Published in final edited form as:

J Proteomics. 2009 February 15; 72(1): 110–120. doi:10.1016/j.jprot.2008.10.008.

A label free quantitative proteomic analysis of the *Saccharomyces cerevisiae* nucleus

Amber L. Mosley, Laurence Florens, Zhihui Wen, and Michael P. Washburn*

Stowers Institute for Medical Research, 1000 E. 50th St., Kansas City, MO 64110, United States

Abstract

To gain insight into the nuclear proteome of *Saccharomyces cerevisiae*, nuclei were isolated and fractionated via sucrose gradient sedimentation. The resulting fractions were analyzed using multidimensional protein identification technology and the detected proteins were quantified using normalized spectral counts. A large number of low abundance proteins, many of which are involved in transcriptional regulation, were recovered. Sucrose gradient elution profiles of known protein complex components demonstrated that this approach may provide insight into the question of what percentage of the total population of a protein is in one complex, versus another protein complex, or exists as a free protein.

Keywords

Multidimensional protein; identification technology; Normalized spectral; abundance factor; Quantitative proteomics; Nucleus; Transcription factors; Protein complexes

1. Introduction

Label free quantitative proteomic technologies are growing in value for the proteomics community. In one approach named spectral counting the number of total spectra, not just the unique peptides, that leads to identification of a protein is used as a quantitative measure of protein abundance [1–8]. Spectral counting is the foundation for methods proposed to determine absolute protein expression values [3,5]. Raw spectral counts are influenced by the length of a protein since longer proteins typically yield more spectral counts than shorter proteins. In order to accommodate this and to handle run to run variation in spectral counts, we developed the normalized spectral abundance factor (NSAF) which divides the spectral count (SpC) of a protein by the length (L) of a protein and then normalizes this value to the sum of all SpC/L for all proteins in an experiment [9,10]. The NSAF approach has proven valuable for label free quantitative proteomic analysis of complex mixtures [10–13] and we have recently shown that NSAF values share very similar statistical properties to transcript abundance values from Affymetrix GeneChip data [14]. Like other spectral counting base approaches [3,5,15], an important aspect of NSAF values is the ability to compare the abundance of different proteins against each other [11,13].

It remains a challenge to comprehensively analyze complex proteomes since all the proteins from a cell are not typically detected and identified in a proteomics analysis of whole cell extracts. Proteomic analyses will yield detections and identifications of the most abundant

proteins in a sample, and low abundance proteins are always challenging to detect and identify. Studies where cells are fractionated into compartments and organelles have yielded valuable datasets, especially for establishing the localization of novel proteins in higher organisms, for example [16–18]. In addition, multiple studies have used label free quantitative proteomics approaches to improve the assessment of the localization of proteins to particular cellular compartments. Other quantitative proteomic analyses of organelles include an analysis of the human nucleolus [19] and of the nuclear proteome during apoptosis [20]. The coupling of quantitative proteomics approaches and organelle analysis is yielding important biological insights.

Of great interest to many biological researchers is the content and abundance of proteins in the nucleus. In the yeast, *Saccharomyces cerevisiae*, global localization analyses have been carried out on whole cells using immunolocalization, and these studies have revealed novel proteins that localize to the yeast nucleus [21,22]. Other studies have assessed the abundance of proteins in yeast using immunodetection [23] and quantitative proteomics approaches [24]. However, relatively few proteomic studies have focused on *S. cerevisiae* nuclei. There is a large body of work analyzing protein complexes from the yeast nucleus via proteomic approaches and mechanistic analyses. These studies have been important for determining a mechanistic understanding of conserved transcriptional regulatory proteins. Insightful studies have been carried out on the *S. cerevisiae* nuclear pore, and have led to important structural insights and molecular models for this critical component of a cell [25–27]. However, the entire yeast nucleus has yet to be analyzed using state of the art proteomic tools.

In this work, nuclei from the yeast *S. cerevisiae* were isolated and fractionated by sucrose gradient sedimentation. Sucrose fractions were each independently analyzed via multidimensional protein identification technology (MudPIT) [9,28] and NSAF values were obtained to quantitatively analyze the content of each fraction. Three biological replicates of isolated *S. cerevisiae* nuclei were separated into 21 fractions, each fraction analyzed by MudPIT, and NSAF values generated. In addition, we have assessed the cofractionation of protein complex components in an approach similar to protein correlation profiling [16,29]. As a result, we have generated a quantitative analysis of the *S. cerevisiae* nucleus where the abundance and cofractionation of proteins are shown.

2. Materials and methods

2.1. Isolation of *S. cerevisiae* nuclei

Yeast (BY4741) were grown to an OD₆₀₀=1.0 in YPD. Cells were pelleted by centrifugation, washed, and then resuspended in SB buffer (40 mM HEPES, pH 7.5, 1.4 M Sorbitol, 0.5 mM MgCl₂) containing 2 mM beta-mercaptoethanol and 1 mM PMSF. The cells were then treated with zymolase at 30 °C. Cell wall digestion was monitored by microscopy. After cell wall digestion, the spheroplasts were pelleted by centrifugation and resuspended in FB buffer (20 mM PIPES, pH 6.5, 18% Ficoll 400, 0.5 mM MgCl₂). The spheroplasts were then disrupted in a Dounce homogenizer in order to release the nuclei. The homogenized FB solution was layered over GB buffer (20 mM PIPES, pH 6.5, 20% Glycerol, 0.5 mM MgCl₂) and then subjected to centrifugation at 11,500 rpm for 30 min at 4 °C in order to pellet the nuclei. The nuclei were subsequently resuspended in FB buffer and the wash was repeated three times in order to remove cytoplasmic contaminants.

2.2. Fractionation of yeast nuclei by sucrose gradient centrifugation

In order to further fractionate the nucleus to assist in protein identification, nuclei corresponding to 2.5 mg of protein were resuspended in freshly prepared PSM buffer (20 mM KP_i, pH 7.0, 1 mM MgCl₂, 250 mM sucrose, 1 mM DTT). Nuclei were then subjected to

digestion with 50 µg DNase I (Sigma, amplication grade) and 250 µg of heparin solution (10 mg/mL heparin in 50% glycerol) for 5 min at room temperature in order to lyse thenuclear membrane as well as solubilize chromatin. The digested solution (approximately 1.5 mL) was then applied to a sucrose step gradient containing 1.5 mL of 0.5 M (17%) sucrose solution, 1.5 mL of 1.0 M (34%) sucrose solution, 1.0 mL of 1.5 M (51%) sucrose solution, 1.0 mL of 1.75 M (60%) sucrose solution, 2.5 mL of a 2.0 M (68%) sucrose solution, 1.0 mL of a 2.25 M (80%) sucrose solution, and 0.5 mL of a 2.5 M (86%) sucrose solution (Fig. 1A). The sucrose solutions were prepared as previously described [30]. Following centrifugation, 0.5 mL fractions were collected from the top to the bottom of the gradient. Approximately 0.2 mL of these fractions were TCA precipitated in the presence of 0.2% Triton X-100 as a carrier. The protein pellets were washed three times in acetone and dried. Protein separation was verified by SDS-PAGE analysis of approximately 5% of the TCA precipitated sample on a 10–20% gradient gel followed by silver staining (Fig. 1B, upper panel). In addition, the DNA content of the fractions was analyzed following proteinase K digestion of 50 µL of the corresponding fraction and ethanol precipitation of the DNA. DNA was visualized through ethidium bromide staining (Fig. 1B, lower panel).

2.3. Multidimensional protein identification technology (MudPIT)

TCA-precipitated proteins were resuspended in 100 mM Tris-HCl, pH 8.5, 8 M urea, reduced with 5 mM TCEP (Tris(2-Carboxylethyl)-Phosphine Hydrochloride, Pierce), and alkylated with 10 mM IAM (Iodoacetamide, Sigma). As described in [28], a two-step digestion procedure was used. Endoproteinase Lys-C (Roche) was added to 1:100 for at least 6 h at 37 °C, then the sample was diluted to 2 M urea with 100 mM Tris-HCl, pH 8.5. Calcium chloride was added to 2 mM and the digestion with 1:100 trypsin (Promega) was incubated overnight at 37 °C while shaking. The reaction was quenched by adding formic acid to 5% and the peptide mixture was loaded onto a fused silica microcapillary column made of a 100 µm tip packed with 8 cm of reverse phase material (Aqua, Phenomenex), coupled by a filtered union with a 250 µm column packed with 3 cm of 5-µm Strong Cation Exchange material (Partisphere SCX, Whatman), followed by 1 cm of 5 µm C₁₈ reverse phase [31]. The loaded microcapillary column was placed in-line with a quaternary 1100 series HPLC pump (Agilent). Overflow tubing was used to decrease the flow rate from 0.1 mL/min to about 200–300 nL/min. Fully automated 12 step chromatography runs were carried out [9]. Three different elution buffers were used: 5% acetonitrile, 0.1% formic acid (Buffer A); 80% acetonitrile, 0.1% formic acid (Buffer B); and 0.5 M ammonium acetate, 5% acetonitrile, 0.1% formic acid (Buffer C). Peptides were sequentially eluted from the SCX resin to the reverse phase resin by increasing salt steps, followed by an organic gradient. The last two chromatography steps consisted in a high salt wash with 100% Buffer C followed by the acetonitrile gradient. The application of a 2.5 kV distal voltage electro sprayed the eluting peptides directly into a LTQ linear ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (ThermoScientific). Full MS spectra were recorded on the peptides over a 400 to 1600 *m/z* range, followed by five tandem mass (MS/MS) events sequentially generated in a data-dependent manner on the first to fifth most intense ions selected from the full MS spectrum (at 35% collision energy). Mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system (ThermoScientific).

2.4. Data analysis

SEQUEST [32] was used to match MS/MS spectra to peptides in a database of 11,982 amino acid sequences, consisting of 5815 *S. cerevisiae* proteins (non-redundant entries from NCBI 2007-03-12 release), 176 usual contaminants such as human keratins, IgGs, and proteolytic enzymes, and, to estimate false discovery rates (FDR), 5991 randomized sequences (keeping the same amino acid composition and length) for each non-redundant protein entry. No enzyme specificity was imposed during searches, setting a mass tolerance of 3 amu for precursor ions

and of 0 amu for fragment ions. Methionine oxidation was searched as a differential modification, while all cysteines were considered fully carboxamidomethylated. Results from different runs were compared and merged using DTASelect and CONTRAST [33]. Spectrum/peptide matches were only retained if peptides were at least 7 amino acids long and fully tryptic. The DeltCn had to be at least 0.08, with minimum XCorrs of 1.8 for singly-, 2.5 for doubly-, and 3.5 for triply-charged spectra, and a maximum Sp rank of 10. Finally, combining all runs, proteins had to be detected by at least two such peptides, or one peptide with two independent spectra. Proteins that were subset of others were removed.

NSAF7 (an in-house developed software) was used to create the final reports on all detected proteins across the different runs (Supplemental Tables 1, 2, 3, and 4), calculate their respective normalized spectral abundance factor (NSAF) values, and estimate false discovery rates (FDR).

Spectral FDR was calculated as:

$$\text{Spectral FDR} = \frac{2 \times \text{SHUFFLED_SpectralCounts}}{\text{Total_SpectralCounts}} \times 100.$$

Protein level FDRs were estimated as:

$$\text{Protein FDR} = \frac{\text{SHUFFLED_Proteins}}{\text{Total_Proteins}} \times 100.$$

Across the three biological replicates, the average spectral FDR was $0.22\% \pm 0.06$, while the average FDR at the protein level was $4.1\% \pm 0.8$ (Supplemental Table 1).

To estimate relative protein levels, NSAFs were calculated for each non-redundant protein, as described in [9–11]:

$$(\text{NSAF})_i = \frac{(\text{SpectralCount}/\text{Length})_i}{\sum_{k=1}^N (\text{SpectralCount}/\text{Length})_k}.$$

We implemented a new algorithm on this dataset to refine spectral counts (hence NSAF) to deal with peptides shared between multiple proteins. Therefore, for each run, dNSAFs were calculated based on distributed spectral counts, in which shared spectral counts were distributed based on spectral counts unique to each isoform [41]. Supplemental Tables 1, 2, 3, and 4 contain both spectral counts and NSAF values for all proteins. Proteins with three or fewer spectral counts should be considered of low or very low abundance. Furthermore, NSAF values are not to be considered measures of absolute abundance.

2.5. Comparison of sucrose fractionation enrichment with whole cell abundance levels

Protein abundance measurements from whole cells are available from Ghaemmamghami et al. [23]. Because NSAF and “molecules per cell” are not readily comparable, we normalized both types of values by dividing them by the median of each dataset. This procedure resulted in both whole cell abundances and NSAF values to be centered around 1 (Supplemental Table 6). To measure differences in protein levels, we then calculated the fold-change between whole cell measurements and relative protein levels after enrichment and fractionation of nuclear proteins on sucrose gradient. We considered changes of at least 3 fold in either direction to be significant (Supplemental Table 6).

3. Results and discussion

3.1. Sucrose gradient and identification of low abundance proteins

In order to determine the composition of *S. cerevisiae* nuclei and address the dynamic range problem, yeast cells were first subjected to cell wall digestion with zymolase. Following treatment, the yeast nuclei were released by Dounce homogenization and then further purified as described in the Materials and methods section. Prior to loading on an 8-mL sucrose step gradient, yeast cells were digested with DNase I and heparin. This step has previously been used to break up nuclei into sub-nuclear compartments including the nuclear membrane, nucleoplasm, and the nucleolus [30]. The nuclear lysate was then applied to sucrose gradient centrifugation in order to separate out sub-nuclear compartments as well as large protein complexes which are present in the nucleus. Following centrifugation, each fraction was analyzed by a 12-step MudPIT analysis. A total of 61 fractions were analyzed (20 from sucrose gradient one (SG1), 20 from SG2, and 21 from SG3). The information from a representative sucrose gradient is shown in Fig. 1. The sucrose gradient itself consists of a series of buffers with different concentrations of sucrose layered on top of each other (Fig. 1A). Each fraction contained proteins as shown by silver staining (Fig. 1B, upper panel) and protein detection (Fig. 1C) and the higher sucrose concentration fractions contained DNA (Fig. 1B, lower panel). When combining all three sucrose gradients, a total of 2722 proteins were detected, identified, and quantified from 467,134 spectral counts (Supplemental Table 1). Usual contaminants accounted for 21 of these proteins, while another 27 were products of retrotransposon TYA Gag and TYB polymerase genes. Over 85% of the 2674 yeast proteins were recovered from at least two out of three biological replicates demonstrating the reproducibility of the fractionation method.

To begin to assess this dataset, we compiled a database of existing *S. cerevisiae* protein abundance, functional category, and localization information from different databases. Protein abundance is given as a measurement of molecules per cell as determined by Ghaemmaghami et al. [23] and reported at <http://yeastgfp.ucsf.edu>. The protein functional categories were determined through the analysis of various *S. cerevisiae* databases including the *Saccharomyces* genome database (SGD; <http://www.yeastgenome.org/>), MIPS (<http://mips.gsf.de/>), as well as <http://yeastgfp.ucsf.edu>, which has all localization data obtained by Huh et al. in their genome wide GFP-localization screen [21]. From this database, the abundance information of 3357 proteins is available (Supplemental Table 5). When considering all these proteins, the majority have abundances between 1000 and 10,000 copies per cell (Fig. 2A). Nuclear proteins of various cellular functions, however, display a more diverse range of abundances. Proteins in the classes “nuclear proteins of unknown function” as well as “chromatin remodeling and modifying factors” in this database display abundances between 1000 and 10,000 copies per cell (Fig. 2B and C). However, of the 123 sequence-specific transcription factors in this database, the majority have abundances between 100 and 1000 copies per cell (Fig. 2D). In fact, many nuclear proteins have diverse abundances, in this database of 3357 proteins, 25 proteins have less than 100 copies per cell, 870 proteins have between 100 and 1000 copies per cell, and there are three proteins that have more than 1,000,000 copies per cell. This illustrates the challenge of the dynamic range problem in proteomics. How does one obtain mass spectrometry information on proteins from a cell that are 100,000 times less abundant than the most abundant protein in a sample? It is especially challenging to obtain information on transcription factors from cellular extracts since they are of such low abundance. One possible solution to this is to fractionate a cell into portions enriched for cellular components of interest.

We then began our assessment of the sucrose gradient dataset by using the database described above to determine what percentage of proteins in each abundance (copies per cell) range were detected and identified in this study (Table 1). In a previous study by Liu et al., nine MudPIT

analyses were carried out on a *S. cerevisiae* extract and a total of 6.3% of the proteins with less than 100 copies per cell and 15.5% of the proteins with between 100 and 1000 copies per cell were detected and identified [4]. In this study we detected and identified 24%, 24%, and 12% of the proteins with less than 100 copies per cell in each of the three sucrose gradients for a total of 32% when all runs were combined (Table 1). Furthermore, we detected and identified 31.3%, 29.9%, and 27.5% of the proteins with between 100 and 1000 copies per cell in each of the three sucrose gradients for a total of 39.9% when all runs were combined (Table 1). Included in the list of proteins that have less than 100 copies per cell is Swi1 a component of the Swi/Snf chromatin remodeling complex. We detected and identified 4 peptides and 9 spectral counts in sucrose gradient 1, 6 peptides and 14 spectral counts in sucrose gradient 2, and 6 peptides and 11 spectral counts in sucrose gradient 3 (Supplemental Table 1), leading to an averaged NSAF value of $2.1 \cdot 10^{-5} \pm 0.5 \cdot 10^{-5}$. This illustrates that although this study includes a large number of total runs, the analysis of each sucrose gradient provided good access to low abundance proteins.

3.2. Functional classification of the detected proteins

We next sought to functionally classify the proteins detected and identified in the current body of work. For example, we detected 62.8% of the 1962 proteins that are considered nuclear (Fig. 3). We detected 47% of the proteins annotated as sequence specific transcription factors, 84.9% of the chromatin remodeling factors, and 92.7% of the components of RNA polymerases I, II, or III (Fig. 3). The proteins identified in this study were sorted into localization categories and the NSAF ranges of proteins in each category determined (Fig. 4A and B). Of the 2674 proteins identified in this study, the largest portion of them, 1233, are classified as localizing to the nucleus and/or the nucleolus, followed by cytoplasmic (546), and mitochondrial (297) (Fig. 4A). However, when considering the range of abundances as estimated by NSAF values (Fig. 4B), the most abundant proteins in the dataset are cytoplasmic, which constituted 43.6% of the total abundance, followed by nuclear (34.2% of the total abundance), and mitochondrial proteins (10.3% of the total abundance). The total abundance is determined by summing all the NSAF values for all the proteins in a particular localization category or functional class. The high percentage of cytoplasmic proteins could represent proteins that are found in both the nucleus and cytoplasm, such as the three proteins that make up glyceraldehyde-3-phosphatase dehydrogenase, or could represent cytoplasmic contaminant proteins that were not removed during isolation of yeast nuclei.

The 1233 proteins annotated as localizing to the nucleus or nucleolus were separated into their respective functional categories, and the number of proteins and NSAF range for the proteins in each category was determined (Fig. 4C and D). The categories with the largest numbers of proteins were involved in transcription (430 proteins), cell cycle and DNA processing (351) (Fig. 4C). The protein synthesis category had the NSAF range with the highest median and mean from 46 proteins (Fig. 4D). However, the sum of NSAF values from the cell cycle and DNA processing (11% of total) and transcription (12% of total) categories demonstrated that these categories contributed the most to the total abundance of nuclear proteins detected. As determined by NSAF values, the three most abundant proteins localized to the nucleus were histones H2A, H4, and H2B (Supplemental Table 1). Of the remaining proteins that have been classified as non-nuclear, the category with the largest NSAF range and greatest contribution to overall NSAF abundances is the protein synthesis category (Fig. 4F), which includes a large number of ribosomal proteins (Supplemental Table 1). Although ribosomes are enriched at the ER, they are assembled within the nucleus at the nucleolus and thereby represent a significant fraction of nuclear proteins. In addition, since the ER is connected to the nuclear envelope in yeast, and is therefore difficult to remove while isolating the nucleus, ribosomal proteins constitute, not surprisingly, a large portion of this dataset. In general, the fractionation protocol

provided excellent detection and identification of nuclear proteins, which contributed 34% of the total abundance of proteins to the dataset.

TAP-intensity based quantitative information [23] was available for 2115 proteins detected in our analysis. These quantitative values are expressed as molecules per whole cell, while the NSAF values we measured are relative protein levels after nuclear enrichment and sucrose gradient fractionation. Thus we were not expecting these two measurements to correlate well, and indeed the overall Pearson correlation coefficient was a low 0.31 (Supplemental Table 6). When only proteins known to localize to the nucleus or nucleolus are used for calculation of the Pearson correlation coefficient, the correlation increases to 0.53 (Supplemental Table 6). Previous studies performed by Liu et al. compared the measured TAP intensities to raw spectral counts from MudPIT analyses of whole cell yeast lysates and found a Pearson correlation coefficient of 0.58, which is similar to what we observed for the proteins found in the nucleus [4]. When we calculated the Pearson correlation coefficient on the remaining proteins which are not known to localize to the nucleus (including cytoplasmic, mitochondrial, and ER proteins) the correlation was much lower at 0.24. The most abundant protein identified in our preparation was the translational elongation factor Teflp. Liu et al. had previously noted that the levels of Teflp observed by spectral counting were much higher than what was reported by Ghaemmaghami et al, leading the authors to the conclusion that Teflp was a potential outlier whose expression level was altered in the Ghaemmaghami et al. study due to the presence of the TAP tag [4,23]. Interestingly, if we remove the values for Teflp from our dataset, the Pearson correlation increases from 0.31 to 0.50 for the whole dataset and from 0.24 to 0.51 for the non-nuclear protein group (Supplemental Table 6), illustrating the negative effects of potential outliers on calculation of the correlation coefficient.

We next used the whole cell abundance measurements as a benchmark against which to test whether the nuclear enrichment increased nuclear protein levels. We sorted the proteins based on their known or predicted localization to the nucleus/ nucleolus (1019 proteins) as opposed to other cellular compartments (1096 proteins), and further sorted these two main groups of proteins based on the fold change observed between whole cell abundance levels and NSAF values measured after sucrose gradient fractionation (Fig. 5 and Supplemental Table 6). About 31% of the nuclear proteins were clearly enriched after the fractionation procedure compared to whole cell levels, with 76.5% of these proteins belonging to the “Transcription” or “Cell Cycle” classes. On the other hand, of the 171 nuclear proteins (16.8%) whose levels were higher in whole cell measurements, 77.2% were annotated as localizing to other cellular compartments besides the nucleus. The levels of such proteins should therefore be higher when considering whole cell measurements, which is what we observed. Finally, 18.4% of non-nuclear proteins showed higher levels after nuclear fractionation. We tested whether particular functional classes were over-represented in this group of proteins using the FatiGO+ software [40] inputting our own functional annotations (Supplemental Table 5). Proteins involved in “Protein Synthesis” constituted 13.9% of this group and were significantly over-represented when compared to a background list of proteins using a Fisher’s exact test (FDR adjusted p value: $1.05e-05$). This is in agreement with our previous observation that ribosomal proteins constitute a large portion of this dataset.

3.3. Sucrose gradient fractionation profiles of protein complexes

We assessed the dataset based on the recovery of protein complexes from our analysis compared to the known components from the database. We focused on transcriptional regulatory complexes like the Rpd3/Sin3 histone deacetylase complex [34] of which we detected all 12 components, the minichromo-some maintenance protein complex (MCM) [35] of which we detected all six components, the CCR4-NOT complex of which we detected seven of the nine components [36], and the COMPASS histone methyltransferase complex

[37] of which we detected and identified six out of eight components (Table 2). From this, we can conclude that we are recovering the majority of the known components of major protein complexes.

Next, we assessed whether known components of protein complexes co-fractionated along the sucrose gradient. To begin, prominent 'contaminants' are shown in Fig. 6 with the cytoplasmic translation elongation components shown in Fig. 6A and the cytosolic ribosomal subunits shown in Fig. 6B. The dominant cytoplasmic translation elongation protein found in nearly every fraction is the translation elongation factor 1, Teflp (Fig. 6A). The cytosolic ribosomal subunits are detected in many fractions, with a clear increase in abundance in fractions 8 through 20 (Fig. 6B) in agreement with these complexes being very large multisubunit assemblies, hence fractionating through the high-density layers of the sucrose gradient.

We next plotted the fractionation profiles of known components of nuclear protein complexes. Many of the nuclear protein complexes involved in chromatin remodeling came out in fractions 6 through 8 (data not shown). We focused on four protein complexes with interesting profiles (Fig. 6). The majority of the components of the Rpd3 histone deacetylase complex [34] co-fractionate in fractions 5–9 (Fig. 6C). Separate peaks for Rpd3p and Ume1p are found in fractions 11–12 and Pho2p was also recovered in fraction 16 (Fig. 6C). This suggests that there are pools of these proteins that are not associated with Rpd3. Their recovery in higher density fractions suggests that they may be associated with other large cellular components, and they are not likely free proteins. Next, the fractionation profile of the MCM (mini-chromosome maintenance) complex [35] is shown in Fig. 6D: all components were detected in fractions 7–8 with higher abundance and then again in lower abundance in fraction 12, where all the components were seen except MCM3, but this may be due to the lower abundance of all the components. This suggests that there could be at least two different sub-assemblies of the MCM complex, which requires further study.

Next, the fractionation profile of the CCR4-NOT complex [36] is shown in Fig. 6E. The majority of the components were detected with a various profile over the first nine fractions. A component of this complex, Pop2p also displayed prominent localization to fraction 16 along with the protein Mob1p, which has been shown to interact with Pop2 via affinity capture western blotting and yeast two hybrid assay [38], Mob1p, however, has not been shown to interact with other components of the CCR4-NOT complex, indicating that it may be involved with another complex containing Pop2p. Finally, the fractionation profile of six components of the COMPASS histone methyltransferase complex [37] is shown in Fig. 6F where the components predominantly come out over fractions 3–10, and Sdc1p/Cps25p in particular have two peaks in fractions 4 and 9, while Bre2p/Cps60p and Swd2p/Cps35p are detected in fraction 16 as well (Fig. 6F). Swd2p/Cps35p is known to be present in other complexes besides COMPASS and is a mobile subunit of COMPASS, which may explain its localization pattern across the gradient [39]. Taken together, the fractionation profiles of known protein complexes provide potential insights into the functions of particular proteins that fractionate differently than the majority of the protein complex. The distinct and limited peaks of proteins, like that of Mob1p and Pop1p, can be explained with existing literature, but the functional implications are unknown. There are two likely explanations for the separation of known protein complexes into distinct peaks. One is that the complex is breaking apart into distinct subcomplexes which could be present in lower sucrose density fractions than the intact complex. In addition, proteins in known complexes could associate with other interacting partners or high density sub-nuclear components such as the nucleolus, which would also cause enrichment in different fractions across the sucrose gradient.

4. Conclusions

In this work, we characterized an enriched nuclear fraction of *S. cerevisiae*. By attempting to isolate nuclei, we were able to enrich for low abundance proteins and were able to gain access to transcriptional regulatory proteins of great interest. We were able to detect and identify many more low abundance proteins that have been measured to be less than 100 copies per cell or between 100 and 1000 copies per cell when compared to a previous proteomic study where a yeast whole cell lysate was analyzed up to nine times [4]. Using the approach described in this study would allow for the comparative analysis of yeast nuclei under different conditions to assess the dynamics of yeast nuclei subjected to different stimuli or could be applied to fractionate sub-nuclear compartments from nuclei in higher organisms.

By fractionating isolated nuclei, we were able to analyze the fractionation profiles of known transcriptional regulatory complexes. This process was rarely perfectly clean where each component of a protein complex only came out in one or two fractions. In many cases, known protein components of complexes had multiple peaks in the sucrose gradient fractionation. However, this may have functional implications for proteins thought to be primarily in one complex or another. An interesting general question requiring further study is when a protein participates in multiple complexes, what percentage of the total population of a protein is in one complex versus another, or exists as free protein? For example, in the case of the CCR4-NOT component Pop2p, more than twice as much Pop2p was found in a fraction with Mob1p than in the fractions with the other CCR4-NOT components (Fig. 6E). On the other hand, in the case of Swd2p/Cps35p, more than 75% of this protein fractionates with the COMPASS components over several fractions, and approximately 25% comes down in fraction 16 (Fig. 6F). Here the quantitative analysis of the fractionation of proteins over a gradient begins to address the question of what relative proportion of a total protein population participates in multiprotein assemblies.

Acknowledgements

This work was supported by a postdoctoral fellowship from the NIGMS to A.L.M. (F32 GM075541) and funding from the Stowers Institute for Medical Research.

Appendix

Appendix A. Supporting information available

Six supplemental tables are provided with this study. Supplemental Table 1 contains a summary of all the data, while the detailed information for each of the three sucrose gradients reported here can be found in Supplemental Tables 2 to 4, respectively. Supplemental Tables 1, 2, 3, and 4 contain both spectral counts and NSAF values for all proteins. Proteins with three or fewer spectral counts should be considered of low or very low abundance. Furthermore, NSAF values are not to be considered measures of absolute abundance. Supplemental Table 5 contains the database compiled for use in some of the analysis of this work as described in the beginning of the Results and discussion section. Supplemental Table 6 contains the comparison of protein levels after nuclear enrichment with the data obtained from the Ghaemmaghami et al. study [23]. All supplemental tables are provided as pdf files due to space constraints but are available as Excel files upon request.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.jprot.2008.10.008.

REFERENCES

1. Blondeau F, Ritter B, Allaire PD, Wasiak S, Girard M, Hussain NK, et al. Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling. *Proc Natl Acad Sci U S A* 2004;101:3833–3838. [PubMed: 15007177]
2. Fu X, Gharib SA, Green PS, Aitken ML, Frazer DA, Park DR, et al. Spectral index for assessment of differential protein expression in shotgun proteomics. *J Proteome Res* 2008;7:845–854. [PubMed: 18198819]
3. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 2005;4:1265–1272. [PubMed: 15958392]
4. Liu H, Sadygov RG, Yates JR III. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 2004;76:4193–4201. [PubMed: 15253663]
5. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 2007;25:117–124. [PubMed: 17187058]
6. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 2005;4:1487–1502. [PubMed: 15979981]
7. Zhang B, VerBerkmoes NC, Langston MA, Uberbacher E, Hettich RL, Samatova NF. Detecting differential and correlated protein expression in label-free shotgun proteomics. *J Proteome Res* 2006;5:2909–2918. [PubMed: 17081042]
8. Zybailov B, Coleman MK, Florens L, Washburn MP. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal Chem* 2005;77:6218–6224. [PubMed: 16194081]
9. Florens L, Garozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL, et al. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* 2006;40:303–311. [PubMed: 17101441](San Diego, Calif).
10. Zybailov B, Mosley AL, Sardi ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* 2006;5:2339–2347. [PubMed: 16944946]
11. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, et al. Quantitative proteomic analysis of distinct mammalian mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci U S A* 2006;103:18928–18933. [PubMed: 17138671]
12. Sardi ME, Cai Y, Jin J, Swanson SK, Conaway RC, Conaway JW, et al. Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc Natl Acad Sci U S A* 2008;105:1454–1459. [PubMed: 18218781]
13. Zybailov BL, Florens L, Washburn MP. Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol Biosyst* 2007;3:354–360. [PubMed: 17460794]
14. Pavelka N, Fournier ML, Swanson SK, Pelizzola M, Ricciardi-Castagnoli P, Florens L, et al. Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol Cell Proteomics* 2008;7:631–644. [PubMed: 18029349]
15. Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, et al. Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* 2008;9:102. [PubMed: 18304323]
16. Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, et al. A mammalian organelle map by protein correlation profiling. *Cell* 2006;125:187–199. [PubMed: 16615899]
17. Gilchrist A, Au CE, Hiding J, Bell AW, Fernandez-Rodriguez J, Lesimple S, et al. Quantitative proteomics analysis of the secretory pathway. *Cell* 2006;127:1265–1281. [PubMed: 17174899]
18. Kislinger T, Cox B, Kannan A, Chung C, Hu P, Ignatchenko A, et al. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* 2006;125:173–186. [PubMed: 16615898]
19. Andersen JS, Lam YW, Leung AK, Ong SE, Lyon CE, Lamond AI, et al. Nucleolar proteome dynamics. *Nature* 2005;433:77–83. [PubMed: 15635413]

20. Hwang SI, Lundgren DH, Mayya V, Rezaul K, Cowan AE, Eng JK, et al. Systematic characterization of nuclear proteome during apoptosis: a quantitative proteomic study by differential extraction and stable isotope labeling. *Mol Cell Proteomics* 2006;5:1131–1145. [PubMed: 16540461]
21. Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, et al. Global analysis of protein localization in budding yeast. *Nature* 2003;425:686–691. [PubMed: 14562095]
22. Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, et al. Subcellular localization of the yeast proteome. *Genes Dev* 2002;16:707–719. [PubMed: 11914276]
23. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, et al. Global analysis of protein expression in yeast. *Nature* 2003;425:737–741. [PubMed: 14562106]
24. de Godoy LM, Olsen JV, de Souza GA, Li G, Mortensen P, Mann M. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol* 2006;7:R50. [PubMed: 16784548]
25. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, et al. Determining the architectures of macromolecular assemblies. *Nature* 2007;450:683–694. [PubMed: 18046405]
26. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, et al. The molecular architecture of the nuclear pore complex. *Nature* 2007;450:695–701. [PubMed: 18046406]
27. Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J Cell Biol* 2000;148:635–651. [PubMed: 10684247]
28. Washburn MP, Wolters D, Yates JR III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19:242–247. [PubMed: 11231557]
29. Andersen JS, Wilkinson CJ, Mayor T, Mortensen P, Nigg EA, Mann M. Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* 2003;426:570–574. [PubMed: 14654843]
30. Dove JE, Brockenbrough JS, Aris JP. Isolation of nuclei and nucleoli from the yeast *Saccharomyces cerevisiae*. *Methods Cell Biol* 1998;53:33–46. [PubMed: 9348503]
31. McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ, Yates JR. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LCMS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int J Mass Spectrom* 2002;219:245–251.
32. Eng J, McCormack AL, Yates JR III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994;5:976–989.
33. Tabb DL, McDonald WH, Yates JR III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 2002;1:21–26. [PubMed: 12643522]
34. Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 2005;123:581–592. [PubMed: 16286007]
35. Davey MJ, Indiani C, O'Donnell M. Reconstitution of the Mcm2–7p heterohexamers, subunit arrangement, and ATP site architecture. *J Biol Chem* 2003;278:4491–4499. [PubMed: 12480933]
36. Chen J, Rappsilber J, Chiang YC, Russell P, Mann M, Denis CL. Purification and characterization of the 1.0 MDa CCR4-NOT complex identifies two novel components of the complex. *J Mol Biol* 2001;314:683–694. [PubMed: 11733989]
37. Schneider J, Wood A, Lee JS, Schuster R, Dueker J, Maguire C, et al. Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol Cell* 2005;19:849–856. [PubMed: 16168379]
38. Komarnitsky SI, Chiang YC, Luca FC, Chen J, Toyn JH, Winey M, et al. DBF2 protein kinase binds to and acts through the cell cycle-regulated MOB1 protein. *Mol Cell Biol* 1998;18:2100–2107. [PubMed: 9528782]
39. Lee JS, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, et al. Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* 2007;131:1084–1096. [PubMed: 18083099]

40. Al-Shahrour F, Minguez P, Tarraga J, Montaner D, Alloza E, Vaquerizas JM, et al. BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Res* 2006;34:W472–W476. [PubMed: 16845052]
41. Liu WL, Coleman RA, Grob P, King DS, Florens L, Washburn MP, et al. Structural changes in TAF4b-TFIID correlate with promoter selectivity. *Mol Cell* 2008;29:81–91. [PubMed: 18206971]

Fig. 1. Demonstration of protein and DNA fractionation by sucrose gradient centrifugation. (A) Schematic representation of the sucrose gradient on which nuclear extracts were fractionated. (B) Upper panel: Five percent of the total protein from each fraction was analyzed by 1-D SDS-PAGE and then visualized by silver staining. Lower panel: DNA from the corresponding fractions noted at the top of the upper panel. (C) The number of detected proteins in each of the fractions was averaged across the three biological replicate sucrose gradient fractionations. The number of detected *S. cerevisiae* proteins are reported by the grey bars, while the number of contaminants and shuffled proteins (as defined in Materials and methods) are reported in the stacked white and black bars, respectively.

Fig. 2.

Assessment of protein abundance in various functional categories. Protein abundance is given as a measurement of molecules per cell as determined by Ghaemmaghami et al. [23] and reported at <http://yeastgfp.ucsf.edu>. The protein functional categories were determined through the analysis of various *S. cerevisiae* databases including the *Saccharomyces* genome database (SGD; <http://www.yeastgenome.org/>), MIPS (<http://mips.gsf.de/>), as well as <http://yeastgfp.ucsf.edu> which has all localization data obtained by Huh et al. in their genome wide GFP-localization screen [21]. (A) All proteins in the database we used to query the MS/MS dataset, for which abundance values were available ($n=3537$). (B) Nuclearprotein of unknown function ($n=130$); this protein category contains proteins that were shown to localize to the nucleus by GFP but did not fit in any other nuclear functional class. (C) Chromatin remodeling and modifying factors ($n = 130$). (D) Transcription factors ($n=123$).

Fig. 3.

Percentage of *S. cerevisiae* nuclear proteins detected in sucrose gradient fractionation followed by MudPIT analysis. Each column represents the number of proteins either detected (light gray bar) or not (black bar) in each of the indicated groups of nuclear proteins. The protein categories are defined as all nuclear proteins (all nuclear), transcription factors (TFs), chromatin remodeling factors (CRFs), and RNA polymerase I, II, and III proteins (RNA Pol).

Fig. 4.

Categorization of the 2674 proteins identified by sucrose gradient followed by MudPIT according to their cellular localization, function, and relative abundance (NSAF values). The upper panels (A, C and E) report the qualitative information (i.e. number of proteins detected in each category) and share the same scale on their ordinate axes, while the lower panels (B, D, and F) report the quantitative results (i.e. NSAF range) and also share the same vertical axes. The detected proteins were sorted based on their cellular localization (panels A and B, which share the same scale and labels on their abscissa). Proteins annotated as localized to the nucleus and/or nucleolus (panels C and D) and proteins mapped to the other cellular compartments (panels E and F) were further sorted based on their function. The summary statistics are shown for NSAF values of proteins within each localization (B), and within each functional category for proteins localized to the nucleus and/or nucleolus (D), and all other cellular compartments (F) (see Supplemental Table 1). Functional categories were labeled as defined by MIPS: “transcription”, “cell cycle and DNA processing”, “metabolism and energy”, “protein fate”, “cellular transport”, “protein synthesis”, and “unclassified”. Functional classes with only a few entries were merged into a “miscellaneous” category (See Supplemental Table 1). In the box plots, the 25th and 75th percentiles are represented by the upper and lower boundaries of the box, with the median being the line dissecting the box, and the mean being the small square symbol. The 5th and 95th percentiles are shown as error bars, the ‘X’ represents the 1st and 99th percentiles, and the stand alone dashes ‘-’ represent the complete range.

Fig. 5.

Comparison between protein levels measured after nuclear enrichment and whole cell abundance. The number (expressed as a percent of total) of nuclear and non-nuclear proteins for which whole cell abundance levels were higher than NSAF values after sucrose gradient fractionation (fold change ≤ 0.33) is represented in grey, proteins for which both types of abundance values fell within the same range ($0.33 < \text{fold change} < 3$) are in white, and proteins for which abundance levels were higher after nuclear enrichment and fractionation (fold change ≥ 3) are represented by the black bars.

Fig. 6.

Sucrose gradient fractionation of contaminating proteins and transcriptional regulatory protein complexes. For each protein, NSAF values are plotted as a function of the sucrose gradient fractions from a representative experiment. (A) A representative sucrose gradient fractionation of proteins involved in cytoplasmic translation elongation is shown. The highly abundant protein Tef1p is shown in orange with other proteins including Efb1p, Tef4p, Cam1p, Hnbs1p, Eft1p, Yef3p, and Hef3p shown in grey lines for demonstration purposes only. The thick solid black line represents the average profile of all the components except Tef1p. (B) A representative sucrose gradient fractionation of cytosolic ribosomal subunits is shown. The thick solid black line represents the average profile for all the components. Plotted in grey are all the cytoplasmic ribosomal subunits for demonstration purposes only. Known components of the Rpd3 (C), MCM (D), CCR4-NOT (E), and COMPASS (F) complexes are shown as examples. The key for each graph is given in the legend inset. The thick solid black line in (C), (D), and (F) represents the average of all the NSAF values for all plotted components. In (E) the solid black line represents the average of all the NSAF values for all proteins except Pop2p and Mob1p, and the grey line represents the average NSAF values for Pop2p and Mob1p.

Table 1
Enrichment of low abundance proteins by sucrose gradient fractionation followed by MudPIT

Molecule per cell	Identified in Liu et al. [4] (% of total in database)	Identified in this study (% of total in database)			
		SG1	SG2	SG3	SG (1+2+3)
<100	6.3	24.0	24.0	12.0	32.0
100–1000	15.5	31.3	29.9	27.5	39.9
1000–10,000	34.9	49.2	52.5	45.2	60.7
10,000–100,000	85.6	76.2	84.3	76.4	86.8
>100,000	95.7	92.9	94.3	92.9	94.3

Comparison of sucrose gradient dataset with results from multiple unfractionated samples analyzed by MudPIT[4].

Table 2

Recovery of components of known nuclear protein complexes

Protein complexes ^a	Components identified in SGs ^b	Components in database	% Identified
ADA/NuA4/SAGA/SAGA-like(SLIK)	31	33	93.9%
CCR4-NOT	7	9	77.8%
CHRAC (chromatin accessibility)	4	4	100%
COMPASS	6	8	75%
FACT	4	4	100%
Histone acetyltransferases	5	6	83.3%
INO80	9	9	100%
ISW1	4	4	100%
MEDIATOR	22	27	81.5%
MCM (minichromosome maintenance)	6	6	100%
NAD-dependent histone deacetylase	5	7	71.4%
Paf	6	6	100%
RPD3/SIN3	12	12	100%
RSC(Remodel the structure of chromatin)	15	15	100%
Sister chromatid cohesion	7	9	77.8%
SWI/SNF transcription activator	9	10	90%
SWR1	7	7	100%

^aProtein complex information obtained from the *Saccharomyces* genome database (SGD;<http://www.yeastgenome.org/>) and the MIPS *Saccharomyces cerevisiae* genome database (<http://mips.gsf.de/genre/proj/yeast/>).

^bSucrose gradient (SG).