 DOI 10.1002/pmic.200600159 *Proteomics* 2006, *6*, 6146–6157

RESEARCH ARTICLE

# Quantitative proteomic analysis of the budding yeast cell cycle using acid-cleavable isotope-coded affinity tag reagents

*Mark R. Flory*[1], *Hookeun Lee*[2], *Richard Bonneau*[3], *Parag Mallick*[4, 5], *Kyle Serikawa*[5, 6], *David R. Morris*[5, 6] *and Ruedi Aebersold*[2, 7, 8]

[1] Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, CT, USA
[2] Institute for Molecular Systems Biology, ETH Zurich, Switzerland
[3] Department of Biology, New York University, New York, NY, USA
[4] Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA
[5] Department of Hematology and Oncology, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[6] Department of Biochemistry, University of Washington, Seattle, WA, USA
[7] Faculty of Natural Sciences, University of Zurich, Zurich, Switzerland
[8] Institute for Systems Biology, Seattle, WA, USA

Quantitative profiling of proteins, the direct effectors of nearly all biological functions, will undoubtedly complement technologies for the measurement of mRNA. Systematic proteomic measurement of the cell cycle is now possible by using stable isotopic labeling with isotope-coded affinity tag reagents and software tools for high-throughput analysis of LC-MS/MS data. We provide here the first such study achieving quantitative, global proteomic measurement of a time-course gene expression experiment in a model eukaryote, the budding yeast *Saccharomyces cerevisiae*, during the cell cycle. We sampled 48% of all predicted ORFs, and provide the data, including identifications, quantitations, and statistical measures of certainty, to the community in a sortable matrix. We do not detect significant concordance in the dynamics of the system over the time-course tested between our proteomic measurements and microarray measures collected from similarly treated yeast cultures. Our proteomic dataset therefore provides a necessary and complementary measure of eukaryotic gene expression, establishes a rich database for the functional analysis of *S. cerevisiae* proteins, and will enable further development of technologies for global proteomic analysis of higher eukaryotes.

Received: February 28, 2006
Revised: July 5, 2006
Accepted: August 12, 2006

**Keywords:**
Cell cycle / Electrospray ionization-tandem mass spectrometry / Isotope-coded affinity tags / Proteome profiling / *Saccharomyces cerevisiae*

**Correspondence:** Dr. Mark R. Flory, Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, CT 06459, USA
**E-mail:** mflory@wesleyan.edu
**Fax:** +1-860-685-2141

**Abbreviations: BCA,** biochionic acid; **GO,** gene ontology; **ICAT,** isotope-coded affinity tag; **SAM,** Significance Analysis of Microarrays; **SBEAMS,** Systems Biology Experiment Analysis Management System; **SCX,** strong cation exchange; **SGD,** *Saccharomyces* genome database

## 1 Introduction

A major effort of modern biological research encompasses interpretation of information contained in sequenced genomes in terms of the structure, function, and control of biological systems and processes. Measurement of changes in mRNA transcript levels is now routinely done in high-throughput fashion using array technology [1]. Whole transcriptomes of several organisms have been analyzed comprehensively using microarrays [2, 3]. However, several

reports attempting to correlate gene expression indices under steady-state conditions suggest that transcript levels may lack significant correlation with proteins levels and may therefore not serve as a complete measure of gene expression [4–7]. A combination of transcript and protein measurements [8, 9], and perhaps also measures of translational control [10], will likely be required to describe gene expression in biological systems more comprehensively. The dynamics of systems under induced, rather than steady state, conditions have been studied in specific contexts using stable isotope tags, for example, in androgen-stimulated prostate cancer cells [11, 12]. By contrast, the major thrust of this paper is to describe proteomic expression dynamics, measured at intervals in a synchronized system over time. This is the first report to our knowledge describing such a set of measurements in a eukaryotic system.

Accurate and quantitative recording of dynamic changes in the levels of proteins, the catalysts, and effectors of essentially all biological functions has been hampered until recently by the vast complexity and dynamic range of proteomes; there also does not exist an easily synthesized complement to proteins akin to the cDNA probes used for analysis of global mRNA expression. The traditional method for determining relative protein expression levels, employing 2-D electrophoretic separation of proteins, is laborious, requires high expertise for reproducibility, misses proteins outside limited pH and size ranges, and possesses limited resolution. Moreover, analysis of extracts from the budding yeast *Saccharomyces cerevisiae* indicates 2-D gel approaches fail to identify proteins of low abundance [13].

Quantitative proteomic, MS/MS methods have recently been used to overcome these limitations for the comparative measurement of state conditions in yeast [14–16] and quantitative proteomic measurements across seven life cycle stages of the malaria parasite *Plasmodium falciparum* [17]. The ICAT (isotope-coded affinity tag) reagent family extends upon these advances by allowing for differential isotopic labeling to be performed *in vitro* on protein samples from any biological system [18]. In brief, total labeled protein samples are combined, subjected to enzymatic digestions with trypsin, separated by offline strong cation exchange, and avidin chromatography. Purified, labeled peptides are then analyzed by RP MS/MS to reveal both the relative intensity and identity of peptide pairs. In addition to broad applicability to biological systems, the ICAT approach is amenable to a high-throughput format allowing rapid analysis of complex proteomic samples. This system has been implemented to demonstrate proteomic changes related to changes in galactose metabolism in yeast [19] and response of *Halobacterium* to changes in light and oxygen [20], and to quantitatively profile fractions from myeloid leukemia cells [21], androgen-stimulated prostate cancer cells [11, 12], and interferon-stimulated liver cells [22]. Moreover, the ICAT family of reagents has continued to evolve with the addition of tags that recognize different protein chemical groups and that allow deeper penetration of proteomes. In particular, the acid-cleavable ICAT reagent is weighted with $^{13}C/^{12}C$ facilitating RP coelution of differentially labeled peptide pairs. This is an advantageous feature as it allows the mass spectrometer additional time for determination of peptide sequence without compromising measurement of ICAT-labeled peptide-pair relative abundances. In addition, the cleavable reagents allow for removal of the biotin moiety prior to MS/MS analysis providing a less bulky tag that is more amenable to mass spectrometric analysis. Given these advantages, acid-cleavable ICAT reagents afford an approximate two-fold increase in proteomic coverage *versus* the original ICAT reagents [23].

A major challenge inherent to global proteomic analysis is the validation of measurements of both peptide identification and relative abundance of peptide pairs. To facilitate rapid and accurate statistical validation of ICAT data, a suite of software tools has been developed including: PeptideProphet (statistical validation of peptide assignments) [24], ASAPRatio (simultaneously accounts for peak shape, multiple charge states, and multiple quantitative measures of the same peptide) [25], and SBEAMS (Systems Biology Experiment Analysis Management System, a relational database for storing multiple proteomic experiments). These software tools are now available at http://www.systemsbiology.org (PeptideProphet), https://www.sbeams.org/projects/sbeams/ (SBEAMS), and http://cvs.sourceforge.net/viewcvs.py/sashimi/ASAPRatio/ (ASAPRatio).

In this report, we describe our implementation of the ICAT method; in particular, we use new acid-cleavable ICAT reagents and several software tools for statistical validation to quantitatively profile the *S. cerevisiae* proteome during synchronous transit of a full cell cycle. Five timepoint samples were collected at 30-min intervals; analysis of these timepoint samples collectively quantified the largest number of yeast proteins to date by quantitative MS, enabling the first kinetic, global, quantitative analysis of a eukaryotic proteome by MS. We display selected datapoints that were sampled quantitatively at every timepoint and that demonstrated up-regulation at one specific timepoint. We find little overall correlation between our global protein expression data and publicly available transcription data, emphasizing the critical need for proteomic measures of gene expression. Our dataset of reproducible, quantitative proteomic measurements complement similar measurements of mRNA from microarrays and should provide a rich dataset for the development of computational tools facilitating deeper proteomic sampling.

## 2 Materials and methods

### 2.1 Strains and culture conditions

Yeast strain growth, α-factor synchronization, release and collection were performed as described previously [26], and cell aliquots for this study and for analysis of translation state

 

[26] were in fact collected simultaneously from the same culture. For proteomic analysis, each 25-mL time-course sample and 25-mL sample from a culture growing asynchronously was poured into a 50 mL tube containing ~5 mL crushed ice. The tube was swirled briefly and then centrifuged for 2 min at 3000 × *g* to pellet the cells. Pellets were then rapidly frozen in a dry ice bath and stored at −80°C.

### 2.2 Verification of cell cycle synchrony

Samples were simultaneously collected for bud counts and total RNA isolation, and budding index quantitation and CLN2 Northern blotting were used to confirm synchrony as previously described [26].

### 2.3 Protein preparation, ICAT labeling, and offline chromatography

Yeast cell pellets were thawed on ice and washed quickly with ice-cold PBS containing 1 mM PMSF. Washed cell pellets were then vigorously vortexed for 1 min in 10% TCA and placed on ice for at least 1 h to allow complete protein precipitation. Whole-cell protein pellets collected by centrifugation were washed twice with ice-cold 90% acetone, and then dried at −20°C for 10 min. Dried pellets were then resuspended in ICAT lysis buffer (6 M urea, 200 mM Tris pH 8.3, 0.05% SDS, 5 mM EDTA) [18] by titration with a pipette tip and sonication in a water bath. Protein levels from timepoint samples and asynchronous reference samples were quantified using the BCA (biochionic acid) Protein Assay kit (Pierce) according to the manufacturer's recommendations with known concentrations of BSA as concentration standards. Each sample was adjusted to contain 2 mg of protein in 0.5 mL ICAT lysis buffer. To reduce disulfide linkages, tri(2-carboxyethyl)phosphine (TCEP) hydrochloride was added to a final concentration of 5 mM and samples were incubated for 60 min at 37°C. Labeling of protein with acid-cleavable $^{13}C$/$^{12}C$-based ICAT (Applied Biosystems, Foster City, CA) reagents was performed according to the manufacturer's recommendations except that the incubation time was extended to 3 h at 37°C. Samples (5 μL) from each timepoint and reference sample were collected just before and after ICAT labeling for detection of the characteristic mobility shift of ICAT-labeled separated by PAGE and detected by CBB staining. ICAT-labeled protein samples were diluted to 15 mL with 20 mM Tris, pH 8.3, 5 mM EDTA, and 20 ng/μL trypsin (Promega). Following incubation overnight at 37°C, peptide solutions were spun at 2000 g for 5 min to pellet insoluble material.

The supernatant was then loaded onto a PolyLC polysulfoethyl A (200 × 4.6 mm, 5 μm, 300 angstrom; Western Analytical, Murrieta, CA) column for strong cation exchange (SCX) in 1 mL increments using a standard autoload method on an Integral 100A HPLC instrument (PerSeptive Biosystems, Foster City, CA) operating at a flow rate of 0.5 mL/min buffer A (5 mM $KH_2PO_4$, 25% ACN, pH 2.9). Peptides were eluted at a flow rate of 0.8 mL/min using a two step binary gradient of 0–25% buffer B (600 mM KCl, 5 mM $KH_2PO_4$, 25% ACN, pH 2.9) for 0–15 min and 25–100% buffer B for 15 to 25 min, and 0.8 mL fractions were collected. Fifty SCX fractions were collected, and 35 fractions demonstrating the highest $A_{214}$ absorbance from each timepoint sample were selected for further analysis. ACN from these fractions was removed by drying under vacuum, and labeled peptides were then manually purified using avidin syringe columns and subjected to acid cleavage according to the manufacturer's recommendations (Applied Biosystems). Samples were then dried under vacuum and resuspended in ~12 μL in water containing 0.1% formic acid.

### 2.4 MS

The configuration for capillary μRPLC has been described [27]. Briefly, the system consists of a binary HPLC pump (HP1100, Agilent Technologies, Wilmington, DE), a microautosampler (Famos, Dionex LC Packings, San Francisco, CA), a precolumn (100 μm id × 2.0 cm length), and a microcapillary column (75 μm × 15 cm). Fused-silica capillary tubing with an integrated borosilicate frit (Integrafrit, New Objective, Cambridge, MA) was used for the precolumn. For the capillary column, one end of polyimide-coated fused-silica capillary (Polymicro Technologies, Phoenix, AZ) was manually pulled to a fine point ~5 μm with a microflame torch. The columns were in-house packed with C18 resin (5 μm, 200 Å Magic C18AQ, Michrom BioResources, Auburn, CA) using a pneumatic pump (Brechbuehler, Spring, TX) at constant helium gas pressure of 1500 psi.

Sample volumes of 1~6 μL were loaded onto the precolumn at a flow rate of 5 μL/min for 5 min. After sample loading and clean up, a binary solvent composition gradient with water containing 0.1% formic acid and ACN was applied to separate the peptide mixture. A linear binary gradient of 5–35% ACN at a flow rate of 200 nL/min was generated over 150 min, followed by isocratic elution at 80% ACN for 5 min. Eluting peptides from the capillary column were selected for CID by LCQ DecaXP IT mass spectrometer (ThermoElectron, San Jose, CA) using a protocol that alternated between a MS scan and four MS/MS scans. The four most abundant precursor ions in each survey scan were selected for CID. Each scan lasted an average of ~1.6 s. The specific $m/z$ value of the peptide analyzed by the MS/MS scan was excluded from reanalysis for 3 min.

### 2.5 Data analysis

For protein identification, all MS/MS spectra were analyzed using SEQUEST, a computer program that compares experimental data with theoretical spectra generated from

 

known protein sequences in the *S. cerevisiae* genome database (SGD) [28]. PeptideProphet (http://peptideprophet.sourceforge.net/) was used to estimate the accuracy of peptide assignments to MS/MS spectra made by SEQUEST [24]. Proteomic data were uploaded to the PeptideAtlas and SBEAMS (Systems Biology Experiment Analysis System) databases for proteomics data storage, sorting, and analysis, and are available *via* this platform to the public at: http://www.peptideatlas.org/repository/publications/flory2005/. Guided instructions for accessing and viewing all raw data, including CID spectra and survey scan peaks, and for accessing the SBEAMS platform to display and sort proteomic data summaries according to multiple constraints are provided as Supplementary Material. Although it is possible to sort and view data associated with any PeptideProphet cut off score, we used a cut off PeptideProphet threshold of 0.7 corresponding to a type 1 (false positive) error rate less than 5%, which applies to all (even single peptide) identifications. Accurate quantification of reconstructed ion chromatograms from ICAT-labeled peaks pairs was determined using ASAPRatio software (http://cvs.sourceforge.net/viewcvs.py/sashimi/ASAPRatio/) that facilitates automated statistical analysis of protein abundance ratios [25]. Described in detail elsewhere [25], the algorithm employs multiple statistical methods to calculate the accuracy of the abundance measurement for each identified protein whether it derives from measurement of single or multiple isotopically labeled peptide pairs. An uncertainty (error) measure for each ASAPRatio-derived ICAT value, the number of peptides measured for each protein, and other features of the quantitative data can be obtained at http://www.peptideatlas.org/repository/publications/flory2005/. To identify proteins whose abundances were most significantly correlated with cell-cycle stage, we first extracted ASAP protein quantification data from SBEAMs. Significance Analysis of Microarrays (SAM) [29] identified 300 proteins as significantly ($p$<0.00001) correlated with timepoint. The Samster package [30] was used to extract the correlated proteins and their abundances from SAM. Next, in-house software was used to discover the subset of proteins both significantly correlated with cell-cycle stage and overabundant. Seventy-six proteins were found to be specifically overabundant in a particular timepoint. To visualize all significant proteins and also those specifically overabundant, we first performed unsupervized clustering [31] on proteins. Next, we used Maple Tree (http://maple-tree.sourceforge.net/) to plot the clustering results. Correlations between ICAT peptide quantitations and microarray mRNA levels were explored using the R-statistical platform [32]. Functional annotation of yeast gene products was determined using the GO (gene ontology) database and related tools (http://www.yeastgenome.org/help/GO.html) [33, 34]. Budding yeast microarray expression data were examined using Expression Connection at http://db.yeastgenome.org/cgi-bin/expression/expressionConnection.pl. Codon bias was calculated *via* CodonW (http://www.molbiol.ox.ac.uk/cu/).

## 3 Results

### 3.1 Global proteomic measurements

Figure 1 shows the work flow for processing and analysis of the five *S. cerevisiae* global proteome samples; samples were collected at 30-min intervals following release from α-factor arrest, and asynchronous control *S. cerevisiae* protein samples were used to establish baseline expression levels (Fig. 1A). Multiple assays confirming cell synchrony, including budding index and CLN2 transcript analysis, were performed (Fig. 1A). We, in fact, used aliquots from the same culture used for analysis of cell cycle-dependent translation state [26]. Several indices were monitored to ensure representative sampling from each timepoint sample collected at 0, 30, 60, 90, and 120 min following release from α-factor arrest. First, the amount of protein analyzed for each timepoint was carefully standardized using the bichionic acid (BCA) protein assay. The BCA assay is advantageous as it tolerates urea in the sample and provides consistent quantitation of yeast whole-cell lysate samples across sample replicates (M. R. F., unpublished observation). Cell lysis was monitored visually using light microscopy. Second, ICAT labeling and trypsin digestion was assayed by gel electrophoresis. Labeling was detected qualitatively by a molecular weight bandshift (Fig. 1A) and the absence of intact protein after trypsin digestion documented the completeness of the reaction (data not shown). Third, HPLC $A_{214}$ absorbance traces were collected to ensure reproducible fractionation *via* SCX (Fig. 1B). Specifically, UV absorbance traces were collected for each timepoint, and the shapes of all the traces were nearly identical to that shown in Fig. 1B. A major peak approaching $A_{214}$ = 1.0 is always observed for the first gradient step and second peak is observed to elute during the initial part of the second gradient step. Spikes and troughs within these major peaks, also shown in Fig. 1B, are notably similar across all five timepoints, agreeing with our expectations for reproducibility predicated on the assumption that most proteins will not change dramatically in abundance across the cell cycle, and that overall bulk traces from cation exchange should therefore appear essentially identical (see comments below). For each timepoint, we collected 50 SCX fractions, and analyzed the 35 fractions showing the highest $A_{214}$ readings. Finally, the integrity of RP microcapillary LC runs was monitored by examining traces of basepeak ion intensity, or TIC, for each run (Fig. 1C) and by implementing the software tool Pep3D [35] to plot selected LC-ESI-MS data translated to the mzXML data format [36] as a 2-D density plot. The Pep3D image (Fig. 1C, right) shows one example of raw ESI-LC-MS data from this analysis plotted with each pixel representing a $MS^1$ datapoint. The intensity of each $MS^1$ measurement is indicated by corresponding pixel intensity. For example, a dark pixel indicates high signal intensity for a precursor ion with a particular $m/z$ ratio eluting at a particular point during RP separation (time). Precursor ions selected for CID ($MS^2$) are highlighted in blue.

 

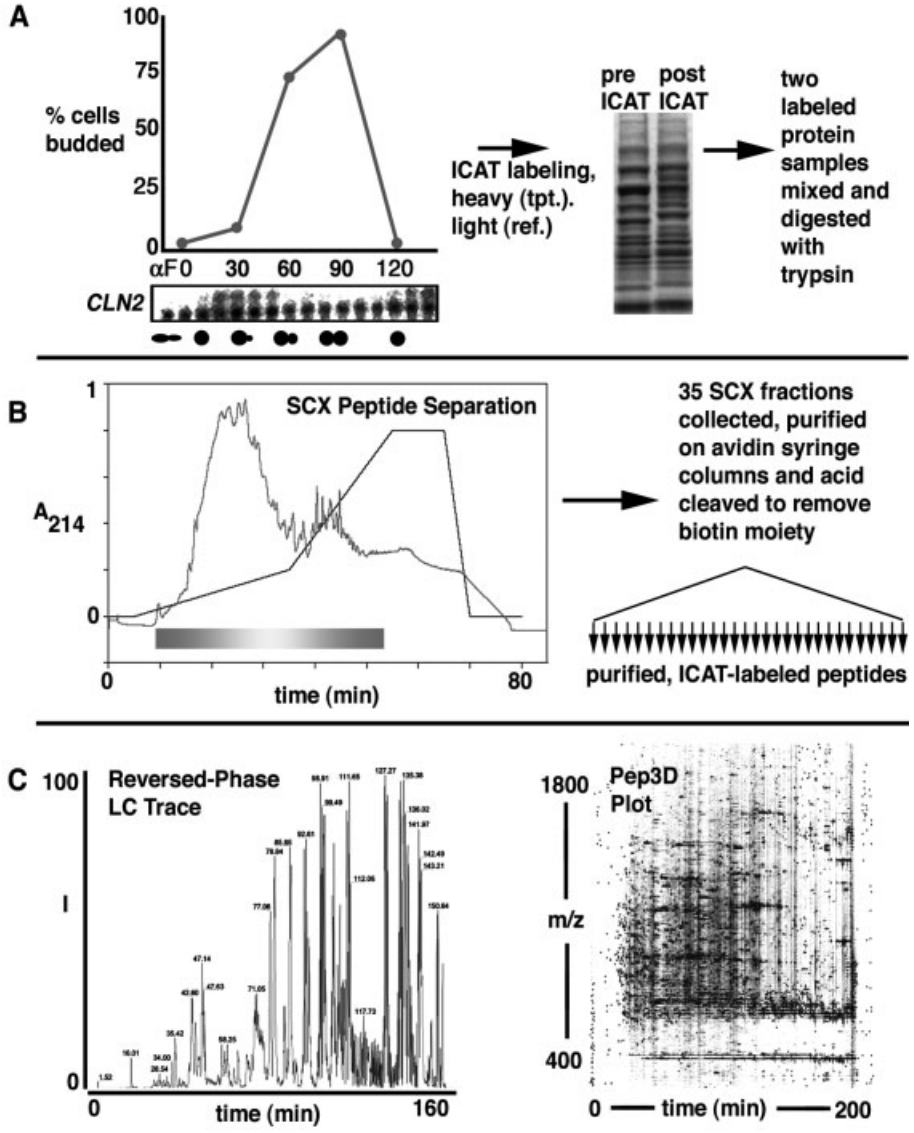


**Figure 1.** Collection and measurement of protein from *S. cerevisiae* cells synchronously transiting the cell cycle. (A) Cells were released from arrest induced by synthetic mating hormone (α-factor) and allowed to synchronously proceed through the cell cycle. Budding index and CLN2 transcript expression (upper band in Northern blot shown) were measured to confirm synchrony (left, [26]). An asynchronous reference sample was collected to provide a control protein extract for ICAT analysis. Band shifting induced by the addition of the ICAT tag analyzed by PAGE and Coomassie staining were detected prior to digestion with trypsin. (B) SCX high-pressure liquid chromatographic separation of labeled peptides from one timepoint sample (left). Yellow/orange blend, fractions collected for subsequent analysis. Avidin affinity chromatography was done on individual, selected fractions from SCX chromatography (right). (C) RP trace (left) from analysis of one avidin-purified fraction of labeled peptides. Pep3D display (right) plots ESI-MS data with intensity of all $MS^1$ features, or precursor ions, shaded in gray (intensity correlates with darkness of pixel). The position of each precursor ion on the plot is dependent on its *m/z* and time of elution during RP separation (time). Precursor ions selected for $MS^2$ CID are indicated by blue highlighting.

Both of these latter methods allow for visual inspection of raw data from the mass spectrometer prior to processing and also facilitate detection of chemical contaminants that could compromise data quality.

Figure 2 demonstrates reconstructed ion trace chromatograms from $MS^1$ survey scans demonstrating peak areas, and by extrapolation the relative quantitative amounts, of differentially ICAT-labeled peptide pairs. An example of an ICAT-labeled peptide pair is shown in Fig. 2. Quantitative information from isotopic labeling derives from a calculation of the area under selected reconstructed $MS^1$ (survey scan) peaks. It is important to note that the two peaks corresponding to each heavy ICAT ($^{13}C$) and light ICAT ($^{12}C$)-labeled peptide pair coelute during RP separation. This is in contrast to earlier versions of the ICAT tag that exhibits a staggered RP elution. Coelution of the second-generation $^{13/12}C$-labeled ICAT reagents allows the mass spectrometer to perform an increased number of peptide sequencing attempts (four *versus* one) for each ICAT-labeled peptide pair detected without compromising quantitative ($MS^1$) measurements [23]. This facet of the newer generation $^{13/12}C$-labeled ICAT reagents results in deeper coverage of complex proteomic samples as more CID spectra are generated that can be matched to theoretical spectra from the *Saccharomyces* Genome Database (SGD). An example of a CID ($MS^2$) spectrum and lists of corresponding b- and y-ion series ions detected are shown in Fig. 2.

Proteomic data were uploaded to the SBEAMS database, which facilitates sorting and analysis of proteomics data, and are available to the public at http://www.peptideatlas.org/repository/publications/flory2005/. All quantitative measurements and peptide identifications are subject to statistical

 

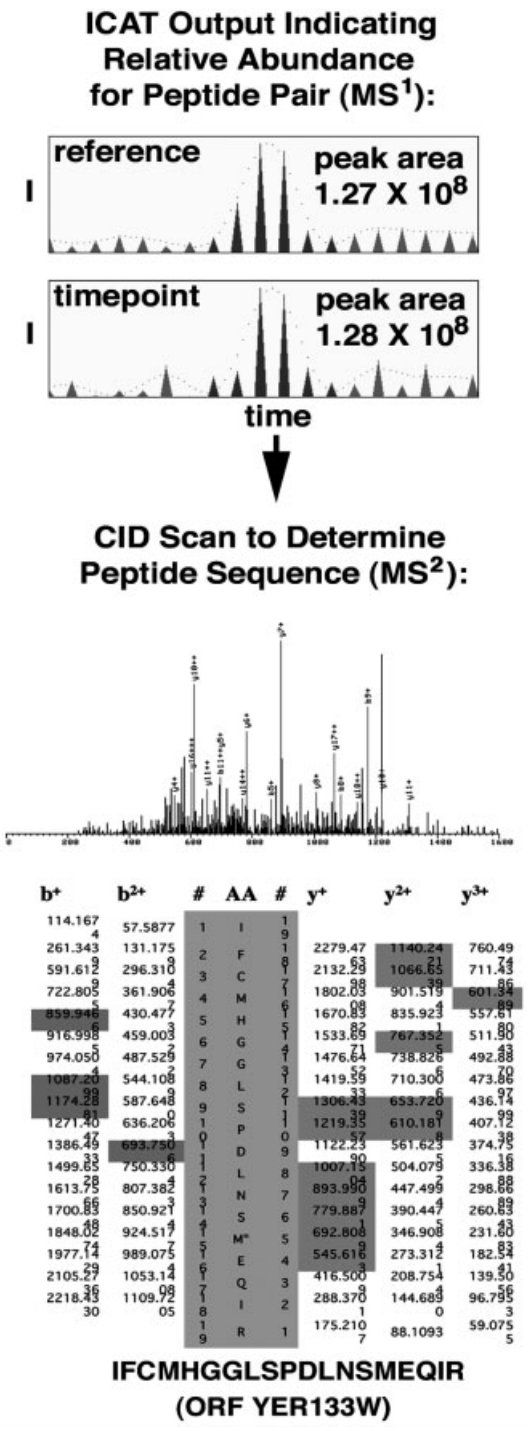


**Figure 2.** Representative output files from ICAT analysis indicating relative abundance and peptide sequence for a labeled peptide pair. Top, reconstructed ion trace chromatograms from $MS^1$ survey scan with selected area used to generate abundance measurement in red (I, intensity). Middle, experimental CID spectra from $MS^2$ analysis of selected precursor ions. Bottom, tabulation of expected b- and y-ion series members showing b-ion species (pink) and y-ion species (blue) detected. Identified peptide sequence with N-terminus at left and corresponding ORF designation are indicated.

validation and are assigned a numerical measure of uncertainty. Quantitative measurements of relative abundance were analyzed using ASAPRatio software, the details of which have been described previously [25]. ASAPRatio, which uses a combination of numerical and statistical methods to determine protein abundance ratios most effectively, obviates the need for manual inspection and interpretation of ion trace chromatograms, a job that is easily biased by personal subjectivity and that is not particularly feasible given the scale of global proteomic analyses. ASAPRatio also provides a method to effectively combine multiple measurements of the same peptide pair, and measurements of different peptide pairs mapping to one parent protein, to arrive at an overall ratio and associated error value for each identified protein. Available to the public at http://www.peptideatlas.org/repository/publications/flory2005/ are displays of all reconstructed ion trace chromatogram peak shapes, a measure of statistical uncertainty (error) for each ratio, the number of peptides used to generate each ratio, and several other values important to statistical validation of overall quantitative ICAT ratios generated for each protein.

All peptide identifications, generated when experimental $MS^2$ spectra were matched by SEQUEST [28] to theoretical peptide spectra from the SGD, were given a statistical measure of certainty (ranging from 0, representing a poor match, to 1.0, indicating a robust match) by PeptideProphet. PeptideProphet, the details of which are described elsewhere [24], employs an expectation maximization algorithm to effectively discriminate correctly from incorrectly identified peptides. This algorithm employs multiple search scores from SEQUEST and the number of tryptic termini to statistically validate both "single-hit" peptides identified once and "multiple-hit" peptides identified on multiple occasions. We selected for further analysis only those peptides identified with a PeptideProphet with a $p>0.7$, which translated to a type 1 error (false positive) rate of less than 5% for cysteine-containing peptides with at least one tryptic end. Both ASAPRatio and PeptideProphet, previously shown to be extremely valuable for high-throughput proteomic analysis of lipid raft preparations from human cells [37, 38], were essential to the proteomic studies described here.

In our analysis, we observed 2754 proteins across all timepoints, representing 48% coverage of the proteome assuming the *S. cerevisiae* proteome derives from 5726 ORFs [39]. In light of recent global gene fusion tagging analysis demonstrating that only 80% of the *S. cerevisiae* proteome is expressed under normal growth conditions [40], our coverage approaches 60% of detectable proteins. A total of 697 proteins were observed in every timepoint. Table 1 shows a matrix describing the overlap of peptides and corresponding proteins observed between each timepoint (Table 1). A total of 214 558 peptides (7382 distinct) were identified, including 1004 peptides seen in all five timepoints. Of the 9207 total (2754 distinct) protein quantifications across all fractions, approximately 75% (6451) were quantified from multiple

 

**Table 1.** Matrix indicating number of total and overlapping proteins (top number in each cell) and peptides (in parentheses) detected during ICAT-based proteomic analysis of the *S. cerevisiae* cell cycle at 0, 30, 60, 90, and 120 min following release from $\alpha$-factor arrest. ICAT-labeled peptides (and corresponding proteins) included in this matrix exhibited at least one tryptic end and a PeptideProphet score of at least 0.7, corresponding to a false-positive error rate of less than 5%.

| | 0 | 30 | 60 | 90 | 120 |
|---|---|---|---|---|---|
| 0 | 1658 (3710) | 1098 (1833) | 1186 (2305) | 1114 (1743) | 894 (1363) |
| 30 | | 1528 (3083) | 1059 (1817) | 1142 (2058) | 926 (1641) |
| 60 | | | 1455 (3056) | 1054 (1698) | 872 (1374) |
| 90 | | | | 1718 (3841) | 964 (1762) |
| 120 | | | | | 1236 (2505) |

peptide observations (Fig. 3A). Within the subset of data identified from multiple peptide observations, we observe the mean and median variance in quantification to be 20% (Fig. 3B), in keeping with previously published values for solid phase isotopic labeling [41]. There were 51 935 observations of peptides with unoxidized methionines and 27 428 oxidation of peptides with oxidized methionines. However, only 580 of these observations represent peptides in multiple oxidation states. Peptides observed in multiple observation states contributed to the quantification of only 11 of the 9207 protein quantifications. On average, 40% of peptides were observed in a single fraction. Correspondingly, the median peptide was observed in two fractions. Less than 15% of peptides were observed in more than five fractions. To assess the overlap in peptide identification between different SCX fractions, we computed: (intersection of identifications/union of identifications). The median and average peptide identification overlap between fractions were 2.4 and 5.5%, respectively (Fig. 3C). The maximum peptide identification overlap between fractions was 65%.

These data indicate that we achieved a robust, although not fully comprehensive, sampling of the budding yeast proteome. GO slim analysis indicates that our profiling is biased toward catalytic functions, metabolic processes, and cytoplasmic components. However, we also measured a broad variety of proteins with cell cycle-related functions, including several components of the core DNA replication machinery ($N = 21$), multiple proteasome components and regulators ($N = 24$), core components of the SPB (spindle pole body or yeast centrosome equivalent, $N = 4$; SPC110, TUB4, CNM67, KAR3), SPB/spindle regulators ($N = 4$; NDC1, RTS1, TPD3, DUO1), a cyclin ($N = 1$, CLB2), mitotic exit regulators ($N = 3$, NET1, NAN1, CDC14), and multiple chromatin remodeling factors ($N = 52$) (complete protein lists are available as Supplementary Microsoft Excel Workbook (Sheet 1)). Although yeast histone proteins are known to be up-regulated during S-phase [42], we did not robustly detect these proteins ($N = 2$, HHO1, SPT6). We sampled proteins from a variety of cellular compartments. For example, we measured proteins not only from the cytoplasm and nucleus, but also from smaller membrane-bound organelles such as the ER and mitochondrion (data not shown). We used the codon adaptation index (CAI) to assay the depth of sampling for rare proteins. This method assumes that less frequently used codons are more likely to be found in low-abundance proteins in a given organism. In general, negative codon usage values as determined by CodonW (as used by the SGD) represent increasingly rare proteins with increased magnitude, whereas positive values indicate proteins of higher abundance. A close match is obtained between the predicted codon bias distribution for the budding yeast proteome and that which we sampled *via* ICAT MS. As expected, we do observe a slight shift in our sampled dataset toward more abundant proteins, which tend to be more readily detected given their multiple orders of magnitude enrichment in an unbiased yeast whole-cell protein extract (data not shown). Together, these data indicate that we sampled a wide range of proteins from multiple cellular compartments, and we find, as expected, that our global profiling method biases toward detection of more abundant cellular proteins.

### 3.2 Abundance ratios associated with selected identified proteins

In an effort to identify cycling proteins, we used the software tool SAM to identify 300 proteins with significantly differentiating ICAT ratio measurements predictive of timepoint (see Supplementary Fig.). Of these 300 proteins, we selected 76 that clearly showed increased abundance at one of the five timepoints. The abundance ratios associated with these 76 proteins are shown in Fig. 4 using an output format similar to that typically used to display microarray data. Increasing red intensity indicates increased relative abundance, increasing green intensity indicates decreased relative abundance, and black indicates a ratio near one or minimal difference in abundance between the experimental (timepoint) sample and the control (asynchronous) sample. Analysis of this set of proteins indicates proteins with relatively high cellular abundances, which would be expected to comprise the group of proteins measured quantitatively in every timepoint. Functional activities associated with these proteins

 

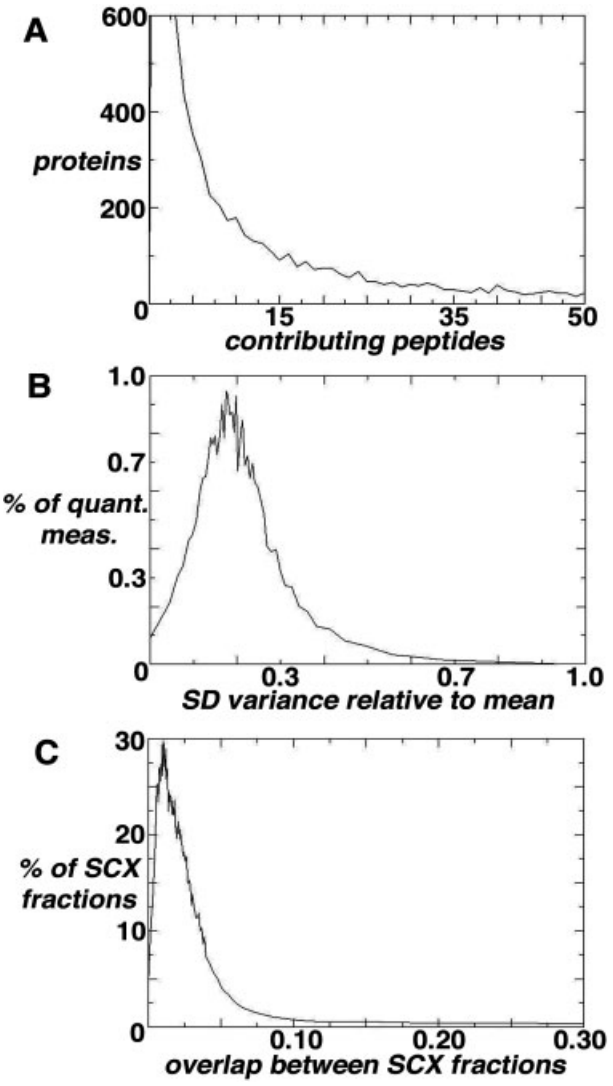


**Figure 3.** Graphical summaries of three statistical aspects of the proteomic dataset. (A) Number of peptide identifications corresponding to each identified parent protein. (B) Distribution of error measurements, expressed as SD variance from the mean, for quantitative ICAT measurements. (C) Frequency of redundant (overlapping) peptide identifications in multiple cation exchange fractions.

include those involving metabolic activities, RNA processing and transport, and ribosome biogenesis, and a small number of proteins (*e.g.*, RSC8) involved in remodeling of chromatin architecture. Lower abundance proteins, such as the cyclins, were not measured consistently across all timepoints, precluding us from globally comparing our data with other measurements of proteins' abundance in budding yeast, for example, that employing high-throughput Western blotting analysis of fusion proteins [40]. As discussed below, implementation of more effective software tools for extraction of quantitative information from our $MS^1$ data should increase the depth to which our dataset can be mined.

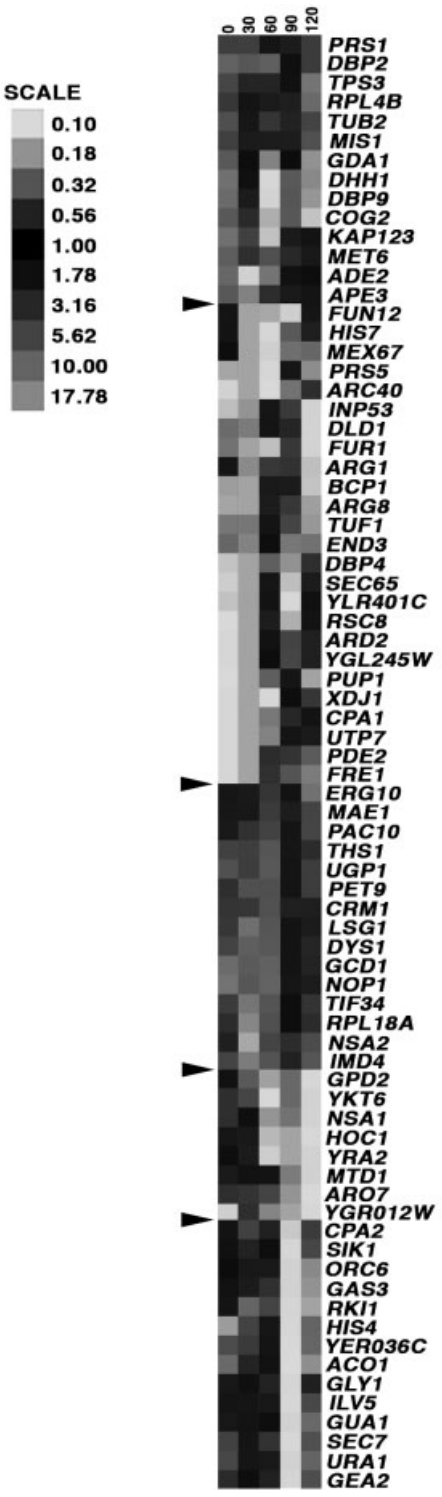


**Figure 4.** Display of 76 representative proteins identified (PeptideProphet score >0.7) with quantitative ICAT-based ratios measured in all five timepoints. Proteins are grouped from top to bottom in five groups, each of which shows increased abundance at one particular timepoint. Black arrowheads delineate the five timepoint groupings. Scale indicates correlation between color and ICAT ratio.

 

### 3.3 Comparison of proteomic data with public yeast microarray data

A fundamental question in analytical biology concerns the degree to which proteomic profiles correlate with microarray measures of transcript abundance, during steady state and for induced responses. To address this question, we compared measurements of protein levels (ICAT expression ratios) from the five timepoints from our proteomic dataset (0, 30, 60, 90, and 120 min timepoints) with the measures of mRNA abundance ratios at closely corresponding timepoints (0, 28, 63, 91, 119 min timepoints) from a microarray-based analysis of budding yeast synchronized by and released from α-factor arrest [2]. Despite many attempts using different computational strategies, we find that the correlation between our protein quantifications and the previously measured mRNA levels is insignificant, with a Pearson correlation coefficient of −0.01. Figure 5 shows the comparisons separated by timepoint. For ease of viewing, only ICAT and microarray measurements demonstrating a two-fold or greater change *versus* asynchronously growing cells are shown. We also did not detect any significant correlation when similar comparisons were made with the timepoints staggered, for example, ICAT T30 *versus* microarray T0, to account for possible time delay between transcription and translation and/or differences in the synchrony of the cell cultures (data not shown). These findings argue that, at least in the model eukaryote *S. cerevisiae*, mRNA level is not necessarily a strong predictor of protein level, and indicate that measurements of protein are essential for a complete description of gene expression.

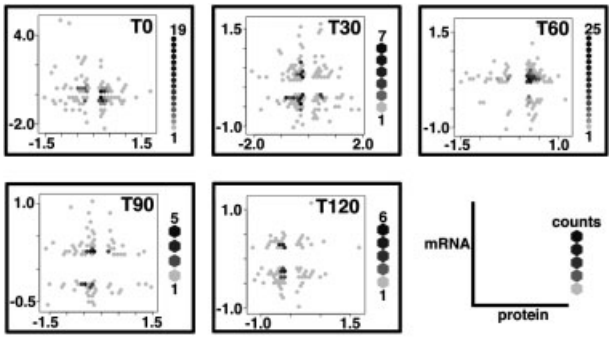


**Figure 5.** Correlation plot of log ratio ICAT measurements from analysis of *S. cerevisiae* proteomic expression (this report) and corresponding log ratio microarray measurements from similar analysis of *S. cerevisiae* transcript expression [2]. Each plot corresponds to a timepoints 0, 30, 60, 90, and 120 from our ICAT analysis, and comparisons were made to timepoints 0, 28, 63, 91, 119, respectively, from microarray experiments using similarly prepared cultures [2]. Only measurements showing at least a two-fold change are plotted. The inset at lower right indicates that protein measurements from ICAT analysis are plotted along on the *x*-axis and mRNA measurements from microarray analysis are plotted along the *y*-axis. "Counts" intensity indicates the number of correlating pairs at a given binned position (hexagons).

We also compared our 0-timepoint ICAT measurements (cells arrested in α-factor) to microarray measurements collected from cells similarly arrested in α-factor (experiment 13 from [43]). This study examined the genome-wide transcriptional profile of the budding yeast response to pheromone signaling, under a variety of experimental conditions. Among the relatively small subset of genes exhibiting a positive correlation for transcriptional and proteomic expression increases (($N = 15$) showing at least a two-fold up-regulation for both measures), we detected the several well-characterized pheromone-response genes including FIG1, FAR1, FUS1, FUS3, FUS7, KAR5, and PRM1 (Fig. 6 and Supplementary Microsoft Excel Workbook, Sheets 2 and 3). FUS1 mRNA measured originally by Northern blotting is strongly up-regulated during the budding yeast response to pheromone, and the physiological importance of Fus1p for cell fusion, during the mating response is indicated by the severe morphological defects exhibited by cells carrying a deletion of the FUS1 gene [44, 45]. While the FUS1 transcript has previously been shown to be induced up to a 100-fold, we measure here only an approximate three-fold induction of protein expression *via* our ICAT measurements. While this potentially indicates a significant discrepancy between the magnitude of FUS1 mRNA and protein up-regulation, it is likely also a result of signal dampening by baseline noise inherent to $MS^1$ measurements (see Fig. 6, baseline peaks in "reference" reconstructed ion chromatogram). Regardless, detection of Fus1p suggests that proteins whose expression is most dramatically increased under certain physiological conditions, such as the mating response, are subject to combined increases of both transcriptional and translational outputs. Like FUS1, several of the additional 15 genes showing coordinated increases in mRNA [43] and protein levels (this study) (Fig. 6) also have important roles in facilitating the mating response. Our results suggest that while this set of pheromone-responsive genes are coordinately up-regulated at the levels of mRNA and protein, the activities and/or levels of the many other proteins up-regulated during the mating response may be increased primarily *via* post-transcriptional and/or post-translational mechanisms.

## 4 Discussion

Critical to accurate analysis of gene expression is the development of robust and accurate techniques for measuring expression of proteins, the effectors of most biological processes. In this report, we use second-generation $^{13}C/^{12}C$-based, acid-cleavable ICAT reagents to quantitatively profile cell cycle-related changes in proteomic expression in a model eukaryote, the budding yeast *S. cerevisiae*. This is the first report, to our knowledge, demonstrating a series of kinetic measurements of a global eukaryotic proteome using stable isotopic labeling and MS/MS. This is in contrast to previous studies examining gene expression *via* microarray or proteomic measurements, or a combination of both measures, to

 

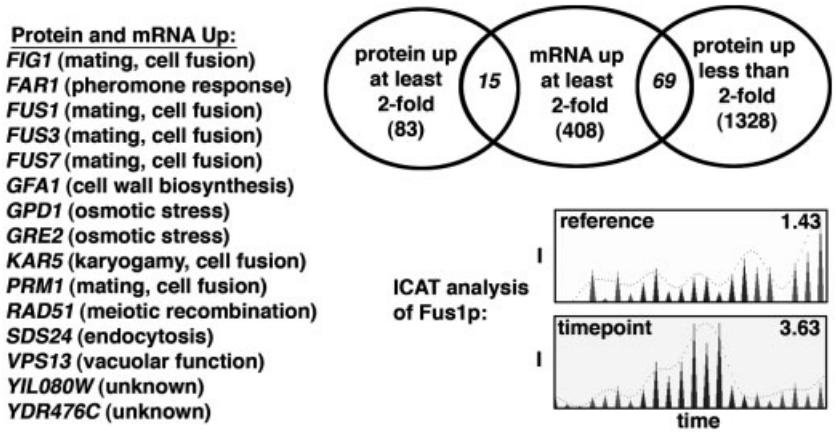


**Figure 6.** Comparative analysis of genes whose expression is up-regulated coordinately at the levels of transcription and translation in response to mating pheromone. We identified 15 genes showing at least a two-fold up-regulation of both protein (this study) and mRNA transcript (experiment 13 from [43]) levels in response to mating pheromone. Several detected genes are known to function in mating, cell fusion, and/or in the signaling response to mating pheromone. One such detected gene, FUS1, is a well-established pheromone-induced gene. We detected an approximate three-fold induction of Fus1p protein levels following treatment with mating pheromone as shown in the reconstructed ion chromatogram traces for a representative ICAT-labeled Fus1p peptide pair.

describe a state change at one point in time. Using the new ICAT reagents in combination with two dimensions of off-line chromatography and RP LC-MS/MS, we achieve the deepest quantitative proteomic sampling to date of budding yeast, namely 2754 proteins or an estimated 48% of the *S. cerevisiae* predicted proteome. With recent measurements indicating that only 80% of the *S. cerevisiae* proteome is expressed under normal growth conditions [40], our coverage may approach 60% of detectable proteins. Analysis of identified protein ontology indicates the identification of a wide range of protein classes from multiple cellular compartments.

The lack of correlation between our proteomic data and public microarray data is in agreement with previous reports indicating a lack of concordance between proteomic and transcriptional measures. One of the first examples of global discordance between transcriptional and proteomic measures involved the demonstration that most *S. cerevisiae* proteins with critical roles in the response to DNA damage are actually not induced transcriptionally in response to DNA damaging agents [46]. Three possible reasons for this lack of correlation in high-throughput expression studies have been postulated: post-transcriptional mechanisms, differences in the *in vivo* half-lives of proteins (*e.g.*, degradation rate differences), and/or error and noise in experimental measurements of mRNA and protein levels [47]. One possible post-transcriptional mechanism, translational control, has recently been investigated in a global fashion during the *S. cerevisiae* using microarrays to analyze polyribosome-enriched mRNAs that are undergoing active translation *versus* less efficiently translated mRNAs associated with monoribosomes. Unexpectedly, little translational control was evident during the cell cycle at large [26], but subsequent higher resolution measurements indicated a more prominent role for translational control in the mitogen-activated protein kinase signal pathway [48]. These data suggest that protein degradation rates may play a significant role in affecting proteomic changes across the cell cycle in complex eukaryotes, including humans, which display large diversity within the ubiquitin ligase family [49]. It should prove interesting to use quantitative proteomics to track degradation rates of proteins in higher eukaryotes in a systematic and global manner. Although we could not find a significant correlation between our measurements of protein expression and measures of mRNA expression collected under similar conditions in separate studies, simultaneous measurements of budding yeast mRNA and protein expression profiles from a single culture in the same experiment should ultimately be performed in the future. This is particularly important given that a direct correlation between transcriptional output and proteomic levels in *P. falciparum* was indeed detected in experiments simultaneously measuring mRNA and protein expression profiles in the same experiment [17]. Similar simultaneous measurement of budding yeast mRNA and protein expression levels would help to determine whether or not this discrepancy between *S. cerevisiae* and *P. falciparum* reflects a fundamental difference in gene expression regulation for these two biological systems. In addition, it is notable that our proteomic measurements and the cell cycle transcript measurements to which we compared our data [2] both derive from single biological samples. It has been well documented that biological replicates decrease the amount of background noise in microarray experiments [50, 51], and that mixtures of biological replicates can also be used to minimize background in such studies [52]. Similar approaches using replicate samples would undoubtedly also

 

reduce background noise for global, quantitative proteomic measurements *via* LC-MS/MS, as has been previously discussed [53]. Thus, future studies employing replicate biological samples to compare multiple modes of gene expression such as transcription and protein expression should complement and extend initial single-sample analyses such as those described here.

The dynamic range of proteins in yeast, estimated at six orders of magnitude [7] becomes even a more severe problem in more complex biological systems, and further technological refinements will be required to effectively survey proteomes of higher eukaryotes in a global manner. We achieved an approximate 50–60% coverage of the yeast proteome, the best effort to date, but find that only ~1000 proteins overlap between timepoints, and that fewer (697) are seen in all timepoints. This depth, while sufficient to track a significant fraction of proteins expressed at moderate or high levels, is insufficient to build complete, quantitative, molecular models of cellular behavior. Recent computational analysis of proteomic data, acquired using so-called "shotgun" sampling methods such as those employed in this study, indicate that highly iterative sampling is required to achieve saturating coverage of all peptides in complicated mixtures. For example, modeling predicts that at least ten samples are required to reach 95% coverage of protein identifications in a soluble yeast lysate [54]. It is therefore likely that we lack some sequence data from some $MS^1$ features detected during our mass survey runs, given that the IT duty cycle, although biased to allow additional CID sequencing events, is still not capable of sequencing all meaningful precursor peaks. To combat this problem, recent efforts have focused on compiling peptide identification data from diverse proteomics experiments into an annotated database that is expandable and searchable [55]. A complimentary approach to extend proteomic sampling coverage in our datasets and in future time-course analyses may involve deconvolving unsequenced peptide masses from survey scans and relating them to identical features robustly identified in other timepoints *via* $MS^2$ CID. This would allow construction of a more complete dataset with a larger number of proteins represented in multiple, or all, timepoints. We await software developments to facilitate this improvement in our coverage to the point that more sophisticated pattern analyses may be performed. As a first step toward this end, we are now working with others to develop methods to computationally match $MS^1$ features between different timepoints in our dataset [56]. We also predict that even deeper proteome sampling will be achieved using more accurate TOF MS instruments and new reagents that preclude the need for switching between MS survey mode for quantitative measurements and CID mode for sequencing of peptides. Further fractionation of peptides or proteins prior to analysis by MS may also prove effective for increasing sampling depth, and recent efforts have focused on new instrumentation to separate molecules based on p*I* as an added dimension of chromatographic fractionation. As these technologies evolve, our proteomic measurements provide not only a rich source of information for nucleating new hypotheses regarding mechanisms of the yeast cell cycle, but also a valuable dataset for the development and refinement of new tools for fully comprehensive proteomic measurements.

*We would like to thank Vivian MacKay and Lynn Law for helping with cell harvests and helpful discussions. We would also like to thank Michael Wright for assistance with strong cation exchange techniques, Eric Deutsch and Nichole King for uploading datasets, and Jimmy Eng for helping with numerous aspects of data handling and interpretation. This project has been funded in part with federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, under contract No. N01-HV-28179 and by a grant from the National Cancer Institute (1R33CA-93302). M. R. F. was supported by a Genome Sciences Post-doctoral Fellowship.*

## 5 References

[1] Lockhart, D. J., Winzeler, E. A., *Nature* 2000, *405*, 827–836.

[2] Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R. *et al.*, *Mol. Biol. Cell* 1998, *9*, 3273–3297.

[3] Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L. *et al.*, *Mol. Cell* 1998, *2*, 65–73.

[4] Kawamoto, S., Matsumoto, Y., Mizuno, K., Okubo, K., Matsubara, K., *Gene* 1996, *174*, 151–158.

[5] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, *18*, 533–537.

[6] Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., Garrels, J. I., *Mol. Cell Biol.* 1999, *19*, 7357–7368.

[7] Gygi, S. P., Rochon, Y., Franza, B. R., Aebersold, R., *Mol. Cell Biol.* 1999, *19*, 1720–1730.

[8] Hatizmanikatis, V., Lee, K. H., *Metabol. Eng.* 1999, *1*, 275–281.

[9] Hatzikmanikatis, V., Choe, L. H., Lee, K. H., *Biotechnol. Prog.* 1999, *15*, 312–318.

[10] Pradet-Balade, B., Boulme, F., Beug, H., Mullner, E. W., Garcia-Sanz, J. A., *Trends Biochem. Sci.* 2001, *26*, 225–229.

[11] Wright, M. E., Eng, J., Sherman, J., Hockenbery, D. M. *et al.*, *Genome Biol.* 2003, *5*, R4.

[12] Meehan, K. L., Sadar, M. D., *Proteomics* 2004, *4*, 1116–1134.

[13] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R., *Proc. Natl. Acad. Sci. USA* 2000, *97*, 9390–9395.

[14] Washburn, M. P., Wolters, D., Yates, J. R. III, *Nat. Biotechnol.* 2001, *19*, 242–247.

[15] Washburn, M. P., Ulaszek, R., Deciu, C., Schieltz, D. M., Yates, J. R. III, *Anal. Chem.* 2002, *74*, 1650–1657.

[16] Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., Gygi, S. P., *J. Proteome Res.* 2003, *2*, 43–50.

[17] Le Roch, K. G., Johnson, J. R., Florens, L., Zhou, Y. *et al.*, *Genome Res.* 2004, *14*, 2308–2318.

[18] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. *et al.*, *Nat. Biotechnol.* 1999, *17*, 994–999.

[19] Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R. *et al.*, *Science* 2001, *292*, 929–934.

 

[20] Baliga, N. S., Pan, M., Goo, Y. A., Yi, E. C. *et al.*, *Proc. Natl. Acad. Sci. USA* 2002, *99*, 14913–14918.

[21] Han, D. K., Eng, J., Zhou, H., Aebersold, R., *Nat. Biotechnol.* 2001, *19*, 946–951.

[22] Yan, W., Lee, H., Yi, E. C., Reiss, D. *et al.*, *Genome Biol.* 2004, *5*, R54.

[23] Yi, E. C., Li, X. J., Cooke, K., Lee, H. *et al.*, *Proteomics* 2005, *5*, 380–387.

[24] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., *Anal. Chem.* 2002, *74*, 5383–5392.

[25] Li, X. J., Zhang, H., Ranish, J. A., Aebersold, R., *Anal. Chem.* 2003, *75*, 6648–6657.

[26] Serikawa, K. A., Xu, X. L., MacKay, V. L., Law, G. L. *et al.*, *Mol. Cell. Proteomics* 2003, *2*, 191–204.

[27] Lee, H., Yi, E. C., Wen, B., Reily, T. P. *et al.*, *J. Chromatogr. B, Analyt. Technol. Biomed. Life Sci.* 2004, *803*, 101–110.

[28] Eng, J., McCormack, A. L., Yates, J. R. III, *J. Am. Soc. Mass Spectrom.* 1994, *5*, 976–989.

[29] Tusher, V. G., Tibshirani, R., Chu, G., *Proc. Natl. Acad. Sci. USA* 2001, *98*, 5116–5121.

[30] Mueller, A., O'Rourke, J., Chu, P., Kim, C. C. *et al.*, *Proc. Natl. Acad. Sci. USA* 2003, *100*, 12289–12294.

[31] Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D., *Proc. Natl. Acad. Sci. USA* 1998, *95*, 14863–14868.

[32] Ihaka, R., Gentleman, R., *J. Comput. Graphical Stat.* 1996, *5*, 314.

[33] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D. *et al.*, *Nat. Genet.* 2000, *25*, 25–29.

[34] Dwight, S. S., Harris, M. A., Dolinski, K., Ball, C. A. *et al.*, *Nucleic Acids Res.* 2002, *30*, 69–72.

[35] Li, X. J., Pedrioli, P. G., Eng, J., Martin, D. *et al.*, *Anal. Chem.* 2002, *76*, 3856–3860.

[36] Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M. *et al.*, *Nat. Biotechnol.* 2004, *22*, 1459–1466.

[37] Von Haller, P. D., Yi, E., Donohoe, S., Vaughn, K. *et al.*, *Mol. Cell Proteomics* 2003, *2*, 426–427.

[38] Von Haller, P. D., Yi, E., Donohoe, S., Vaughn, K. *et al.*, *Mol. Cell Proteomics* 2003, *2*, 428–442.

[39] Kellis, M., Patterson, N., Endrizzi, M., Birren, B., Lander, E. S., *Nature* 2003, *423*, 241–254.

[40] Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W. *et al.*, *Nature* 2003, *425*, 737–741.

[41] Bantscheff, M., Dumpelfeld, B., Kuster, B., *Rapid Commun. Mass Spectrom.* 2004, *18*, 869–876.

[42] Moll, R., Wintersberger, E., *Proc. Natl. Acad. Sci. USA* 1976, *73*, 1863–1867.

[43] Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R. *et al.*, *Science* 2000, *287*, 873–880.

[44] McCaffrey, G., Clay, F. J., Kelsay, K., Sprague, G. F., Jr., *Mol. Cell Biol.* 1987, *7*, 2680–2690.

[45] Trueheart, J., Boeke, J. D., Fink, G. R., *Mol. Cell Biol.* 1987, *7*, 2316–2328.

[46] Birrell, G. W., Brown, J. A., Wu, H. I., Giaever, G. *et al.*, *Proc. Natl. Acad. Sci. USA* 2002, *99*, 8778–8783.

[47] Greenbaum, D., Colangelo, C., Williams, K., Gerstein, M., *Genome Biol.* 2003, *4*, 117.

[48] MacKay, V. L., Li, X., Flory, M. R., Turcott, E. *et al.*, *Mol. Cell Proteomics* 2004, *3*, 478–489.

[49] Pickart, C. M., *Annu. Rev. Biochem.* 2001, *70*, 503–533.

[50] Chen, J. J., Delongchamp, R. R., Tsai, C. A., Hsueh, H. M. *et al.*, *Bioinformatics* 2004, *20*, 1436–1446.

[51] Zakharkin, S. O., Kim, K., Mehta, T., Chen, L. *et al.*, *BMC Bioinformat.* 2005, *6*, 214.

[52] Altman, N., *Appl. Bioinformat.* 2005, *4*, 33–44.

[53] Stewart, II, Zhao, L., Le Bihan, T., Larsen, B. *et al.*, *Rapid Commun. Mass Spectrom.* 2004, *18*, 1697–1710.

[54] Liu, H., Sadygov, R. G., Yates, J. R. III, *Anal. Chem.* 2004, *76*, 4193–4201.

[55] Desiere, F., Deutsch, E. W., Nesvizhskii, A. I., Mallick, P. *et al.*, *Genome Biol.* 2005, *6*, R9.

[56] Prakash, A., Mallick, P., Whiteaker, J., Zhang, H. *et al.*, *Mol. Cell. Proteomics* 2006, *5*, 423–432.

 