+1} - \mu_R}{\sigma_R}\right) - \Phi\left(\frac{w_k - \mu_R}{\sigma_R}\right) \right]$$

where  $\beta_R$  is the peak height and quantifies the abundance of each particle, and  $\Phi(\cdot)$  is the cumulative distribution function of Gaussian distribution. The above representation, despite being motivated statistically, is actually connected with phenomenological theory of sedimentation processes in the ultracentrifuge (23). In fact, the mixture of Gaussian distributions, with appropriate re-parameterization, could approximate the solutions to the set of partial differential equations used to describe the consequences of sedimentation and diffusion in the ultracentrifuge.

Applying the above model to gene expression analysis, we need to take into account mRNA baseline abundance and random fluctuation

due to uncontrolled random variations. The revised model may now be written as

$$Y_{jk} = \alpha_j + \sum_{R=1}^4 \beta_{jR} \left[ \Phi\left(\frac{w_{k+1} - \mu_{jR}}{\sigma_{jR}}\right) - \Phi\left(\frac{w_k - \mu_{jR}}{\sigma_{jR}}\right) \right] + \epsilon_{jk},$$

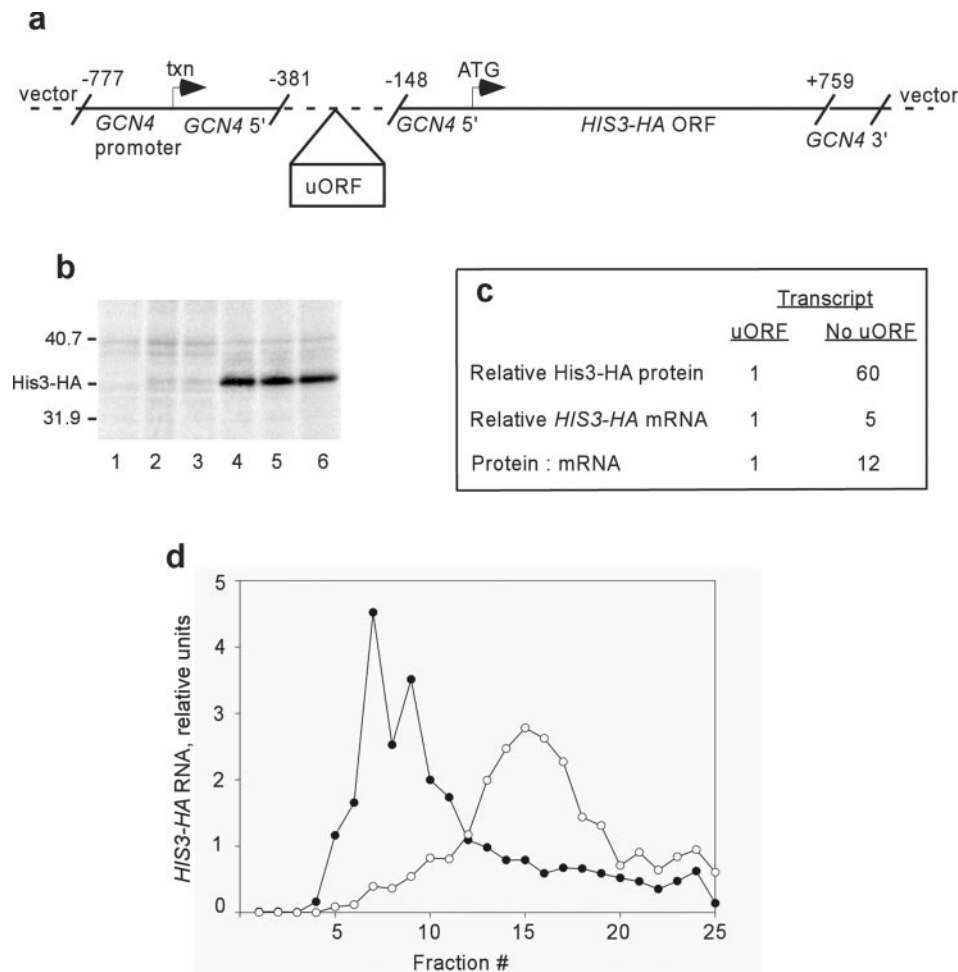
where  $\alpha_j$  is the baseline abundance measurement for the  $j$ th transcript, and  $\epsilon_{jk}$  is a random variable characterizing variations from all other sources. With the least-square technique, we can estimate all of the gene-specific parameters. Using the estimating equation theory, we can estimate standard errors for all estimated parameters.

**Comparison of Polysome Profiles Between Microarray Experiments**—From the MPM modeling, values were calculated for the percent of each transcript in polysomes and its ribosome density, as well as the relative transcript level (determined from the Cy5 signals). These values in each of the two experiments were then used to calculate the  $+/-$  ratios (plus  $\alpha$ -factor to minus  $\alpha$ -factor) for transcript level, translation efficiency and predicted protein synthetic rate for each mRNA (see Fig. 4).

**ICAT Proteome Analysis**—Simultaneously with harvesting cells for lysis and polysome analysis (see above), four 25-ml culture samples were harvested and the pellets immediately quick-frozen. Samples were stored at  $-80^\circ\text{C}$  until use. Each cell pellet was briefly washed in prechilled phosphate-buffered saline containing 1 mM phenylmethylsulfonylfluoride (Sigma, St. Louis, MO) and then transferred to a 1.7-ml Eppendorf tube. For cell lysis and protein precipitation, trichloroacetic acid (Sigma) was added to a final concentration of 10%, and the sample was incubated on ice for at least 60 min prior to centrifugation at  $14,000 \times g$  for 15 min. Resulting trichloroacetic acid precipitates were washed twice with 90% acetone prior to mild desiccation in a  $-20^\circ\text{C}$  freezer for 20 min. Dried protein samples were resuspended in isotope-coded affinity tag (ICAT) label buffer (6 M urea, 200 mM Tris, pH 8, 5 mM EDTA, 0.05% SDS). Protein concentration was measured by bichronic acid assay (Pierce, Rockford, IL), and experimental and control protein samples were each adjusted to a concentration of 2 mg/ml in a volume of 0.5 ml. Cysteiny disulfide linkages were reduced by the addition of 5 mM Tris(2-carboxyethyl)-phosphine hydrochloride (Pierce) for 60 min at  $37^\circ\text{C}$ . Following addition of 500  $\mu\text{g}$  of either isotopically light or heavy cleavable  $^{12}/^{13}\text{C}$  ICAT reagent (Applied Biosystems, Foster City, CA), the protein solutions were incubated at  $37^\circ\text{C}$  for 3 h. ICAT labeling was quenched by addition of 12 mM dithiothreitol. Heavy (experimental) and light (control) ICAT-labeled peptide solutions paired for comparison were combined and diluted 10-fold with 20 mM Tris, pH 8.3, 5 mM EDTA, and 20 ng/ $\mu\text{l}$  sequencing-grade modified trypsin (Fisher Scientific, Pittsburgh, PA) and placed in a  $37^\circ\text{C}$  water bath overnight.

ICAT-labeled peptide mixtures were fractionated by strong cation exchange via high-performance liquid chromatography (HPLC) on a polysulfioethyl A column ( $200 \times 4.6$  mm, 5  $\mu\text{m}$ , 300  $\text{\AA}$ ; PolyLC) using an Integral HPLC instrument (PerSeptive Biosystems, Foster City, CA). Fractions ( $\sim 30$ ) with highest peptide content as indicated by  $A_{214}$  measurements were dried without heat under vacuum for 30 min to remove acetonitrile. Avidin chromatography was performed using manual syringe columns (Applied Biosystems) on each selected fraction according to the manufacturer's recommendations. One-dimensional reversed-phase chromatography with on-line mass spectrometry was performed generally as described (24), but employing a 2-h binary gradient from 5 to 80% acetonitrile during which each mass spectrometry (MS) scan was followed by three MS/MS scans.

Tandem MS data were analyzed by Sequest software to determine protein identity and relative quantitation (25). Statistical robustness of peptide identifications was determined using Peptide Prophet software (26, 27), and relative quantitation of peptide pairs was further refined using ASAP software (28). Proteomic data was uploaded into



**FIG. 1. Correlation between protein synthetic rate from a transcript and ribosome loading.** *a*, diagram of the *HIS3-HA* reporter system. Plasmids pVW05 (uORF-containing) and pVW06 (no uORF) were derived from pMHY1 and pMHY3 (20, 29) by replacing the *GCN4-lacZ* ORF with the *HIS3-HA* coding sequence (the *S. cerevisiae HIS3* ORF with a 3' DNA sequence encoding a C-terminal hemagglutinin epitope). Expression is from the *GCN4* promoter and both constructs have the *GCN4* 3' UTR and a modified *GCN4* 5' leader in which 232 nucleotides of the leader are deleted and replaced with a 33-nucleotide sequence that contains the AdoMetDC uORF or the mutated form (A to G mutation of the initiating ATG codon) (29). Deletion of *UPF3* in VM1601 was necessary to prevent nonsense-mediated degradation of the mRNA, particularly with the uORF in the leader. *b*, scan of the pulse-labeled His3-HA protein after immunoprecipitation and electrophoretic separation on a polyacrylamide gel. Lane 1, Vector control; lanes 2 and 3, two independent isolates transformed with pVW05; lanes 4–6, three independent isolates transformed with pVW06. For labeling, the transformed cells were grown at 30 °C in methionine-free minimal-glucose medium with necessary supplements. Culture samples were labeled in 0.3 ml of the same medium containing 0.5–1.0 mCi <sup>35</sup>S-EasyTag EXPRE<sup>35</sup>S<sup>35</sup>S Protein Labeling Mix (Perkin-Elmer, Wellesley, MA) for 5 min at 30 °C and processed as described (49). HA-tagged protein was immunoprecipitated with mouse monoclonal antibody HA.11 (Covance Research Products, Berkeley, CA) and adsorbed to Protein G Plus-Agarose (Oncogene Research Products, San Diego, CA). Proteins in the redissolved pellet were separated by electrophoresis on a 15% polyacrylamide gel. *c*, relative His3-HA protein synthesis and mRNA levels. Labeled His3-HA protein (*b*) was quantitated with a Storm 840 phosphorimager (Amersham Biosciences). *HIS3-HA* transcript levels were determined by Q-PCR and normalized to *PRP8* and *DED1* mRNA levels (*n* = 20). *d*, transcript profiles from sucrose gradient sedimentation of yeast transformed with the two reporter transcripts. *HIS3-HA* mRNA was determined by Q-PCR (in quadruplicate, see legend to Fig. 2), normalized to one of the “utility control” RNAs added to the fractions before RNA purification (see “Experimental Procedures”), and the level in each fraction expressed as percent of total *HIS3-HA* signal across the gradient. ●, pVW05 (with the uORF); ○, pVW06 (ATG to GTG mutation of the uORF). Assignment of polysome size (number of ribosomes per mRNA) as a function of sedimentation rate in the gradient was based on the *A*<sub>254</sub> profile (see Fig. 2).

the Institute for Systems Biology gene expression platform SBEAMS (Systems Biology Expression Analysis, which can be accessed at [db.systemsbiology.net/sbeams/](http://db.systemsbiology.net/sbeams/)).

## RESULTS

**Protein Synthesis as a Function of Ribosome Density**—It is generally accepted that the synthetic rate of a specific protein

is related to the number of ribosomes actively translating its mRNA. To confirm this relationship, we employed two transcripts of different translational efficiencies dictated by the presence or absence of an inhibitory (20, 29) upstream ORF (uORF) (Fig. 1a). Presence of the uORF strongly inhibited synthesis of His3p (Fig. 1b). After correction for mRNA levels,



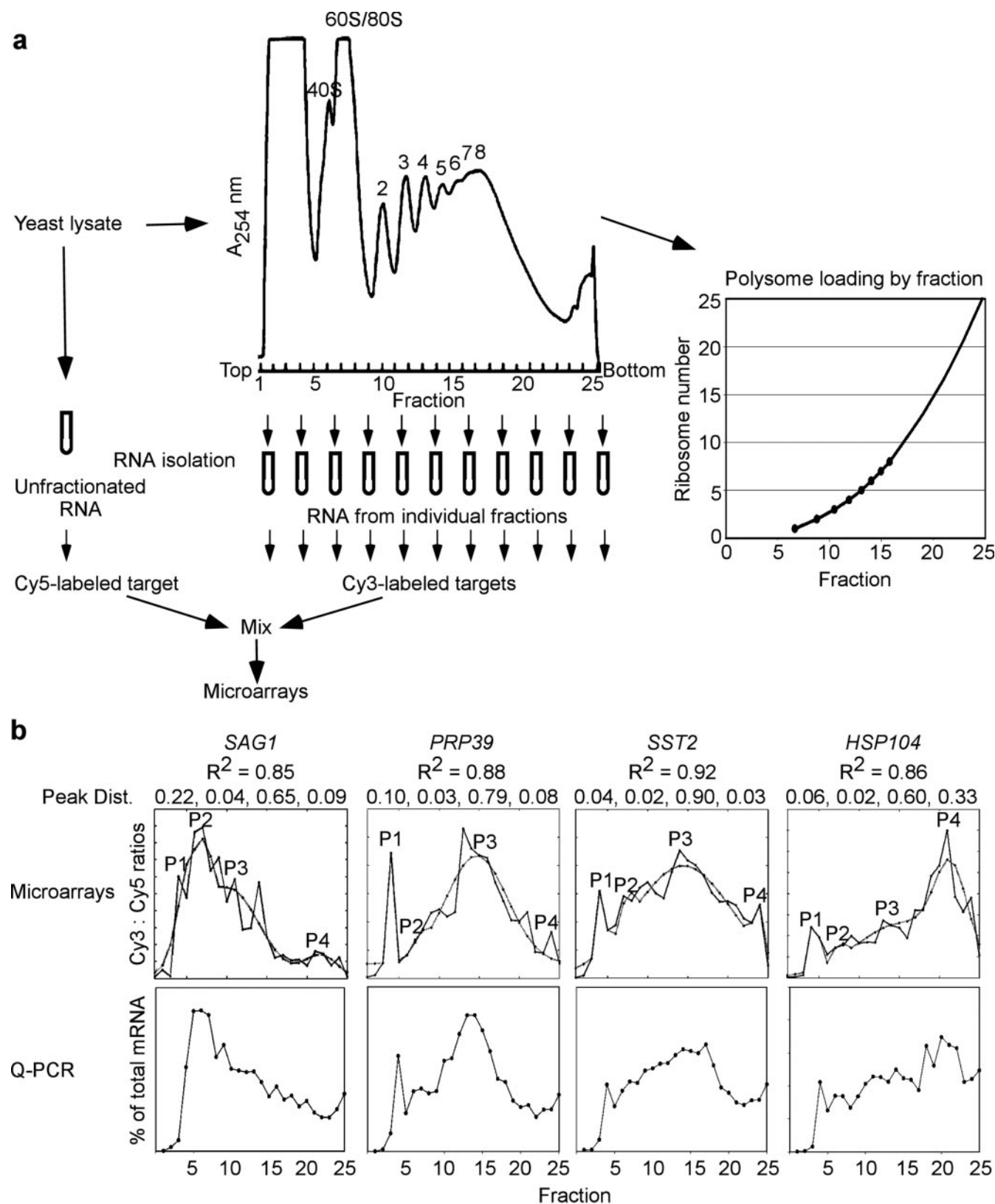


FIG. 2. Polysome profiles for individual mRNA species in gradient fractions. **a**, experimental strategy for high-resolution TSAA. Clarified lysates from steady-state growing strain BY2125 were sedimented through high-salt 7–47% sucrose gradients (21, 22) and collected into 25 fractions (see supplemental material for lysis and centrifugation conditions). Polysome profiles were found to vary by less than one-half fraction

a 12-fold difference in translational efficiency (Protein:RNA, Fig. 1c) was found between the two transcripts. Cytosol preparations from the same cultures were fractionated by velocity sedimentation through sucrose gradients. The presence of the uORF effectively inhibited ribosome loading so that a majority of the transcript contained only one ribosome or less (Fig. 1d). In the absence of the uORF, most of the mRNA was loaded with five to seven ribosomes. These results are consistent with the accepted relationship between ribosome profiles and protein synthetic rates (see "Discussion").

**Transcriptome-wide Analysis of Ribosome Loading**—Ly-sates from cultures of steady-state growing yeast were fractionated by sucrose-gradient centrifugation. The transcripts in the 25 fractions were converted to Cy3-labeled cDNA target and mixed with a constant amount of Cy5-labeled cDNA target from unfractionated RNA for hybridization to microarrays (Fig. 2a). Thus, the Cy3:Cy5 ratios generate profiles across the sucrose gradient for each specific transcript, expressed relative to its abundance in total RNA. The Cy3 and Cy5 signals were normalized to a standard curve and corrected for RNA recovery through the use of exogenously added standards; of the 6022 elements on the microarrays, 4928 had Cy5 signals within the linear range of the standard curve. The Cy3:Cy5 ratios, plotted against fraction number, were analyzed assuming the MPM (see "Experimental Procedures"). This model assumes a transcript to be distributed among four compartments: 1) mRNP particles, 2) monosomes, 3) polysomes, and 4) rapidly sedimenting material at the bottom of the gradient. These assumptions were validated by the fact that 82% (4020) of the individual profiles generated from a steady-state growing culture fit this model with  $R^2 > 0.7$ . Specific examples of the data points and fitted curves are presented in Fig. 2b. It can be seen that gene-specific variations occur in all parameters that describe the profiles, including the relative proportions in the four compartments and the rate of sedimentation of the polysome peak.

For subsequent analyses, 379 transcripts with ORFs fewer than 400 nucleotides in length were omitted, because 60% represented either dubious genes or uncharacterized small ORFs. Peak 4 was also omitted from detailed analysis of the data, because it was not clear whether this material was actually associated with ribosomes. The proportion of tran-

scripts in peak 4 averaged less than 10% across the entire transcriptome and showed no significant trend as a function of either transcript abundance or ORF length. Omitting these transcripts did not significantly affect the conclusions drawn here.

For validation, independent sucrose-gradient experiments were performed and mRNA profiles were examined gene-by-gene. Of 24 genes examined by Q-PCR or Northern blots, only one failed to validate the array results. The results from four of these genes are shown in Fig. 2b. Given that these results were from independent experiments and analyzed by quite different technologies, the strong agreement demonstrates the robustness of this approach to fractionation and analysis.

The transcriptome of steady-state growing yeast is, on average, well translated with the transcripts averaging nearly 80% association with polysomes (Fig. 3a). However, at the level of individual mRNA species, the proportion of a transcript located in peak 3 (polysomes) varies widely from gene to gene. The association of individual transcripts with polysomes ranges from 0 to 100% (Fig. 3b), with a slight tendency to decrease with the less-abundant transcripts.

The average number of ribosomes associated with a particular transcript (polysome size) is proportional to the rate of synthesis of the encoded protein. Polysome size is dictated by: 1) the length of the ORF and 2) the rate of loading of ribosomes onto an mRNA (translation initiation), relative to their rate of linear progression along the message. ORF length is available from the sequence of the yeast genome, and polysome size was derived from the dataset generated in this study. Ribosome density (number of ribosomes per 1000 nucleotides of ORF length) is quite disperse across the transcriptome (Fig. 3c), with no strong trend as a function of transcript abundance. One curious property of ribosome density, which has also been observed by others (30), is a quite significant decrease with increasing ORF length (more than 4-fold across the distribution, Fig. 3d).

We have also examined the relationship between codon adaptation index (CAI) (31) and ribosome density (supplemental Fig. 1S). Transcripts with more favorable CAI values (greater than 0.2) show a tendency toward higher ribosome densities, consistent with higher translational efficiency. Applying an index value (the AUG CAI or AUG Context Adapta-

among six gradients. The absorbance trace established the positions of mRNAs loaded with 1 to 8 ribosomes, and the positions of higher oligomers were estimated by extrapolation of a curve fit to these points. For genome-wide microarray hybridization, the poly(A)-containing RNA in each fraction was converted to Cy3-labeled cDNA target with reverse transcriptase and mixed with Cy5-labeled target made from unfractionated RNA. Hybridization to the microarrays was carried out with four replicas per fraction. *b*, *top row*, Data from microarray analysis and curves fit to the data points by the MPM (see "Experimental Procedures"). *Solid line*, Cy3:Cy5 ratios calculated from signal intensities. *Dotted line*, Curve fit to the data points. The  $R^2$  value and relative distribution ("Dist.") (derived from curve fitting) among peaks 1 through 4, respectively, are shown above the graph for each mRNA. For modeling, peak 1 (mRNP particles) was allowed to vary within fractions 1–5, peak 2 (monosomes) was fixed at fraction 7, peak 3 (polysomes) could vary in position within fractions 9–20, and peak 4 was fixed within fractions 21–25. The areas calculated beneath Gaussian curves fit to the peaks yielded the proportional distributions. Results are shown for transcripts from the following genes: *SAG1* (YJR004C), *PRP39* (YML046W), *SST2* (YLR452C), and *HSP104* (YLL026W). *Bottom row*, Determination of relative mRNA levels across a polysome sucrose gradient. Total RNA (~2  $\mu$ g) from the peak fraction in the polysome region of the gradient and the equivalent volume from all other fractions was converted to cDNA using 0.5  $\mu$ g oligo-dT<sub>25</sub> with a G/C/A 3' anchor and SuperScript II (Amersham Biosciences) and used in Q-PCR ("Experimental Procedures").

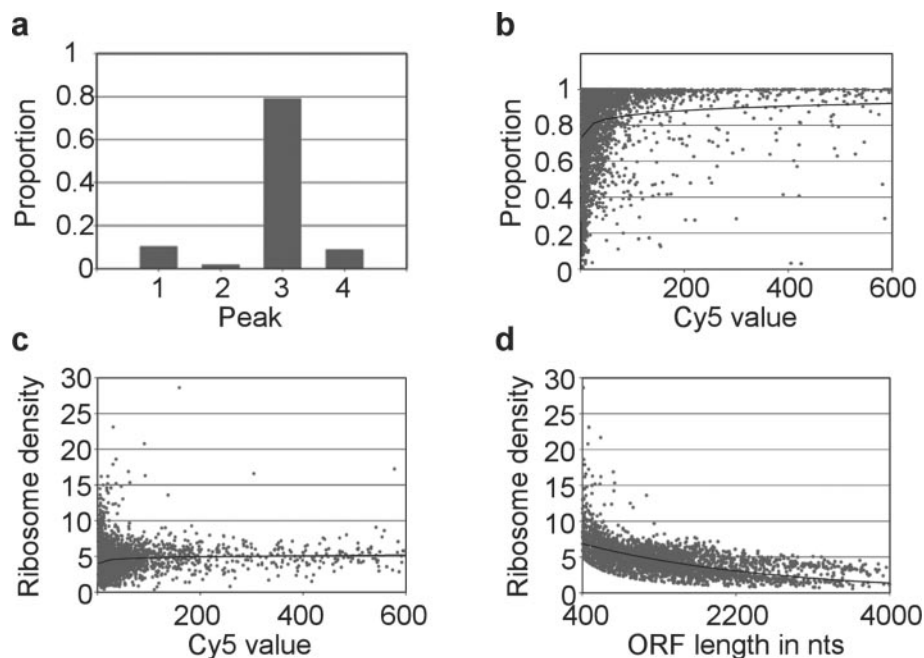


FIG. 3. **Global distribution of mRNA in steady-state growing yeast among the four compartments defined by the MPM.** *a*, values shown (normalized to 1) are the average distribution in each compartment for the 4704 mRNAs with expression values within the linear range of the standard curves on the microarrays and with  $R^2$  values  $>0.7$ , indicating good fits with the model. *b*, the proportion of each transcript in peak 3 (polysomes) plotted against relative abundance of each of the 4704 mRNAs. The x-axis has been truncated at 600 to improve the display for the majority of the data points. *c*, ribosome densities on mRNAs in peak 3 (polysomes). Data shown are for the 4066 transcripts with at least 50% mRNA in peak 3 and ORF lengths greater than 400 nucleotides (see text). Ribosome density was calculated for each mRNA as ribosomes per 1000 nucleotides of ORF length, using the average number of ribosomes per mRNA in peak 3 (obtained from modeling). The x-axis has been truncated at 600 to improve the display for the majority of the data points. *d*, ribosome density plotted against ORF length. ORF lengths were obtained from the Saccharomyces Genome Database ([genome-www.stanford.edu/Saccharomyces](http://genome-www.stanford.edu/Saccharomyces)). The x-axis has been truncated at 4000 to improve the display for the majority of the data points.

tion Index (32)) to the 12 nucleotides surrounding the AUG for all the ORFs in this database revealed no significant trend over the total compilation of data (not shown).

**Transcripts in Monosomes and mRNP Particles**—Because many of the well-established mechanisms of translational control involve sequestration of mRNAs into mRNP particles (peak 1) or monosomes (peak 2) (13, 15, 33, 34), transcripts enriched in these compartments are likely candidates for translational control. For example, the transcripts of the *GCN4* and *HAC1* genes (supplemental Fig. 2S) both have well-established mechanisms of translational regulation (35, 36), and in steady-state growing cells they are found 85 and 80%, respectively, in the combined mRNP and monosome compartments, far above the average value of 10% (Fig. 3a). The transcripts over-represented in the mRNP and monosome compartments tend to be low abundance (supplemental Fig. 3S), suggesting that translational control occurs more frequently with the less-abundant proteins. No RNA structural feature was found to correlate with appearance in this fraction, including CAI, AUG CAI, and estimated lengths of 5' and 3' untranslated regions (data not shown).

The *RPL41A* transcript has an ORF of only 78 nucleotides and occurs primarily in monosomes, with a small fraction in disomes (supplemental Fig. 2S). This represents maximum

loading of this short mRNA (37). In this case, the monosome is actively translating the encoded protein and, contrasting *RPL41A* with *GCN4*, it is apparent that caution must be exercised in interpreting monosome peaks.

**Changes in Gene Expression During the Response of Yeast to Pheromone**—Because translation state array analysis (TSAA) provides simultaneous measurements of the level and translational efficiency of any detectable transcript, it is possible to globally compare gene expression between cells with different phenotypes. As a test, we examined the changes in gene expression in yeast after acute exposure to mating pheromone. Ratios of treated to untreated cells were calculated for transcript levels, translational efficiencies, and estimated protein synthesis rates, providing a global view of changes in gene expression across the entire transcriptome (Fig. 4). Out of 3874 transcript profiles from TSAA that could be modeled with an  $R^2$  of at least 0.5 under both conditions, 3058 showed less than a 2-fold change in estimated rate of protein production (gray data points, shown without drop lines). Of 816 genes that were altered at least 2-fold in estimated protein expression, the change was driven solely by transcript level in 76% (617 transcripts, red data points on the diagonal at translation efficiency = 0.5–2.0). The remaining 24% of these transcripts showed at least a 2-fold alteration in

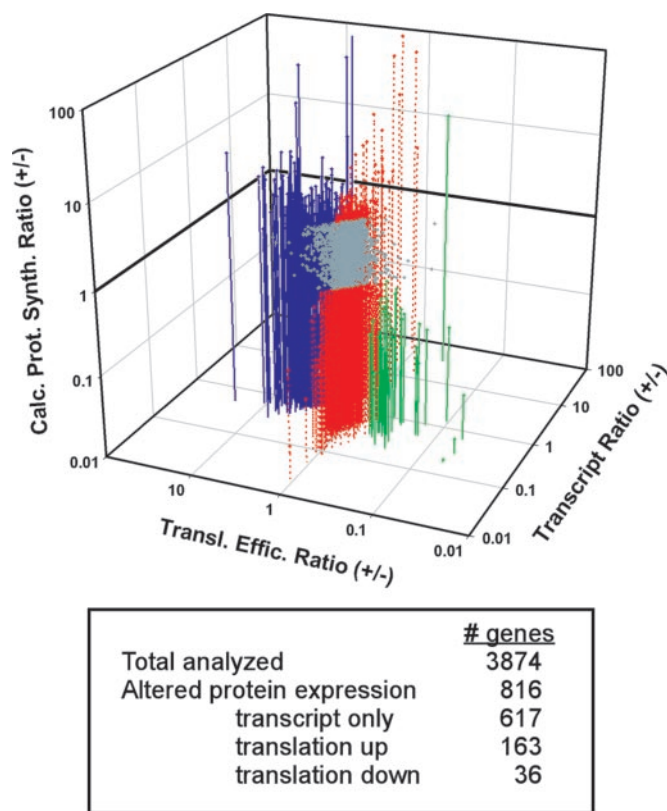


FIG. 4. **The response of the yeast transcriptome to mating pheromone as measured by TSAA.** TSAA was used to compare cells before and after 30-min treatment with  $\alpha$ -factor, as described under "Experimental Procedures." The data are presented as the ratio of plus  $\alpha$ -factor to minus  $\alpha$ -factor for those transcripts that could be modeled with  $R^2$  values  $>0.5$  and that contained ORFs  $>400$  nucleotides in length. The ratios of transcript levels (normalized Cy5 signals), translational efficiencies and calculated protein synthesis rates are plotted (see "Experimental Procedures" for the bases of these calculations). The calculated changes in protein expression were considered "biologically significant" only if the ratio of synthesis rates was less than 0.5 or greater than 2.0; data points in the "insignificant" range are colored *gray*. The *red* data points are those values where the changes in translational efficiencies are in the insignificant range. Those genes for which the ratio of translation efficiencies is  $<0.5$  or  $>2.0$  are colored *green* and *blue*, respectively.

translational efficiency; 163 were translationally up-regulated (blue data points) and 36 were down-regulated (green data points). As can be seen from Fig. 4, many of the translationally controlled genes also changed in transcript level. Customary transcript array analysis would have ignored those that were regulated solely at the translational level and would have erred quantitatively with those transcripts that showed mixed regulation (see "Discussion").

It is also of note in Fig. 4 that very few transcripts show opposing changes in transcript level and ribosome loading in response to  $\alpha$ -factor. These homodirectional changes in transcript level and translation are consistent with what was found recently in response to two other external stimuli (38).

*Functional Classification of Regulated Genes Revealed from*

*Comparing Quantitative Proteomic Analysis with TSAA*—A quantitative comparison of the proteomes of steady-state growing yeast before and after treatment with  $\alpha$ -factor was carried out using the ICAT methodology (39). From a total of 607 tagged proteins that provided reliable data for identification and quantitation, 47 were found to be up-regulated and 79 down-regulated by the criteria used in Table I. After independently clustering the results of TSAA and ICAT analysis by function, those functional groups that were significantly over-represented in both datasets are listed in Table I along with the corresponding  $p$  values.

Twenty-five genes in the "carboxylic acid transport" category were identified as decreasing in expression significantly in response to  $\alpha$ -factor and clustered with highly favorable  $p$  values (Table I). Two of these genes were identified by both TSAA and ICAT. This category was dominated by genes involved specifically in amino acid transport (19 out of the 25). The entire gene list from TSAA was searched for additional genes in amino acid transport, and six more were identified and added. Two of the genes in the general carboxylic acid transport category identified from the ICAT analysis were eliminated because they did not specifically function in amino acid transport. The TSAA and ICAT results for the complete set of amino acid transport genes are plotted in Fig. 5a. The entire category shows a strong bias toward down-regulation and, of the six additional transcripts added to the dataset based solely on functional category, only one, *AGP3*, showed a predicted increase in expression rate, while the other five followed the overall trend of this category.

A group of 23 genes, which were found to be elevated in expression by either TSAA or ICAT (four of these genes were identified in both analyses), clustered into the category of "protein catabolism" (Table I). Inspection of this group of genes showed a strong bias toward the *PRE*, *PUP*, *RPN*, and *RPT* families of genes that encode components of the proteasome. Of the 19 up-regulated protein catabolism genes discovered through TSAA, 12 were from one of these four gene families, as were six of the eight identified by ICAT (three were found by both TSAA and ICAT). The TSAA dataset was examined for additional genes from these four families, and 18 more were added, resulting in a total of 33 genes. TSAA and ICAT results for these four families of proteasome genes are plotted in Fig. 5b. Of these 33 genes, only one was significantly down-regulated (*i.e.* below a ratio of 0.75), and in this case, *RPN2*, both TSAA and ICAT showed decreased expression in response to  $\alpha$ -factor.

Not surprisingly, one of the most common groups of genes detected by TSAA comprised those in the pheromone response category (Table I). This group was not highly represented in the results of the ICAT analysis (compare with the protein catabolism category). Because many pheromone response genes have regulatory functions, it is likely that the levels of their encoded proteins are generally low, making reliable detection difficult.



TABLE I  
Functionally related clusters of regulated transcripts identified both by TSAA and ICAT analysis<sup>a</sup>

GO categories	Number of genes detected by			p values	
	TSAA	ICAT	Both	TSAA	ICAT
Expression down-regulated to 0.75 or lower					
Carboxylic acid transport	23	4	2	8.9E-08	0.0001
Vitamin/cofactor transport	3	2	2	0.009	0.004
Nuclear protein export	12	3	2	0.0002	0.009
DNA replication fidelity	9	3	1	0.02	0.001
Expression up-regulated to 1.5 or higher					
Protein catabolism	19	8	4	0.009	2.1E-06
Response to pheromone	22	2	1	5.0E-07	0.07

<sup>a</sup> Those genes whose estimated rates of protein expression from TSAA (Fig. 4) were down-regulated to <0.75 or up-regulated to >1.5 by  $\alpha$ -factor, relative to untreated cells, were analyzed (869 genes up and 1081 genes down) using the gene ontology (GO) tools available on the Saccharomyces Genome Database ([www.yeastgenome.org](http://www.yeastgenome.org)). An identical analysis was carried out with the regulated proteins revealed through ICAT analysis (49 up- and 79 down-regulated). Those genes that were significantly over-represented in both the TSAA and ICAT datasets are given above, along with the probabilities that the observed frequencies could have occurred by chance.

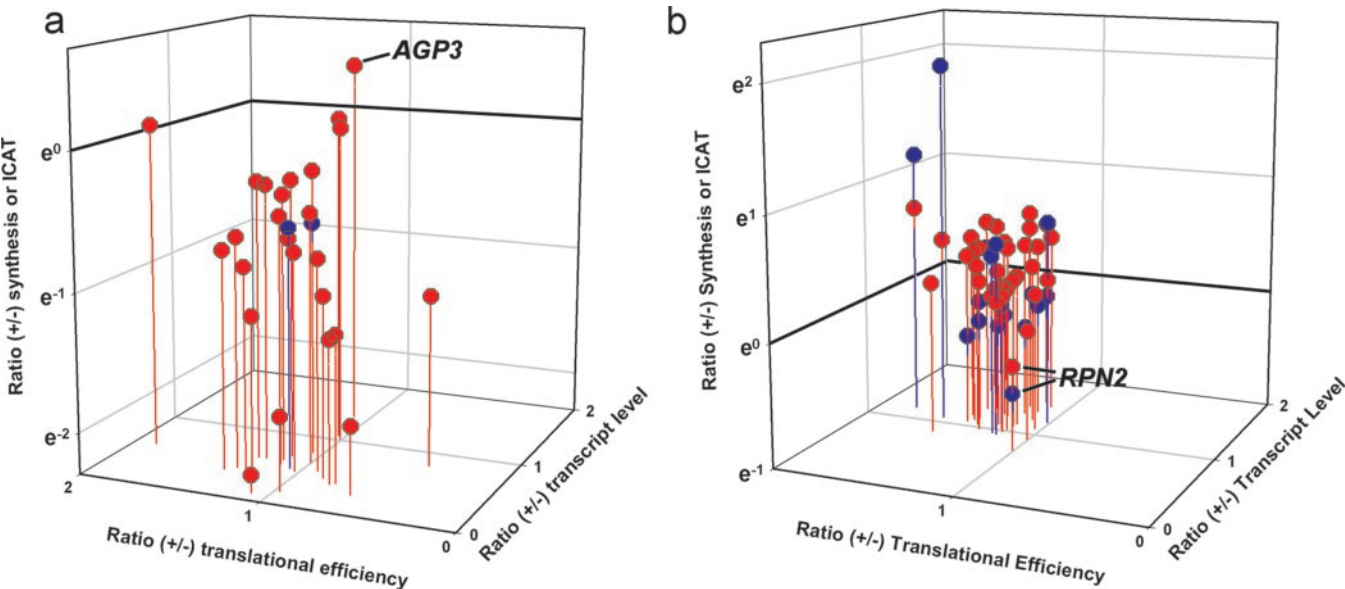


FIG. 5. **Changes in transcript level, translation, and protein level induced by  $\alpha$ -factor for genes encoding amino acid transport and proteasome components.** The ratios, plus  $\alpha$ -factor to minus  $\alpha$ -factor, obtained from TSAA (red symbols) for transcript levels, translational efficiencies, and calculated protein synthesis rates (natural logarithms) are plotted as in Fig. 4. The blue symbols are the natural logarithms of the ICAT ratios for the proteins detected by this methodology, plotted against translational efficiency and transcript level. The grid lines for synthesis ratio = 1, or  $e^0$ , are shown in bold. Genes that run counter to the general trend of each functional category in either TSAA or ICAT are labeled. a, all 25 genes in the “amino acid transport” category detected in TSAA. Two genes were detected in both TSAA and the ICAT experiment. The outlier, *AGP3*, were detected only in TSAA. b, all members of the *PRE1-10*, *PUP1-3*, *RPN1-13*, and *RPT1-6* gene families that were detected in either the TSAA or ICAT experiment (33 genes total: 16 by TSAA only, 17 by both ICAT and TSAA). *RPN2* was an outlier in both the ICAT and the TSAA data.

DISCUSSION

The composition of a proteome is the end product of regulated gene expression; the activities of its component proteins are largely responsible for defining cellular phenotype under a particular physiological state. We (16, 20) and others (17–19, 30) have developed experimental approaches to gain insight into the expression of a proteome by evaluating the association of a transcriptome with the translational apparatus. Prior studies used a somewhat arbitrary division of the resulting mRNA profiles into two pools, “poorly translated” and “well translated” transcripts, thereby losing much of the

rich information contained in the complete polysome distribution. Using recent approaches (this article and Refs. 30 and 38), which assay multiple fractions from sucrose gradients, one can define with precision, across an entire transcriptome, the proportion of transcripts actively engaged with polysomes and, within the polysome compartment itself, the number of ribosomes associated with each transcript. Furthermore, as demonstrated here, translation of the transcriptomes of two different cell populations can be compared.

The rate of synthesis of a particular protein can be expressed as the number of translationally active transcripts

times the translational efficiency (number of completed protein molecules produced per mRNA per unit time) (15). Because, with a few exceptions, the macroscopic rate of nascent peptide elongation seems to be constant across the transcriptome of a cell (15), the linear ribosome density (ribosomes per 1000 nucleotides) should provide a good comparison of rates of peptide completion between transcripts (discussed further below). To test the validity of this approach, we employed two transcripts that produced the same protein (His3p) with different translational efficiencies. As expected, the better translated transcript was loaded on average with five to seven ribosomes, while the poorly translated mRNA was located on monosomes and small polysomes (Fig. 1d). The ribosome density was calculated in each fraction, multiplied by the quantity of the transcript in the corresponding fraction and summed. From these calculations, the rate of His3p synthesis from the mRNA lacking the uORF was estimated to be ~5-fold higher than the poorly translated transcript. Given the necessary assumptions, this is a reasonable agreement with the experimental observation of a 12-fold difference in translational efficiency (Fig. 1c). These assumptions were: 1) one ribosome on each transcript containing the uORF is stalled at termination of the uORF (40) and not actively engaged in translation of His3p, and 2) with an ORF length for *HIS3-HA* of 759 nucleotides, most of these transcripts are bound by fewer than 10 ribosomes (13 ribosomes per 1000 nucleotides; see Fig. 3c). The average signal in fractions 20–25 for the uORF-containing construct (Fig. 1d) therefore provided a baseline value that was subtracted from all fractions of both gradients.

One of the striking features of the two datasets reported here is the extraordinary diversity in translation state across a transcriptome. Transcripts can vary from being localized nearly 100% in untranslated mRNP particles to essentially complete association with polysomes. In addition, the linear density of ribosomes along an mRNA was found to range from a maximum of approximately one ribosome per 30 nucleotides, which is the length of mRNA protected by a single eukaryotic ribosome from nuclease digestion (41), to less than one per 1000 nucleotides. Ribosome density is determined by the relative rates of initiation (ribosome loading) and peptide chain elongation (ribosome movement). Although translation is generally controlled at initiation (15), fluctuations in observed ribosome density for some transcripts could arise in principle from variations in elongation rate. Also, it should be underscored that these values are average densities across a transcript and that, for some mRNAs, changes either in elongation rate along the length of the message or in termination rate could result in localized regions of altered ribosome density.

This study of yeast cultures responding to  $\alpha$ -factor demonstrated several features of the TSAA methodology. First and foremost, analysis of translation state appears to be robust and reproducible from sample to sample. This is clearly illus-

trated by the fact that only a fraction of the transcripts evaluated (325 out of 3874) showed significant differences in translation state between the extracts from treated and untreated cells. Second, because independent analysis of these cells by TSAA and ICAT yielded overlapping functional groups of co-regulated genes, it seems that the assumptions that went into estimating protein synthesis rate from ribosome loading onto transcripts were warranted, at least for those proteins and transcripts that were detected by both methods. In comparing TSAA estimates of changes in rates of synthesis with proteome measurements, the ICAT ratios were usually smaller. This was expected because estimates of ribosome loading reflect the instantaneous rate of protein synthesis, while proteome measurements integrate the balance between synthesis and degradation over the course of the experiment. Therefore, except for proteins with very short half-lives, one would anticipate only qualitative agreement between the two measurements. Examples of wide discrepancy between the two approaches would suggest instances of regulated protein stability.

The combined TSAA and ICAT analysis revealed a strongly coordinated up-regulation of proteins of the proteasome, including components of both the 20S catalytic complex (the *PRE* and *PUP* genes) and the 19S regulatory complex (the *RPN* and *RPT* genes). There are several known examples of specific ubiquitination and degradation by proteasomes in *S. cerevisiae* that are related to pheromone response (42–46). Perhaps the increased synthesis of proteasome subunits detected by ICAT and TSAA represents preparation for recovery from pheromone response, through degradation of key regulatory proteins such as Far1p (44, 45) and Ste7p (46), and re-entry into the mitotic cell cycle. Progression through the cell cycle is regulated in part by ubiquitin-proteasome proteolysis (47).

An extensive study of the response of transcript levels to  $\alpha$ -factor has been published (48). With our current experimental paradigm, a robust dataset of transcript levels (Cy5 values) was created containing 100 replicas each in the presence and absence of  $\alpha$ -factor. Using 3-fold increases or decreases in transcript level as a measure of biological significance, changes were seen in 376 genes of the 5227 where the Cy5 measurement was greater than background in at least one of the two conditions. Of these 376 transcripts, 101 were identified as changing in the previous study (48). In addition, 79 of these with altered transcript levels also changed at least 2-fold in translation efficiency. Of the 3874 transcripts for which ribosome loading was analyzed in this study, 325 showed at least a 2-fold change in translation efficiency but only 44 of these also had significant (3-fold) transcript changes.

The outcome of this study is clear: each species of mRNA is unique, not only with respect to the protein it encodes, but also in its interaction with the translational machinery. Translation of the transcriptome is highly diverse both qualitatively and quantitatively, and it is impossible to assume a simple,

linear relationship between the level of an mRNA and the rate of synthesis of its encoded protein. Furthermore, although translational control seems to be quite selective, ribosome loading can change with physiological state and, together with altered protein stability, can produce dramatic discrepancies between transcript levels and apparent rates of protein synthesis.

**Acknowledgments**—We are very grateful to Roger Bumgarner and his colleagues at the University of Washington Center for Expression Array Analysis for their help and guidance in performing the hybridizations and processing the raw data. We are also deeply indebted to T. Harkin and H. Samartzidou of Amersham Biosciences for allowing us to be a beta test site for their Universal ScoreCard internal standards. Without the precision provided by these standards, these experiments would not have been possible. We also express our gratitude to Kellie Plow for performing many of the Q-PCR confirmations of the array results.

\* This work was supported by grants from the National Cancer Institute (CA89807) and the National Human Genome Research Institute (HG02283). 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.

§ The on-line version of this manuscript (available at <http://www.mcponline.org>) contains supplemental material. The original data from this work are deposited at [tsaa.info](http://tsaa.info).

|| To whom correspondence should be addressed: Department of Biochemistry, University of Washington, Seattle, WA 98195. Tel.: 206-543-1694; Fax: 206-543-4822; E-mail: [dmorris@u.washington.edu](mailto:dmorris@u.washington.edu).

## REFERENCES

- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297
- Cho, R. J., Campbell, M. J., Winzler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrieli, A. E., Landsman, D., Lockhart, D. J., and Davis, R. W. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73
- Ideker, T., Thorsson, V., Rani, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R. E., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934
- Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., and Brown, P. O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257
- Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S., and Young, R. A. (2001) Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **12**, 323–337
- Diehn, M., Eisen, M. B., Botstein, D., and Brown, P. O. (2000) Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. *Nat. Genet.* **25**, 58–62
- Brown, V., Jin, P., Ceman, S., Darnell, J. C., O’Donnell, W. T., Tenenbaum, S. A., Jin, X. K., Feng, Y., Wilkinson, K. D., Keene, J. D., Darnell, R. B., and Warren, S. T. (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**, 477–487
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. G., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L. M., Marti, G. E., Moore, T., Hudson, J., Lu, L. S., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511
- Gygi, S. P., Rochon, Y., Franz, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730
- Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., and Garrels, J. I. (1999) A sampling of the yeast proteome. *Mol. Cell. Biol.* **19**, 7357–7368
- Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, R. R., Plouffe, D., Deciu, C., Winzler, E., and Yates, J. R., III (2003) Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3107–3112
- Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003) Comparing protein abundance and mRNA expression on a genomic scale. *Genome Biol.* **4**, 117.1–117.8
- Morris, D. R. (1997) *Cis*-acting mRNA structures in gene-specific translation control, in *RNA Metabolism and Post-transcriptional Gene Regulation* (Harford, J. B., and Morris, D. R., eds) Wiley-Liss, Inc., New York
- Ruan, H. J., Brown, C. Y., and Morris, D. R. (1997) Analysis of ribosome loading onto mRNA species: Implications for translational control, in *Analysis of mRNA Formation and Function* (Richter, J. D., ed.) Academic Press, New York
- Mathews, M. B., Sonenberg, N., and Hershey, J. W. B. (1996) Origins and targets of translational control, in *Translational Control* (Hershey, J. W. B., Mathews, M. B., and Sonenberg, N., ed) Cold Spring Harbor Press, Cold Spring Harbor, NY
- Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: The second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10632–10636
- Mikulits, W., Pradet, B. B., Habermann, B., Beug, H., Garcia, S. J., and Mullner, E. W. (2000) Isolation of translationally controlled mRNAs by differential screening. *FASEB J.* **14**, 1641–1652
- Kuhn, K. M., DeRisi, J. L., Brown, P. O., and Sarnow, P. (2001) Global and specific translational regulation in the genomic response of *Saccharomyces cerevisiae* to a rapid transfer from a fermentable to a nonfermentable carbon source. *Mol. Cell. Biol.* **21**, 916–927
- Grolleau, A., Bowman, J., Pradet-Balade, B., Puravs, E., Hanash, S., Garcia-Sanz, J. A., and Beretta, L. (2002) Global and specific translational control by rapamycin in T cells uncovered by microarrays and proteomics. *J. Biol. Chem.* **277**, 22175–22184
- Serikawa, K. A., Xu, X. L., MacKay, V. L., Law, G. L., Zong, Q., Zhao, L. P., Bumgarner, R., and Morris, D. R. (2003) The transcriptome and its translation during recovery from cell-cycle arrest in *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **2**, 191–204
- Martin, T. E., and Hartwell, L. H. (1970) Resistance of active yeast ribosomes to dissociation by KCl. *J. Biol. Chem.* **245**, 1504–1508
- Martin, T. E. (1973) A simple general method to determine the proportion of active ribosomes in eukaryotic cells. *Exptl. Cell Res.* **80**, 496–498
- Fujita, H. 1975. *Foundations of Ultracentrifugal Analysis*. John Wiley & Sons, New York
- Yi, E. C., Lee, H., Aebersold, R., and Goodlett, D. R. (2003) A microcapillary trap cartridge-microcapillary high-performance liquid chromatography electrospray ionization emitter device capable of peptide tandem mass spectrometry at the attomole level on an ion trap mass spectrometer with automated routine operation. *Rapid Commun. Mass Spectrom.* **17**, 2093–2098
- Eng, J., McCormack, A. L., and Yates, J. R. 3. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989
- Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392
- Von Haller, P. D., Yi, E., Donohoe, S., Vaughn, K., Keller, A., Nesvizhskii, A. I., Eng, J., Li, X. J., Goodlett, D. R., Aebersold, R., and Watts, J. D. (2003) The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: I. Statistically annotated datasets for peptide sequences and proteins identified via the application of ICAT and tandem mass spectrometry to proteins copurifying with T cell lipid rafts. *Mol. Cell. Proteomics* **2**, 426–427

28. Li, X. J., Zhang, H., Ranish, J. A., and Aebersold, R. (2003) Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal. Chem.* **75**, 6648–6657
29. Mize, G. J., Ruan, H. J., Low, J. J., and Morris, D. R. (1998) The inhibitory upstream open reading frame from mammalian S-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. *J. Biol. Chem.* **273**, 32500–32505
30. Arava, Y., Wang, Y., Storey, J. D., Liu, C. L., Brown, P. O., and Herschlag, D. (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3889–3894
31. Bennetzen, J. L., and Hall, B. D. (1982) Codon selection in yeast. *J. Biol. Chem.* **257**, 3026–3031
32. Miyasaka, H. (1999) The positive relationship between codon usage bias and translation initiation AUG context in *Saccharomyces cerevisiae*. *Yeast* **15**, 633–637
33. Morris, D. R., and Geballe, A. P. (2000) Upstream open reading frames as regulators of mRNA translation. *Mol. Cell. Biol.* **20**, 8635–8642
34. Dreyfuss, G., Hentze, M., and Lamond, A. I. (1996) From transcript to protein. *Cell* **85**, 963–972
35. Gaba, A., Wang, Z., Krishnamoorthy, T., Hinnebusch, A. G., and Sachs, M. S. (2001) Physical evidence for distinct mechanisms of translational control by upstream open reading frames. *EMBO J.* **20**, 6453–6463
36. Ruegsegger, U., Leber, J. H., and Walter, P. (2001) Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. *Cell* **107**, 103–114
37. Yu, X., and Warner, J. R. (2001) Expression of a micro-protein. *J. Biol. Chem.* **276**, 33821–33825
38. Preiss, T., Baron-Benhamou, J., Ansorge, W., and Hentze, M. W. (2003) Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock. *Nat. Struct. Biol.* **10**, 1039–1047
39. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999
40. Raney, A., Law, G. L., Mize, G. J., and Morris, D. R. (2002) Regulated translation termination at the upstream open reading frame in S-adenosylmethionine decarboxylase mRNA. *J. Biol. Chem.* **277**, 5988–5994
41. Kozak, M. (1983) Comparison of initiation of protein synthesis in prokaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47**, 1–45
42. Johnson, P. R., Swanson, R., Rakhilina, L., and Hochstrasser, M. (1998) Degradation signal masking by heterodimerization of MAT $\alpha$ 2 and MAT $\alpha$ 1 blocks their mutual destruction by the ubiquitin-proteasome pathway. *Cell* **94**, 217–227
43. Takeuchi, J., and Toh-e, A. (2001) Genetic dissection of the yeast 26S proteasome: Cell cycle defects caused by the  $\Delta$ prp9 mutation. *Biochimie* **83**, 333–340
44. Henchoz, S., Chi, Y., Catarin, B., Herskowitz, I., Deshaies, R. J., and Peter, M. (1997) Phosphorylation- and ubiquitin-dependent degradation of the cyclin-dependent kinase inhibitor Far1p in budding yeast. *Genes Dev.* **11**, 3046–3060
45. Blondel, M., Galan, J. M., Chi, Y., Lafourcade, C., Longaretti, C., Deshaies, R. J., and Peter, M. (2000) Nuclear-specific degradation of Far1 is controlled by the localization of the F-box protein Cdc4. *EMBO J.* **19**, 6085–6097
46. Wang, Y., and Dohlman, H. G. (2002) Pheromone-dependent ubiquitination of the mitogen-activated protein kinase kinase Ste7. *J. Biol. Chem.* **277**, 15766–15772
47. Glickman, M. H., and Ciechanover, A. (2002) The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. *Physiol. Rev.* **82**, 373–428
48. Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., He, Y. D., Dai, H., Walker, W. L., Hughes, T. R., Tyers, M., Boone, C., and Friend, S. H. (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287**, 873–880
49. Kornitzer, D. (2002) Monitoring protein degradation. *Methods Enzymol.* **351**, 639–647