research articles Journal of proteome research

Orthogonal Separation Techniques for the Characterization of the Yeast Nuclear Proteome

Sharon Gauci,[†,‡] Liesbeth M. Veenhoff,[‡,§] Albert J. R. Heck,[†,‡] and Jeroen Krijgsveld*[,†,‡,||]

Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands, Netherlands Proteomics Centre and Centre for Biomedical Genetics, The Netherlands, and Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received February 6, 2009

The presence of the nucleus is the distinguishing feature of eukaryotic cells, separating the genome from the cytoplasm. Key cellular events, including transcription, DNA replication, RNA-processing and ribosome biogenesis all take place in the nucleus. All of these processes can be regulated through controlled and bidirectional translocation of proteins across the nuclear envelope, making the nucleus a highly dynamic organelle. In this study, we present four orthogonal multidimensional separation techniques for the comprehensive characterization of the yeast nuclear proteome. By combining methods on the peptide level (SCX chromatography, isoelectric focusing) and protein level (SDS-PAGE, phosphocellulose chromatography) coupled with mass spectrometry, we identified 1889 proteins from highly purified nuclei, of which 1032 were previously annotated as nuclear proteins. In particular, the most successful setup was the use of phosphocellulose P11 chromatography in combination with SDS-PAGE and reversed phase chromatography. Phosphocellulose P11 chromatography has been classically used for the purification of functional protein complexes involved in transcription regulation. Here, by its coupling with LC-MS, this method resulted in approximately 1.5 times more protein identifications than the other three combined, thereby contributing significantly to the coverage of nuclear proteins. In addition, the use of this technique resulted in the enrichment of DNA binding proteins and proved to be a valuable tool for the simultaneous analysis of multiple protein complexes. The enrichment for specific nuclear complexes has resulted in high protein sequence coverage, which will be particularly useful for the detailed characterization of subunits.

Keywords: *Saccharomyces cerevisiae* • Nucleus • prefractionation • Phosphocellulose P11 • mass spectrometry • SCX • peptide IEF • protein complex

Introduction

The nucleus is one of the most important and complex organelles, and is a distinctive feature of eukaryotes. It houses most of the cell's genetic material rendering it essential for the proper functioning of a cell. The functions of the nucleus are multiple, starting with the compaction of the genome within the boundaries of the nucleus. The genetic material is made accessible in a controlled manner involving general and sequence-specific transcription factors, to appropriately regulate gene expression depending on growth condition or changes in environmental conditions.[2−8] Proteins governing gene transcription and pre-mRNA splicing are located in the nucleus, as are the proteins regulating export of mRNA to the cytosol. Further communication between cytosol and nucleus occurs through the nuclear pore complex and its cognate transport factors, translocating proteins often in a phosphorylation-dependent manner.[3−5,9] Multiple proteins essential for cellular survival and regulation are driven by this principle, such as transcription factors, as ultimate effectors of signaling cascades,[6] and regulators of the cell cycle.[10−12] Thus, the nucleus is not a closed entity but a very dynamic organelle that responds extensively to extranuclear events. Because of the extensive shuttling between the cytoplasm and nucleus, large numbers of proteins are expected to reside in the nucleus at some point in time. Indeed, approximately 27% of all proteins from budding yeast *Saccharomyces cerevisiae* are annotated in Saccharomyces Genome Database (SGD) to be nuclear.[1,7] Localization of these proteins has been assessed experimentally in several genome-wide studies through random epitope-tagging, [13] plasmid-based overexpression of epitope-tagged proteins[7] and chromosomal expression of green fluorescent protein (GFP)-fusions.[7,8,13] The latter approach has enabled the localization of 70% of previously unlocalized proteins in budding yeast.[8]

* To whom correspondence should be addressed. Jeroen Krijgsveld, Biomolecular Mass Spectrometry and Proteomics Group, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands. Tel: +49-6221-3878560. E-mail: jeroen.krijgsveld@embl.de.

† Utrecht University.

‡ Netherlands Proteomics Centre and Centre for Biomedical Genetics.

§ University of Groningen.

|| Present address: EMBL, Meyerhofstrasse 1, Heidelberg, Germany.

10.1021/pr9000948 CCC: $40.75

Published on Web 04/28/2009

With the completion of the yeast genome,[15] numerous approaches were undertaken to characterize the yeast proteome. Efforts to chart the full proteome[16,17] have been complemented with studies on protein interactions using mass spectrometry-based techniques in combination with affinity tag technology,[18−20] creating a wiring diagram of protein networks. In addition, a number of studies have aimed at characterizing the proteomes of various organelles.[8,21−24] Yet, the proteomic analysis of the nucleus has been somewhat underexplored (except for a recent study),[25] possibly hampered by the fact that the isolation of yeast nuclei involves a number of critical steps before sufficiently pure preparations can be obtained.

Multidimensional separation techniques in combination with mass spectrometry have emerged as a powerful tool for the large-scale analysis of such complex samples.[26−29] Fractionation techniques are aimed to reduce sample complexity therefore permitting the identification of low-abundance proteins.[26,27,30,31] Various orthogonal prefractionation techniques on the protein and peptide level have been utilized for the characterization of a part of the yeast proteome leading to the identification of thousands of proteins.[17,32,33]

One of the most commonly used protein prefractionation techniques is one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), which has enabled the separation of large numbers of proteins according to their size. However, because of its low resolution, low dynamic range and poor loading capacity, other prefractionation techniques are desired. In particular, prefractionation at the peptide level utilizing specific peptide parameters has been widely applied.[26,28,29,31,34] Two such examples include strong cation exchange (SCX) and peptide isoelectric focusing (IEF).

While SCX chromatography separates peptides by charge, peptide IEF exploits the isoelectric point of peptides for focusing in discrete fractions, which is generally not possible with chromatographic techniques.[26,31,35,36] The use of an immobilized pH gradient further ensures a well-defined pH range and the flexibility to choose the required pH for the experiment. Moreover, we recently demonstrated the use of the peptide isoelectric point (p*I*) as an additional identification criterion, increasing the confidence level of protein identifications.[26]

Protein prefractionation has also been utilized as a first step in the purification of various functional protein complexes such as transcription factors.[37] However, because of limited access to high-throughput mass spectrometry, initially these techniques were not evaluated as potential generic prefractionation techniques for complex samples. For example, phosphocellulose P11 chromatography has been used for decades for the purification of general transcription factors or transcriptional regulators.[37,30] Phosphocellulose consists of a bifunctional cation exchanger containing both strong and weak acid groups based on an ester-linked orthophosphate group. These negatively charged phosphate groups make it suitable for the isolation of DNA binding proteins mimicking the phosphate groups present in DNA. Similar to SCX, elution is achieved with increasing salt concentration. A distinguishing feature is that separation is performed under near-native conditions assumed to preserve protein−protein interactions. Recently, Yaneva et al. further reported that a close correlation exists between the stability of a protein−phosphocellulose P11 column interaction and protein complex−DNA interaction in the presence of salt.[30]

In this study, we have used an optimized protocol for obtaining highly purified yeast nuclei as a starting point for studying the yeast nuclear proteome. We have used a number of orthogonal prefractionation methods at the peptide and protein level (SDS-PAGE, peptide IEF, SCX and phosphocellulose P11 chromatography) which were combined with high mass-accuracy mass spectrometry for the large-scale identification of nuclear proteins. Furthermore, we evaluate the use of phosphocellulose P11 chromatography as a prefractionation technique in combination with mass spectrometry for the analysis of complex samples, as well as its value in the enrichment of DNA-binding proteins. The results show that the use of these prefractionation techniques led to the identification of 1889 unique proteins, 1032 of which were previously known to be nuclear, including over 200 (91%) nucleolar proteins. We discuss the likelihood that at least a portion of the remaining proteins is localized in the nucleus as well. Additionally, we demonstrate that phosphocellulose P11 chromatography is a versatile prefractionation tool identifying multiple protein complexes involved in transcription and gene expression, adding a level of information that cannot be obtained from the other fractionation methods.

Materials and Methods

Preparation of Yeast Nuclei. Nuclei were isolated from *S. cerevisiae* BY 4743 as described by Rout and Kilmartin.[38,39] Briefly, 9 L of yeast culture at 1×10^7 cells/mL was harvested, washed and converted to spheroplasts by using Glusulase NEE-154 Du Pont/NEN (Perkinelmer), ZymolyaseTM-20T from *Arthrobacter luteus* (Seikagaku America), and Lysing Enzymes *Trichoderma harzianum* (Sigma-Aldrich, St. Louis, MO) in 1.1 M sorbitol for 3 h at 30 °C. The spheroplasts were harvested by centrifugation and carefully resuspended. Residual lysing enzyme and cytosol from broken spheroplasts were removed by centrifugation over a ficoll layer and careful removal of the sample and cushion layers. The spheroplasts were then lysed in 8% Polyvinylpyrrolidone (PVP) solution by the sheering force generated by a polytron. During this lysis step, intact nuclei were released from the spheroplasts, while other organelles, including ER, were shattered into pieces. Immediately after lysis, the lysate was diluted and the crude nuclei were isolated by centrifugation. The nuclei were purified in a three step sucrose/PVP gradient (layers of 2.01, 2.1, and 2.3 M sucrose/PVP) at the 2.1/2.3 M interface and harvested by centrifugation. Approximately 8 mg of nuclear protein was isolated from 9 L of culture.

Phosphocellulose P11-SDS-PAGE. All procedures were performed at 4 °C. Nuclei (2 mg of protein) were pelleted at 100 000*g* to form a tight pellet. The nuclei were then lysed in 1 mL of phosphocellulose buffer [20 mM HEPES (Merck KGaA Germany), 0.01% Tween (Merck KGaA Germany), 10% Glycerol, 75 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT (Sigma-Aldrich, St. Louis, MO), 1 mM PMSF (Sigma-Aldrich, St. Louis, MO), and 2 Tablets of protease inhibitors cocktail (Roche Diagnostics, Germany). The presence of EDTA makes the nuclei fragile. In this step, the buffer was swirled over the nuclei pellet such that the shearing force lysed the nuclei. After complete disappearance of the pellet was performed after ∼10 min, vortexing was continued for an additional 5 min. A 0.5 mL phosphocellulose P11 column (Whatman, Maidstone, U.K.) was prepared according to manufacturer's instructions on a 0.8 × 4 cm Poly-Prep column (Bio-Rad, Hercules, CA) and equilibrated with the phosphocellulose buffer. The 2 mg sample was loaded onto the P11 column and eluted with a stepwise gradient of 20 column volumes of phosphocellulose buffer supplemented with

0.1, 0.3, 0.5, and 0.85 M NaCl.[30] Diafiltration of the eluted fractions against 50 mM ammonium bicarbonate pH 8 was performed using iCON concentrators 7 mL/9K (Pierce, Rockford, IL). One-third of the resulting fractions were loaded on a mini 12.5% SDS gel and each lane was excised into 23 gel pieces. The proteins were reduced with 10 mM DTT (Roche Diagnostics) at 56 °C followed by alkylation with 55 mM iodoacetamide (Sigma-Aldrich, St. Louis, MO) at room temperature for 1 h. A trypsin digestion was subsequently performed overnight at an enzyme/substrate ratio of 1:50 (w/w). Approximately 2 μg of material from each band was analyzed by mass spectrometry.

1D SDS-PAGE. One hundred micrograms of a nuclear extract was pelleted at 100 000*g* and then solubilized in SDS Buffer (0.5 M Tris, pH 6.8, 5% SDS, glycerol, milli-Q water, Bromophenol Blue, and 10 mM DTT). 1D SDS-PAGE was performed on a 12.5% maxi gel using the BioRad Protean II Electrophoresis system (BioRad, Veenendaal, The Netherlands) using 60 V in the stacking layer, increasing up to 80 V during the separation. The gel was stained using Gelcode blue stain reagent (Pierce, Rockford, IL) overnight and then washed with milli-Q water. The lane was excised into 22 gel pieces and reduced with 6.5 mM DTT (Roche Diagnostics) at 56 °C followed by alkylation with 54 mM iodoacetamide (Sigma-Aldrich, St. Louis, MO) for 1 h at room temperature (RT), to be then digested with trypsin at an enzyme/substrate ratio of 1:50 (w/w). Again, approximately 2 μg of sample from each band was used for further analysis by mass spectrometry.

Strong Cation Exchange Chromatography. Two hundred micrograms of the nuclear extract was pelleted at 100 000*g* and reconstituted in 90% methanol (v/v), vortexed briefly, and left at −20 °C for 1 h. The sample was then centrifuged at 14 000*g* at 4 °C for 20 min and the supernatant was collected. This process was repeated to ensure removal of any residual PVP in the sample. Strong cation exchange was performed using Zorbax BioSCX-Series II columns (0.8 mm (i.d.) × 50 mm (l); particle size, 3.5 μm), an Agilent 1100 Series binary pump and autosampler (Agilent Technologies), and a SPD-6A UV-detector (Shimadzu, Tokyo, Japan). The pellet was reconstituted in 20% acetonitrile (ACN) and 0.05% formic acid. After injection, the first 2 min were run isocratically at 100% solvent A (0.05% formic acid in 8:2 (v/v) water/ACN, pH 3.0), followed by a linear gradient to 90% solvent B in 48 min (500 mM NaCl in 0.05% formic acid in 8:2 (v/v) water/ACN, pH 3.0), followed by 2 min of 100% solvent A. A total number of 50 SCX fractions (1 min each, i.e., 50 μL elution volume) were manually collected and dried in a vacuum centrifuge. Fractions 3−48 were reconstituted in 10% formic acid for further analysis. Around 2 μg of each SCX fraction was analyzed further.

In-Gel Peptide Isoelectric Focusing. Two hundred micrograms of the nuclear extract was pelleted at 100 000*g* and a methanol precipitation was performed twice for the removal of any residual PVP as mentioned in the previous section. The pellet was reconstituted in 8 M urea and 50 mM ammonium bicarbonate pH 8. Proteins were reduced with a final concentration of 10 mM DTT (Roche Diagnostics) at 56 °C followed by alkylation with iodoacetamide (Sigma-Aldrich, St. Louis, MO) at a final concentration of 55 mM at room temperature for 1 h. A Lys-C (Roche Diagnostics) digestion was performed overnight at an enzyme/substrate ratio of 1:100 (w/w). The eluate was further diluted to 2 M Urea/50 mM ammonium bicarbonate pH 8 and a tryptic digest was performed overnight at 37 °C in an enzyme/substrate ratio of 1:50 (w/w). The sample was desalted using STAGE tips[40] and lyophilized. The dried sample was reconstituted in 8 M urea in the presence of the IPG buffer 3−10 NL (GE Healthcare) and applied to a 24 cm IPG dry strip, 3−10 NL (GE Healthcare). With the use of the IPGphor (GE Healthcare), the following focusing protocol was utilized: 30 V, 4 h; 500 V, 2 h; 1000 V, 1 h; 8000 V up to 60 000 V/h. The IPG strip was cut into 48 sections and peptides were extracted 3 times using 0%, 50% and 100% ACN, respectively, in water and 0.1% TFA.[26] Any residual oil and salt in each fraction was removed using a 96 well Oasis SPE HLB μElution plate (Waters, Milford, MA), dried and reconstituted in 10% formic acid. Similarly as with the previous methods, approximately 2 μg of each gel piece was analyzed further.

Nanoflow-HPLC. Nanoflow liquid chromatography was performed on an Agilent 1100 HPLC binary solvent delivery system (Agilent Technologies, Waldbronn, Germany) with a thermostatted wellplate autosampler coupled with an LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany) or an LTQ-FTICR mass spectrometer (Thermo Electron, Bremen, Germany). Peptides were trapped at 5 μL/min in 100% A (0.1 M acetic acid in water) on a trapping column (30 mm × 100 μm packed in-house with Aqua C_{18}, Phenomenex, Torrance, CA) for 10 min. After flow-splitting down to ∼100 nL/min, peptides were transferred to the analytical column (200 mm × 50 μm packed in-house with Reprosil-Pur C_{18}-AQ, Dr. Maisch, GmbH, Ammerbuch, Germany) and eluted with a gradient of 0−40% B (80% ACN/0.1 M acetic acid) in 40 min in a 60 min gradient. Nanospray was achieved using a coated fused silica emitter (New Objective, Cambridge, MA) (o.d., 360 μm; i.d., 20 μm, tip i.d. 10 μm). A 33 MΩ resistor was introduced between the high voltage supply and the electrospray needle to reduce ion current.

Mass Spectrometry. The LTQ-Orbitrap mass spectrometer and the LTQ-FTICR were operated in data-dependent mode, automatically switching between MS and MS/MS. The two most intense peaks above a threshold of 500 were selected for collision induced dissociation (CID) in the linear ion trap at normalized collision energy of 35%. In the LTQ-Orbitrap, full scan MS spectra were acquired with a resolution of 60 000 at 400 m/z after accumulation to a target value of 500 000, while in the LTQ-FTICR, full scan MS spectra were acquired with a resolution of 100 000 at 400 m/z after accumulation to a target value of 1 000 000.

Data Analysis. All MS/MS spectra were converted to DTA files using Bioworks 3.3 (Thermo, San Jose, CA). An in-house developed Perl script was used to convert all spectra into a single file and this was searched using MASCOT search engine (Version 2.2.01, Matrix Science, London, U.K.) against Yeast Saccharomyces Genome Database (SGD) available at http://www.yeastgenome.org with cysteine carbamidomethylation as a fixed modification. Methionine oxidation and deamidation (at Asn and Gln) were chosen as a variable modification. A peptide mass tolerance of 15 ppm and fragment mass tolerance of 0.9 Da were selected and Trypsin was chosen as proteolytic enzyme allowing one missed cleavage. All data were loaded into Scaffold (version 02.01.00, Proteome-Software, Portland, OR) to probabilistically validate peptide and protein identifications. Peptide identifications were accepted when they reached a greater than 95.0% probability as specified by the Peptide Prophet algorithm.[41] Protein identifications were accepted if they could be established at greater than 95.0% probability, as assigned by the Protein Prophet algorithm,[42] with the additional criterion that they contained at least 2 identified peptides.

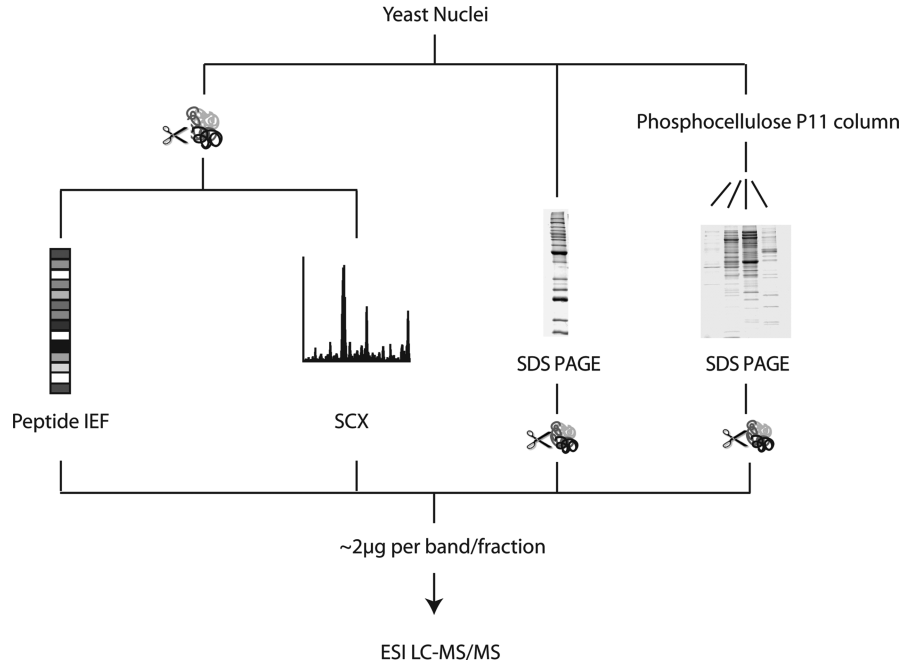


Figure 1. Schematic workflow of the experiment. Following nuclei extraction, two protein and peptide prefractionation techniques were utilized for the large-scale identification of nuclear proteins. In the peptide prefractionation, the nuclei were first lysed and digested to be subsequently separated using SCX or in-gel peptide IEF. While with the protein prefractionation, the lysed nuclei were first separated using a maxi 12.5% SDS gel or applied onto a phosphocellulose P11 column to be subsequently eluted and separated on a mini SDS gel. Each lane was excised followed by in-gel tryptic digestion. All the peptides resulting from these prefractionation techniques were applied in an equal manner and analyzed using reversed phase liquid chromatography coupled with mass spectrometry.

Effectively, this results in a <1% false positive rate. The peptide p*I* was determined using an in-house built p*I* calculator.[43] Peptide p*I* outliers with a p*I* of ±2 p*I* units from the expected average p*I* and a Mascot score below 30 were eliminated. The identified proteins in the phosphocellulose experiment were grouped into protein complexes as annotated in the Yeast SGD database. The elution patterns were determined by choosing only proteins with ≥50% of the identified peptides in one fraction and only if the proteins collectively amount to 40% of the protein complex in one fraction. If only two proteins belonging to one complex were identified, both proteins had to be eluting in one fraction, otherwise they were ignored. Gene ontologies of identified proteins were determined using http://www.yeastgenome.org. The significance of enrichment of protein classes were calculated using the entire yeast proteome as a background.

Results

As a starting point for this study, we have applied a rigorous protocol[38,39] to obtain highly enriched and homogeneous nuclei (Supplementary Figure 1). Specifically, three spheroplasting agents were used, followed by lysis using sheering force to liberate intact nuclei while fragmenting the other organelles. A crude nuclei fraction was then purified over a sucrose/PVP gradient, to remove contaminating organelle fragments, for example, from mitochondria, plasma membrane, Golgi and peripheral ER. Given the expected complexity of the nuclear proteome, prefractionation is an essential step for its in-depth characterization by mass spectrometry. Starting from the highly purified nuclei, we used four different protein and peptide prefractionation techniques: SDS-PAGE, SCX, in-gel peptide IEF and phosphocellulose P11 chromatography with SDS-PAGE, each coupled with high mass accuracy mass spectrometry (Figure 1). For this purpose, we utilized the extract of purified nuclei and evaluated each technique separately as well as collectively for the identification of yeast nuclear proteins, using equal amounts of sample as input for the MS analysis. Additionally, we investigated the use of phosphocellulose P11 chromatography as a prefractionation tool as well as for the enrichment of DNA binding proteins.

SDS-PAGE. SDS-PAGE is one of the most commonly used protein separation techniques, separating proteins by size. For the purpose of this experiment, a maxi 12.5% SDS gel was utilized to load 100 μg of the nuclear extract. This lane was subsequently excised into 23 gel pieces and analyzed using a reversed phase nanoflow LC coupled with an LTQ FT-ICR. A total of 806 unique proteins were identified with a minimum of two peptides per protein (Supplementary Table 1). Among these were 502 known nuclear proteins out of which 163 nucleolar proteins annotated in the yeast SGD database. Given the complexity of the sample, the low dynamic range and the poor loading capacity, SDS-PAGE alone was not sufficient to obtain a high-resolution protein separation to characterize the yeast nuclear proteome. Therefore, to complement these data, other orthogonal protein and peptide prefractionation techniques were chosen.

Strong Cation Exchange Chromatography. SCX is a well-established peptide prefractionation technique separating peptides based on charge, with higher charged peptides eluting at increasing salt concentrations.[31] A tryptic digest of 200 μg of the nuclear extract was loaded onto the SCX column and eluted in a linear salt gradient. A total of 46 fractions were analyzed using a reversed phase nanoflow LC coupled with an LTQ-FTICR. In the first 24 fractions, the singly charged peptides and

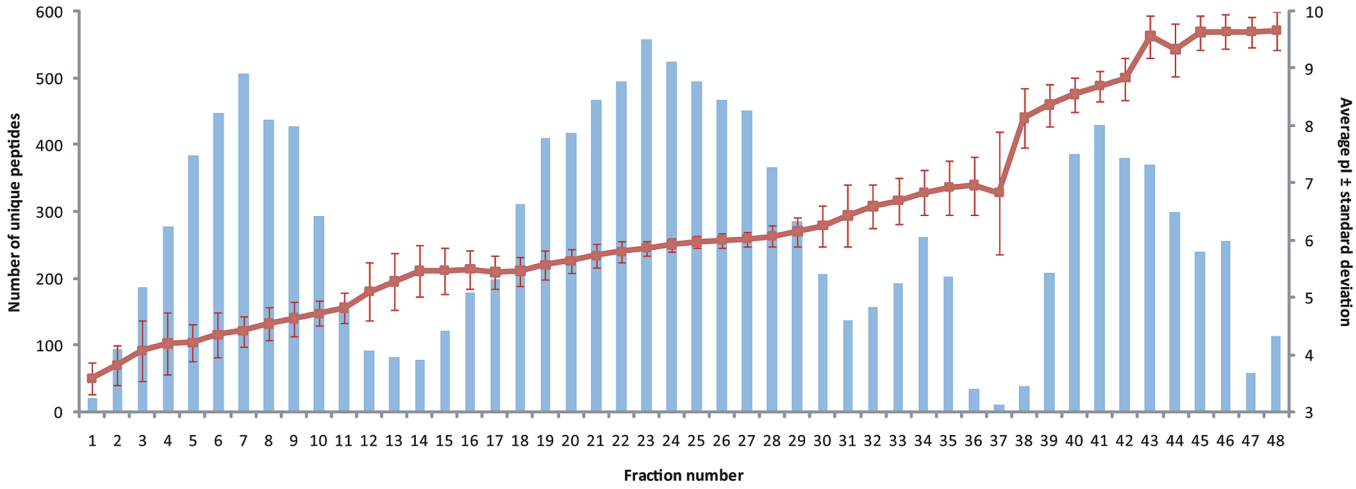


Figure 2. Distribution of peptides using in-gel peptide IEF. Distribution of tryptic peptides from the yeast extract over a 3−10 NL IPG strip. For each of the 48 gel fractions, the number of unique identified peptides (blue bars, left axis) and the average p*I* ± standard deviation of these peptides (red line, right axis) are plotted.

the 2+ peptides were identified followed by the 3+ and 4+ peptides in the remaining fractions. In total, 805 unique proteins were identified with a minimum of two peptides per protein (Supplementary Table 2), out of which 526 and 158 proteins were classified as nuclear and nucleolar proteins, respectively, in the Yeast SGD database.

In-Gel Peptide Isoelectric Focusing. Peptide IEF exploits the peptide p*I* to achieve high-resolution separation in distinct fractions. A tryptic digest of 200 μg of the nuclear extract was loaded on a 3−10 NL 24 cm IPG strip, which was excised into 48 equal parts after focusing. Peptides were extracted and loaded onto a reversed phase nano-LC-LTQ-FTICR for identification. A total of 7205 unique peptides were identified (Figure 2) showing a nonlinear distribution across the IPG strip. The accurate focusing of peptides enabled the removal of p*I* outliers with a p*I* of ±2 p*I* units from the expected average p*I*. In the total set of peptides (Supplementary Table 3), 3849 (29%) peptides were identified solely in one fraction, 4048 (31%) in 2 fractions, 2199 (17%) in 3 fractions and decreasing numbers were identified in up to 10 fractions. The largest number of peptides was identified in fraction 23 with 559 unique peptides corresponding to an average predicted p*I* of 5.86 ± 0.16 (Figure 2). Peptide distribution was found to be uneven along the p*I* strip (Figure 2), consistent with similar trends observed in analyzing *Drosophila*[26] and *E. coli* samples by peptide IEF.[44] This includes a low number of peptides which were focused around pH 7 in the IPG strip corresponding to fraction 37 in this experimental setup.[26] This indicates that the distribution of charged residues of tryptic peptides is comparable in different biological systems.

The average standard deviation throughout the strip was ±0.34 p*I* units giving an indication of the resolving power of the experiment. A high standard deviation was observed in fractions containing lower amounts of peptides such as fractions 36−38. Overall, 643 unique proteins were identified with a minimum of two peptides per protein (Supplementary Table 4), among which we found 415 known nuclear proteins and 139 nucleolar proteins.

Phosphocellulose P11- SDS-PAGE. Phosphocellulose P11 chromatography is a bifunctional cation exchanger containing strong and weak acid groups based on ester-linked orthophosphate functional group. This technique has been widely used

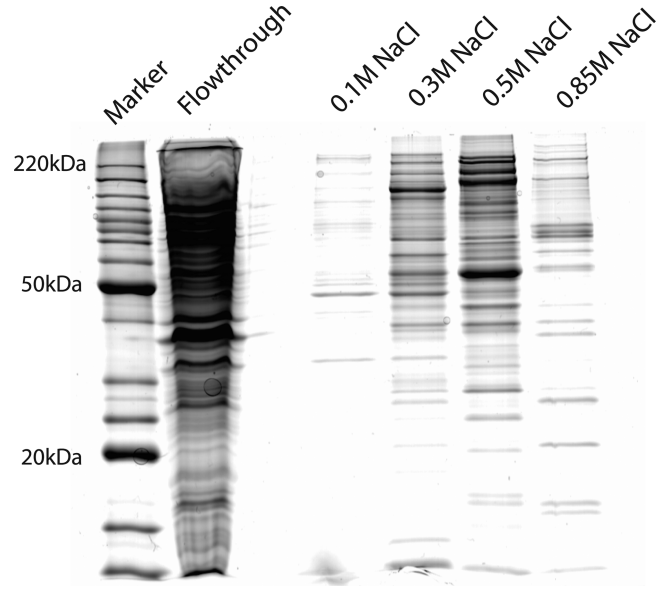


Figure 3. SDS-PAGE of the eluted phosphocellulose P11 fractions. Distinct bands could be visualized in the eluting salt fractions indicating that different proteins were eluting depending on the salt concentrations utilized.

for the isolation of DNA binding proteins by exploiting the negatively charged phosphate groups mimicking those found on DNA. This strategy has been used as a first step in numerous purification protocols of functional protein complexes, since under these conditions most protein−protein interactions remain intact.[30,37,45] Two milligrams of the protein sample was loaded on a 0.5 mL phosphocellulose column (0.8 × 4 cm) and eluted in a stepwise manner in buffer containing 0.1, 0.3, 0.5, and 0.85 M NaCl.[30] Each salt fraction was loaded on a 12.5% SDS-PAGE gel (Figure 3) showing a distinct pattern in each fraction. This indicates that different proteins were eluted with increasing ionic strength. The column was washed with 20 column volumes until the eluate contained no residual protein (data not shown). Furthermore, the overloading in the flow-through indicated that a surplus of proteins did not have a high affinity for the column (Figure 3).

The SDS-PAGE gel was processed for in-gel digestion and protein identification by reversed phase nanoflow LC coupled

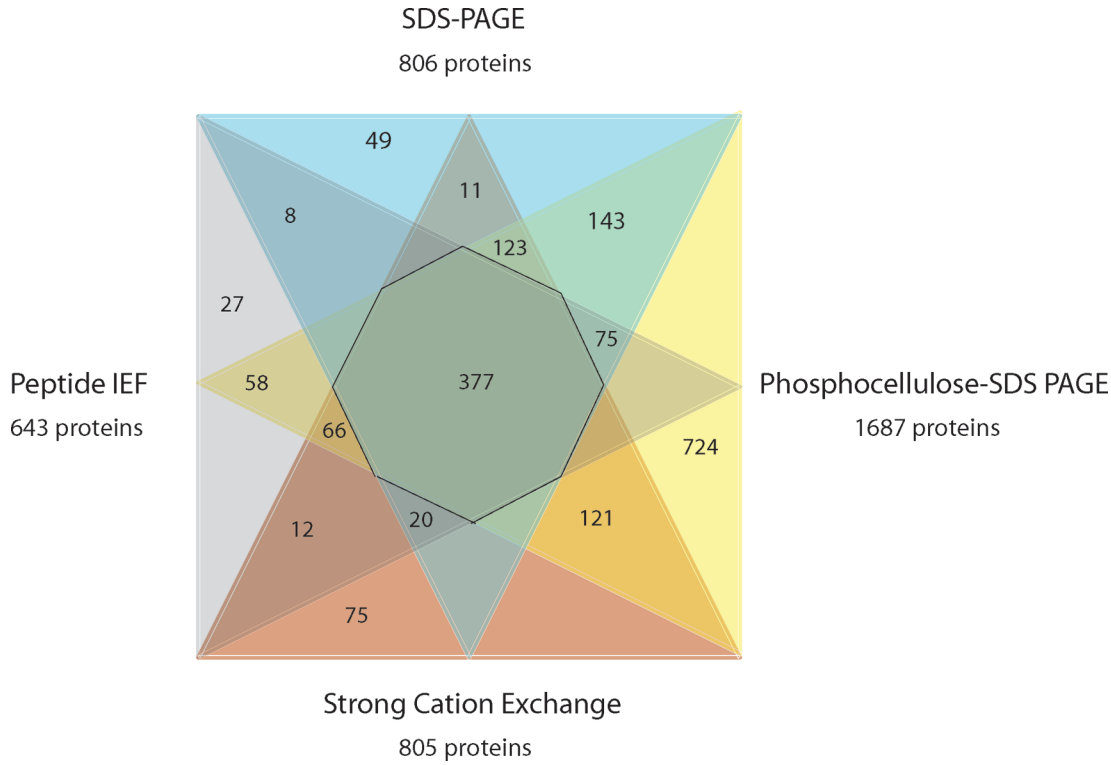


Figure 4. The direct comparison of protein identifications using the four techniques. A Venn diagram illustrating in detail the number of unique and shared proteins identified with each technique. A total of 377 proteins were commonly identified with all techniques, while phosphocellulose P11 chromatography- SDS-PAGE led to the highest number of unique protein identifications.

with an LTQ-FTICR or an LTQ-Orbitrap mass spectrometer. The combination of all these fractions (including the flow-through) led to the identification of 1687 unique proteins with a minimum of two peptides per protein (Supplementary Table 5). A total of 1311 proteins were solely identified in the eluting salt fractions, while 1071 proteins were identified in the flow-through with an overlap of 696 proteins. The salt elution steps led to the identification of 1041, 1057, 968, and 703 unique proteins in the 0.85, 0.5, 0.3, and 0.1 M fractions, respectively. Collectively, phosphocellulose P11 chromatography led to the identification of 930 and 201 proteins annotated to be nuclear or nucleolar, respectively (refer to Supplementary Table 6).

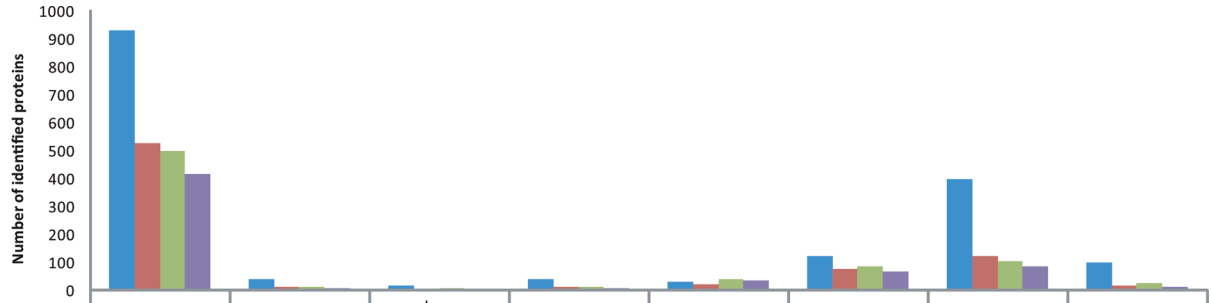
Comparison of the Four Prefractionation Techniques. Collectively, the four techniques led to the identification of 1889 unique proteins (Supplementary Table 7). Figure 4 illustrates in detail the number of unique proteins identified for each technique. As can be observed in this figure, phosphocellulose P11-SDS-PAGE led to the highest number of identifications (1687 proteins), of which 724 exclusively identified with this technique. SCX, SDS-PAGE and peptide IEF added a total of 202 proteins not identified by phosphocellulose P11-SDS-PAGE, of which 75, 49, and 27 were unique for the respective techniques. A total of 377 proteins were identified by all 4 techniques.

With regards to nuclear protein identifications, phosphocellulose P11-SDS-PAGE led to the identification of the highest number of known nuclear proteins listed in the Yeast SGD database (Figure 5). A total of 930 nuclear proteins were identified with this technique compared to 415 with peptide IEF, 502 with SDS-PAGE alone and 526 nuclear proteins with SCX. Collectively, 1032 known nuclear proteins were identified. On average, around 60% of the proteins identified with these four techniques were classified as nuclear proteins (Figure 5).

Analysis of the data further revealed that the 211 nucleolar proteins identified with this study constituted around 91% of the ones listed in the Yeast SGD database (Figure 6). SCX, in-gel peptide IEF, and SDS-PAGE performed equally well in identifying 158, 139, and 163 nucleolar proteins individually, while phosphocellulose P11 resulted in the identification of 201 nucleolar proteins. A total of 116 nucleolar proteins were identified in common between the four techniques, with phosphocellulose P11 contributing 23 unique nucleolar proteins, compared to 3, 3, and 2 with SDS-PAGE, SCX and in-gel peptide IEF, respectively.

Characterization of the Nuclear Proteome. The four respective data sets were combined to create a compendium of yeast nuclear proteins, containing 1889 proteins (Supplementary Table 7). To further interpret this total data set, we inferred the localization of the identified proteins as documented by GO annotation in the Yeast SGD database (Supplementary Table 6 and Figure 5). This shows that 1032 proteins with known or predicted nuclear localization were identified. Proteins from other membrane bound organelles were identified as well, but these appeared in relatively low numbers (Figure 5). Especially the small number of proteins from the ER that is continuous with the nuclear envelope indicated that the preparation of the nuclear extract was very efficient (Figure 5). Yet, the localization of numerous proteins was not annotated and could not be established via GO annotation (labeled "unknown"). Because of the continuous shuttling of regulatory proteins between the nucleus and cytoplasm, it cannot be excluded that these unallocated proteins, as well as some of the cytoplasmic proteins, can actually be present in the nucleus, as will be discussed below.

Of the 1848 proteins assigned as nuclear in the Yeast SGD database, 1032 were identified in our study (Figure 6). Since



	nuclear	mitochondrial	plasma membrane	golgi	ER	ribosomal	cytoplasm	unknown
phosphocellulose-SDS PAGE	930 (55%)	43 (2.6%)	20 (1.2%)	40 (2.4%)	33 (2%)	123 (7.3%)	398 (23.6%)	100 (5.9%)
strong cation exchange	526 (65.3%)	14 (1.8%)	5 (0.6%)	13 (1.6%)	25 (3.1%)	80 (9.9%)	123 (15.3%)	19 (2.4%)
SDS-PAGE	502 (62.3%)	15 (1.9%)	7 (0.9%)	16 (2%)	42 (5.2%)	89 (11%)	106 (13.1%)	29 (3.6%)
Peptide IEF	415 (64.5%)	7 (1.1%)	4 (0.6%)	10 (1.6%)	36 (5.6%)	68 (10.6%)	88 (13.7%)	15 (2.3%)
CUMULATIVE UNIQUE	1032	52	25	46	70	136	427	101

Figure 5. Protein localization for each prefractionation technique. A plot of the number of proteins identified versus the localization according to the yeast SGD database. Phosphocellulose P11 chromatography-SDS-PAGE, SCX, SDS-PAGE alone, and in-gel peptide IEF are illustrated in blue, red, green, and purple, respectively, along with the number and percentage of the total number of proteins for each method in the table below. The cumulative unique number of proteins per localization is also displayed for all four techniques.

Figure 6. The number of identified nuclear and nucleolar proteins listed in the Yeast SGD database. A pie chart illustrating the number of nuclear (A) and nucleolar (B) proteins identified from the yeast SGD database using all four techniques. The proteins identified are illustrated in gray, while unidentified proteins with these techniques are shown in white. The use of these orthogonal prefractionation techniques led to the identification of 61% of all nuclear proteins listed in the yeast SGD database out of which we find 91% coverage of all nucleolar proteins.

we did not identify all known nuclear proteins, and to see if our data were biased against low copy number-proteins, we plotted our data as a function of protein abundance based on estimations by Ghememaghami et al.[46](Supplementary Figure 2). Of the 1032 established nuclear proteins identified in our study, 891 proteins have been associated with a copy number. These were plotted in bins, showing both absolute numbers of identifications as well as relative numbers compared to the total as reported by Ghememaghami et al.[46] Over a dynamic range of 200−10 000 copies per cell, we find a coverage increasing from approximately 40% for the categories of low-abundance proteins to around 70% for the most abundant proteins. The largest number of proteins is present in 1000−4000 copies/cell, a range that is well-represented in our data set. Of note, some of the very high abundant proteins (>14000 copies/cell) are missing in this study, raising the possibility that these are not (always) nuclear. Another factor that could overestimate the number of nonidentified proteins lies in redundancy of the database listing identical proteins by different names. In fact, some of the proteins apparently not identified in the highest abundance bin are identical to proteins with another accession number that in fact were identified. These include HHF2 (Histone H4, identical to HHF1), HHT2 (Histone H3, identical to HHT1), RPS6A (identical to RPS6B) and RPS16A (identical to RPS16B).

Phosphocellulose P11 Chromatography as an Enrichment Technique for DNA Binding Proteins. In this study, we exploited the properties of P11 to identify proteins in the context of the entire nuclear proteome. We classified all proteins identified after P11 chromatography to evaluate its performance in covering classes of proteins expected to reside in the nucleus, in terms of localization, function and molecular process (Supplementary Table 8). Tables 1 and 2 illustrate the most prominent protein functions and localizations, along with the number of proteins identified and the significance of the enrichment for each function. This resulted in a high predominance of nuclear proteins, transcription factors and specifically protein complexes. Among the top-scoring terms are (sub)-nuclear structures, including protein complexes. Diverse functions related to maintenance and transcription of the genome (transcription, gene expression, RNA polymerization, DNA packaging, chromatin modification, etc.) are well-represented (Table 2). Significant levels for enrichment of these classes easily reach p-values of 1×10^{-25} and lower (Supplementary

Table 1. Protein Functions of Identified Proteins[a]

GO_term	identified proteins	% of total assigned	*P*-value
gene expression	586	44%	7.10×10^{-73}
DNA-dependent transcription	154	73%	1.38×10^{-50}
chromosome organization and biogenesis	294	51%	3.14×10^{-48}
transcription	162	68%	4.94×10^{-47}
chromatin modification	146	67%	5.09×10^{-40}
transcription from RNA pol II promoter	116	73%	6.74×10^{-37}
chromatin remodeling	99	66%	8.83×10^{-26}
DNA packaging	70	65%	5.78×10^{-17}
termination of RNA pol II transcription	9	100%	2.63×10^{-3}
regulation of transcription from RNA pol I promoter	9	100%	2.63E-03

[a] Some of the most prominent protein functions together with the number of proteins identified with phosphocellulose P11- SDS-PAGE and the probability value for each function.

Table 2. Component Enrichment of Identified Proteins[a]

GO_term	identified proteins	% of total assigned	*P*-value
macromolecular complex	916	52%	1.84×10^{-206}
nuclear part	650	58%	3.11×10^{-161}
nucleus	918	46%	3.89×10^{-153}
protein complex	675	53%	2.55×10^{-143}
nucleolus	197	65%	2.79×10^{-52}
transcription factor complex	92	70%	2.02×10^{-27}
DNA-directed RNA polymerase II, holoenzyme	61	85%	1.86×10^{-25}
chromosome	127	52%	4.51×10^{-20}
nuclear pore	39	76%	8.10×10^{-13}
histone acetyltransferase complex	32	90%	1.07×10^{-10}

[a] Some of the most important protein component enrichments for the proteins identified with phosphocellulose P11- SDS-PAGE, together with the number of proteins and the probability value for each component.

Table 8). Many of these classes contain hundreds of proteins, which are represented up to around 75%. Interestingly, some of the less-populated structures are fully covered; for example, all components of RNA polymerases and general transcription factor complexes were identified, as well as all factors involved in transcription termination and mRNA export (Tables 1 and 2, Supplementary Table 8). Collectively, this indicates that P11 chromatography combined with SDS-PAGE and LC-MS provides a profound insight and coverage of the nuclear proteome.

Protein Complexes Isolation Using Phosphocellulose P11 Chromatography. While most of the studies utilizing P11 are targeting a single protein complex, we wanted to explore whether phosphocellulose P11 combined with in-depth MS could be used to study multiple complexes simultaneously in a nuclear extract. A first indication that this might be fruitful is the observation that the top-scoring component in our GO-analysis (Supplementary Table 8) is in fact 'macromolecular complex' ($p = 1.84 \times 10^{-206}$) covering 52% of all proteins known to be part of a complex (in cytosol and nucleus). To explore this in further detail, the 1687 proteins identified in subsequent P11 fractions were grouped into the various protein complexes as defined in the Yeast SGD database (Supplementary Table 9).

The number of identified spectra per protein (spectral counting) was utilized as a quantitative indication of the presence of each protein per fraction. This allows the visualization of elution profiles of individual proteins across all fractions, and the comparison of elution profiles of proteins present in complexes. By grouping profiles of proteins per complex, it can be estimated whether complexes remain intact or disassemble during elution. By taking a closer look at the elution patterns of proteins belonging to a particular complex, we encountered three different scenarios: (1) the proteins do not bind to the P11 column and are only found in the flow-through, (2) the proteins elute in one salt fraction, or (3) proteins elute across several fractions. These various modes of elution are illustrated in Figure 7 for a number of complexes. This shows that the MCM-complex and the proteasome core complex do not have affinity for p11 and elute in the flow-through (Figure 7A). The THO and NuA4 histone acetylstransferase complexes are relatively stable and elute in a higher salt fraction (Figure 7B). Members of the exosome and the SAGA complex were found across several fractions (Figure 7C) indicating that they are either unstable under these conditions or that the complex exists in different forms. Therefore, we defined criteria to distinguish stable from instable complexes. First, we required $\geq$50% of the identified peptides of a protein to elute in one fraction to categorize that protein as eluting in that particular fraction. Next, we required $\geq$40% of the proteins in a protein complex to elute together in the same fraction to call the protein complex eluting in that fraction. Protein complexes of which less than 40% of its protein components coeluted in one fraction were considered to be spread across all fractions (Supplementary Table 10).

With the use of these criteria, the analysis resulted in the classification of 181 protein complexes of which 122 specifically enriched in a particular fraction (Supplementary Table 10). Sixty complexes did not bind to the column and eluted in the flow-through, and 62 complexes eluted in a particular salt fraction (1 eluted at 0.1 M salt, 17 at 0.3 M, 32 at 0.5 M and 12 at 0.85 M fractions). Fifty-nine complexes were not specifically enriched in one fraction and were found to be spread out over several fractions. An additional 57 complexes were ignored because they were represented by only 1 protein. On average, 78% of the proteins within a complex were found in the salt fractions. Some complexes were covered by all of their known

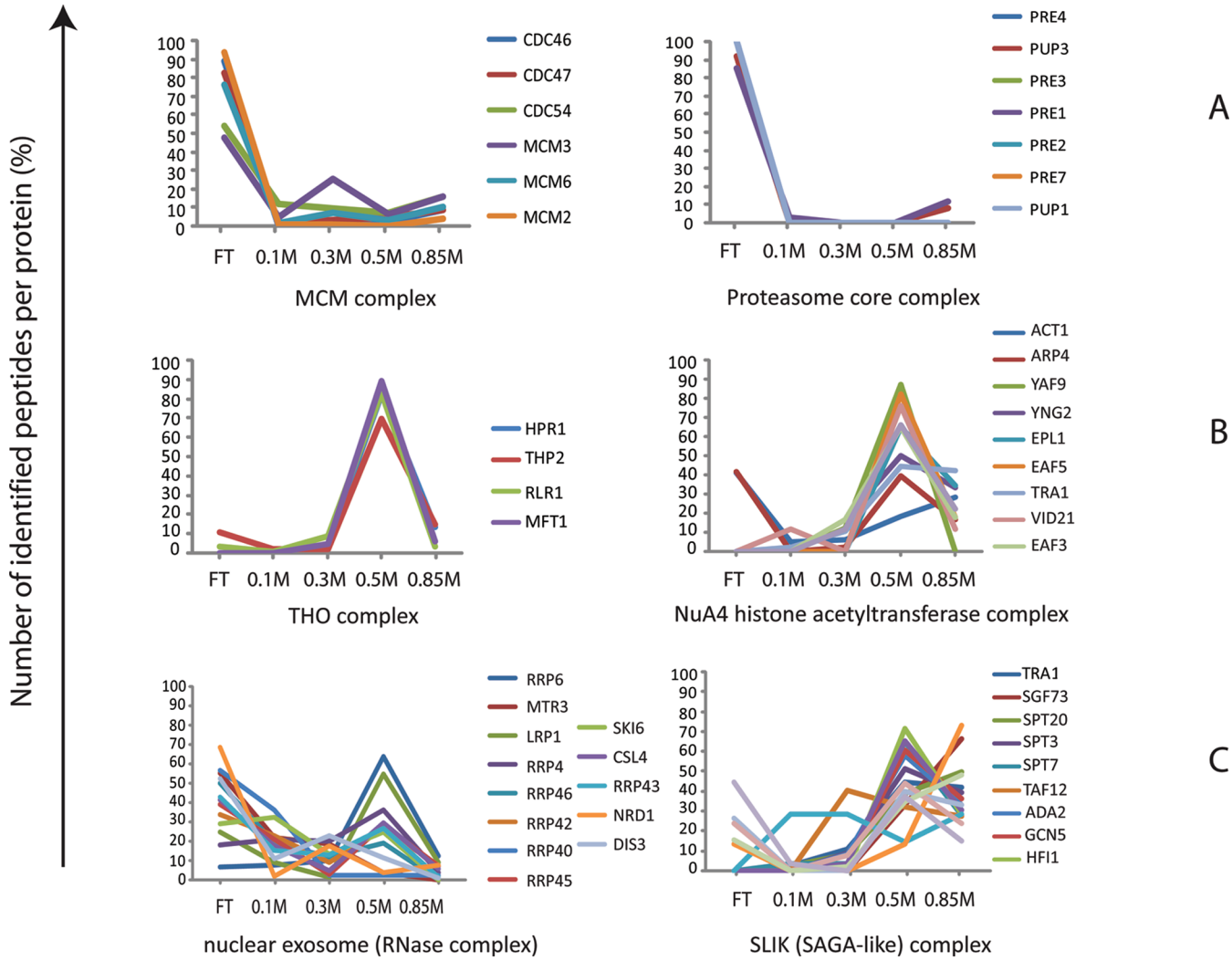


Figure 7. The number of identified peptides per protein (%) versus the elution fractions. Proteins eluting in the flow through (FT), 0.1, 0.3, 0.5, and 0.85 M NaCl were grouped into the complexes listed in the yeast SGD database to visualize the elution patterns of these proteins. Numerous proteins did not bind to the column and were identified in the flow through (A). Examples of proteins eluting in one fraction or eluting across all fractions are illustrated in panels B and C, respectively.

components, for instance, the general transcription factors such as TFIIC, TFIIE and TFIIF. These data reveal several important points: (1) Protein complexes purified previously using P11 elute at the expected salt step. For instance, TFIIE and TFIIF elute at 0.5 M, confirming the reproducibility of the method.[45] (2) There is a clear distinction in the functionality of complexes binding and not binding to phosphocellulose. Complexes that are retained are dominated by complexes involved in transcription regulation and chromatin structure, while complexes not involved in DNA interactions prevail in the flow-through (e.g., proteasome, nuclear pore). The distinction is not absolute, but presumably reflects cationic properties of the complex as a whole. (3) The majority of the complexes does not elute in one fraction, but has components eluting at different salt concentration. In particular, there are several complexes eluting in high-salt fractions that lose subunits in the flow-through (e.g., the replication factor C complex, the nucleosome, the cohesion complex), possibly reflecting their weak interaction. This is not surprising as some subunits may be associated to the complex transiently or by low-affinity interactions. (4) Some complexes elute across many fractions (e.g., mediator, SWI/SNF) indicating that either complexes disassemble due to increased ionic strength, or that one (or more) subunit was also present in an alternative form (Figure 7C). These latter two observations emphasize that the concept that intact complexes can be isolated quantitatively in one fraction is not necessarily true in all cases.

Discussion

The success of large-scale proteomics lies in the identification of large number of proteins in complex samples. This has been facilitated by recent developments in mass spectrometry and its direct coupling to reversed phase chromatography.[27,29,34,50−53] Nonetheless, in general, the complexity and dynamic range of proteins in biological samples is too high to obtain sufficient coverage, thus, requiring additional prefractionation techniques. Because of their specific mechanical and physical properties, yeast nuclei can be isolated with high purity and provide an excellent starting point for the characterization of the yeast nuclear proteome. An important criterion for organelle-specific analysis of any organism is the purity of the compartment in question. Critical to the isolation of highly enriched nuclei is that during the lysis of the spheroplasts intact

nuclei are released while fragmenting the connecting and surrounding cellular structures. The nuclei can then be further enriched on a sucrose gradient. The predominance of nuclear proteins (Figure 5) and the low amount of contaminating proteins originating from the plasma membrane, mitochondria and endoplasmic reticulum indicated that this was an efficient means of purifying and extracting yeast nuclei.

Since the nuclear proteome is still expected to be considerably complex, we utilized four well-established orthogonal prefractionation techniques at the peptide level (SCX, peptide IEF) and protein level (SDS-PAGE, and phosphocellulose P11 chromatography). The latter technique has been utilized as a first step in the isolation of transcription regulators and functional protein complexes,[45,49] but it has never been evaluated as a prefractionation technique in combination with the high dynamic range and high sensitivity mass spectrometry available today.

The combination of these prefractionation techniques led to the identification of 1889 proteins, 1032 (55%) of which were annotated as nuclear proteins in the yeast SGD database. These numbers are close to those obtained in a very recent study[25] identifying 2674 proteins, including 1234 annotated nuclear proteins (46%). Of these, 832 were in common with our study. The overlap between these data sets, but also the proteins that were uniquely identified in either study, could be caused by the differences experimental approaches and analytical platforms, including purification schemes for nuclei. It seems that our nuclear preparation was cleaner, given the significant lower levels of contamination of mitochondrial, ER and Golgi proteins: in our study, 52 out of 1889 proteins (2.7%) were mitochondrial (Figure 5), compared to ~300 out of 2674 (11%) by Mosley et al. For ER and Golgi proteins combined, these numbers are 116 out of 1889 proteins (6%) (Figure 5), and ~280 out of 2674 (10.5%),[25] respectively.

A closer look at our data set at specific subnuclear compartments revealed that 211 nucleolar proteins were identified, covering 91% of the known nucleolar proteins. Nucleolar proteins play an important role in cell growth and proliferation, as well as coordinating the synthesis and assembly of ribosomal subunits.[51,54,55] Moreover, nucleoli are highly conserved through evolution having 90% homology with human nucleolar proteins.[56,57]

Phosphocellulose P11 chromatography contributed the largest number of proteins to the total identifications with 724 unique proteins compared to the other techniques. In fact, phosphocellulose-P11 chromatography led to ~1.5 times the number of identifications (1687) compared to the other three techniques combined (1165). Moreover, 85% of the proteins found by all four techniques combined were identified by P11 alone. One can argue that the sample loaded onto the phosphocellulose P11 column was larger compared to the other techniques; however, equal amounts of protein were utilized for mass spectrometry analysis for each prefractionation technique. Classification by gene ontology revealed that the proteins identified after P11 separation are highly enriched in typical nuclear functionality (Supplementary Table 8). We found hundreds of proteins with very general nuclear functions such as gene expression, transcription as well as nuclear transport. Moreover, we noted that of several protein complexes, particularly those involved in transcription, all subunits were identified. We examined which protein complexes were present in our nuclei isolations and identified by P11 chromatography, and to what extent they remained intact during the separation process. By grouping all the proteins into their respective complexes, (parts of) 181 complexes were identified. In addition, we used the elution patterns of the subunits of each complex to classify complexes as an approximation of stability.

Although separation by P11 chromatography is not an alternative, for example, for TAP-purification of individual protein complexes, it provides a powerful means for additional prefractionation. Because of the dramatic enrichment, we have observed excessively high sequence coverage of various proteins, opening the way to analyze these proteins in greater detail with respect to, for example, PTMs under various growth conditions. In addition, it provides a complementary view to the classical application of P11 for the isolation of individual protein complexes: by combining this with modern and sensitive mass spectrometric detection, the less stable parts of such complexes can be detected over an extended range of salt fractions.

For the classification of all proteins identified in this study, we relied on Gene Ontology classification by cellular location and function as archived in SGD. This was used to visualize the number of established nuclear proteins as well as proteins residing in other compartments (Figure 5). From this classification, a number of issues can be raised. (1) The total number of identified nuclear proteins is considerable (1032), but not complete (1848 in SGD). Possible explanations are that we still have suffered from under-sampling, or that proteins were lost during the isolation procedures, especially with those proteins that have no stable interactions with nuclear structures. Another explanation is that not all nuclear proteins always reside in the nucleus. (2) The number of mitochondrial and especially ER proteins is very modest. This leads to the third observation, being the relatively high number of cytoplasmic proteins. Although some contamination could be due to cytosolic proteins sticking (specifically or aspecifically) to the cytosolic side of the nuclear envelope, it seems unlikely that the degree of contamination of the nuclear extract by soluble proteins from the cytosol is higher than that from membrane structures like the ER. Therefore, there might be other explanations for this observation. One of the most likely possibilities is that some of the cytosolic proteins translocate to the nucleus, at least for a fraction of the protein pool. Shuttling of proteins between the nucleus and the cytoplasm is a well-known phenomenon, and in fact, we identified several proteins flagged to be cytosolic in SGD, but with a documented function in the nucleus, of which the following represent a few examples. For instance, this applies to many proteins involved in mRNA transport and maturation.[59] Ribosomal proteins represent another prominent example, which are translated in the cytosol, but translocate to the nucleolus for ribosome assembly and maturation.[60−62] This may well explain the large number of ribosomal proteins in this study (Figure 5). Other protein complexes partitioning between nucleus and cytosol are the proteasome[63,64] and, for instance, the CCT complex.[65] In addition, individual proteins initially considered to be strictly cytosolic appear to fulfill specific roles in the nucleus. This applies to myosin and various forms of actin which are involved in regulation of transcription[66] and chromatin remodelling.[67] Myosin 1 and actin, as well as various Arps, were identified here. Another example is guanosine 5′-monophosphate synthetase (GMPS) which is a cytosolic protein but for which we showed before it also takes part in a nuclear complex involved in histone deubiquitilation in flies.[68] This might also apply to

another deubiquitinylase (Ubp3) assigned to be cytoplasmic but identified in our nuclear preparation along with its binding partner BRE5, which is in line with a nuclear function of this protein.[69] Other examples of *a priori* cytoplasmic proteins but with a documented function in the nucleus include PRK1 (phosphorylating Histone 3[70]), HXK1,[71] PFK1 and PFK2.[72] The phenomenon of extensive translocation of proteins even extends to typical plasma membrane proteins such as TOR1 and TOR2,[73,74] and YCK1 and YCK2[75] which were all identified here. This nonexhaustive list of examples shows that many typical cytosolic proteins can be nuclear. Since this might apply to only a small part of the total population, this may only be observed in targeted studies, while it is very likely to escape attention in large-scale studies. Many of the GO-annotations are based on such genome-wide studies,[8] leading to assignments that are only partially true or not complete. Thus, with the use of a highly enriched nuclear extract as starting material, we can postulate that many more proteins can be nuclear than appreciated before.

Conclusion

The characterization of the yeast nuclear proteome is a fundamental step toward the understanding of the important cellular processes occurring in yeast. We have shown that the use of four orthogonal peptide and protein prefractionation techniques can be utilized for the large-scale identification of these nuclear proteins. In particular, phosphocellulose P11 chromatography coupled with SDS-PAGE has proven to be a valuable prefractionation technique and an enrichment technique for proteins involved in various important nuclear processes. We foresee that this technique has a high potential for the detailed study of nucleo-cytoplasmic transport, and the mechanisms regulating this process. For instance, translocation can be influenced by or dependent on growth circumstance, stress condition or cell cycle-state, and can be regulated by phosphorylation of either cargo or transporter proteins.[14,76−78] It would be highly informative to study nuclear proteomes obtained under different conditions, preferably in a quantitative manner. The high sequence coverage that we have observed for many proteins would aid in the quantitation process, and would increase the chance to identify post-translational modifications.

Acknowledgment. We would like to thank Nune Kazaryan for providing technical assistance. This work was supported by The Netherlands Proteomics Centre.

Supporting Information Available: Supplementary Table 1, a description of the unique proteins identified using SDS-PAGE together with the number of unique peptides identified per protein and the protein sequence coverage. Supplementary Table 2, a description of the unique proteins identified using Strong Cation Exchange together with the unique number of peptides identified per protein and the protein sequence coverage. Supplementary Table 3 illustrates the peptides identified with peptide IEF binned into 48 fractions and its corresponding p*I*. The average p*I* and standard deviation is displayed per fraction. Supplementary Table 4, a description of all the unique proteins identified using peptide IEF together with the unique number of peptides identified per protein and the protein sequence coverage. Supplementary Table 5, a description of the unique proteins identified using phosphocellulose P11 chromatography in combination with SDS-PAGE together with the unique number of peptides identified per protein and the protein sequence coverage. Supplementary Table 6, protein localization of the proteins identified using SDS-PAGE, Strong Cation Exchange, Peptide IEF and phosphocellulose P11 chromatography. Supplementary Table 7, a combined list of the unique proteins identified using all 4 orthogonal prefractionation techniques, as well as a description of the annotated unique nuclear proteins identified in our study. Supplementary Table 8, the classification of proteins identified using phosphocellulose P11 chromatography utilizing Gene Ontology classification by component, function and process as archived in the Yeast SGD database. Supplementary Table 9, the proteins identified in subsequent P11 fractions grouped into the various protein complexes as defined in the Yeast SGD database. Supplementary Table 10, the elution profiles of the protein complexes in the phosphocellulose P11 experiment. The elution criteria to classify elution patterns of protein complexes were as follows: (1) ≥50% of the identified peptides of a protein was required to elute in one fraction to categorize that protein as eluting in that particular fraction and (2) ≥40% of the proteins in a protein complex was required to elute together in the same fraction to call that protein complex as eluting in that fraction. Protein complexes of which less than 40% of its protein components co-eluted in one fraction were considered to be spread across all fractions. Two proteins were minimally required to form part of a complex, and only if the two proteins were eluting in one fraction was the protein complex considered to be valid. Supplementary Figure 1, light microscopy images of nuclei preparation using a Leica DMLS2 and 10× 0.22 NA (upper panel) and 40× 0.65NA (lower panel) objective lenses; scale bars show 100 and 10 μm, respectively. Supplementary Figure 2, a plot of the nuclear proteins as a function of protein abundance based on estimations by Ghaemmaghami et al.[46] On the right axis, the relative number of nuclear proteins identified (as a percentage) relative to the total number of nuclear proteins (illustrated in green). The left axis illustrates the absolute number of nuclear proteins identified in specific copy number bins (illustrated in blue). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Pandey, A.; Choudhary, M. K.; Bhushan, D.; Chattopadhyay, A.; Chakraborty, S.; Datta, A.; Chakraborty, N. The nuclear proteome of chickpea (*Cicer arietinum* L.) reveals predicted and unexpected proteins. *J. Proteome Res.* **2006**, *5* (12), 3301–11.
(2) Zilberman, D.; Coleman-Derr, D.; Ballinger, T.; Henikoff, S. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **2008**, *456* (7218), 125–9.
(3) Blobel, G.; Wozniak, R. W. Proteomics for the pore. *Nature* **2000**, *403* (6772), 835–6.
(4) Sorokin, A. V.; Kim, E. R.; Ovchinnikov, L. P. Nucleocytoplasmic transport of proteins. *Biochemistry (Moscow)* **2007**, *72* (13), 1439–57.
(5) Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Sali, A.; Rout, M. P. The molecular architecture of the nuclear pore complex. *Nature* **2007**, *450* (7170), 695–701.
(6) Carrera, I.; Treisman, J. E. Message in a nucleus: signaling to the transcriptional machinery. *Curr. Opin. Genet. Dev.* **2008**, *18* (5), 397–403.
(7) Kumar, A.; Agarwal, S.; Heyman, J. A.; Matson, S.; Heidtman, M.; Piccirillo, S.; Umansky, L.; Drawid, A.; Jansen, R.; Liu, Y.; Cheung, K. H.; Miller, P.; Gerstein, M.; Roeder, G. S.; Snyder, M. Subcellular localization of the yeast proteome. *Genes Dev.* **2002**, *16* (6), 707–19.
(8) Huh, W. K.; Falvo, J. V.; Gerke, L. C.; Carroll, A. S.; Howson, R. W.; Weissman, J. S.; O'Shea, E. K. Global analysis of protein localization in budding yeast. *Nature* **2003**, *425* (6959), 686–91.

(9) Adam, S. A. The nuclear pore complex. *GenomeBiology* **2001**, *2* (9), REVIEWS0007.
(10) Li, J.; Meyer, A. N.; Donoghue, D. J. Nuclear localization of cyclin B1 mediates its biological activity and is regulated by phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (2), 502–7.
(11) Diehl, J. A.; Cheng, M.; Roussel, M. F.; Sherr, C. J. Glycogen synthase kinase-3beta regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev.* **1998**, *12* (22), 3499–511.
(12) Furnari, B.; Rhind, N.; Russell, P. Cdc25 mitotic inducer targeted by chk1 DNA damage checkpoint kinase. *Science* **1997**, *277* (5331), 1495–7.
(13) Ross-Macdonald, P.; Coelho, P. S.; Roemer, T.; Agarwal, S.; Kumar, A.; Jansen, R.; Cheung, K. H.; Sheehan, A.; Symoniatis, D.; Umansky, L.; Heidtman, M.; Nelson, F. K.; Iwasaki, H.; Hager, K.; Gerstein, M.; Miller, P.; Roeder, G. S.; Snyder, M. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **1999**, *402* (6760), 413–8.
(14) Hood, J. K.; Silver, P. A. Diverse nuclear transport pathways regulate cell proliferation and oncogenesis. *Biochim. Biophys. Acta* **2000**, *1471* (1), M31−41.
(15) Goffeau, A.; Barrell, B. G.; Bussey, H.; Davis, R. W.; Dujon, B.; Feldmann, H.; Galibert, F.; Hoheisel, J. D.; Jacq, C.; Johnston, M.; Louis, E. J.; Mewes, H. W.; Murakami, Y.; Philippsen, P.; Tettelin, H.; Oliver, S. G. Life with 6000 genes. *Science* **1996**, *274* (5287), 563–7.
(16) Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2003**, *2* (1), 43–50.
(17) de Godoy, L. M.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Frohlich, F.; Walther, T. C.; Mann, M. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **2008**, *455* (7217), 1251–4.
(18) Gavin, A. C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dumpelfeld, B.; Edelmann, A.; Heurtier, M. A.; Hoffman, V.; Hoefert, C.; Klein, K.; Hudak, M.; Michon, A. M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J. M.; Kuster, B.; Bork, P.; Russell, R. B.; Superti-Furga, G. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **2006**, *440* (7084), 631–6.
(19) Gavin, A. C.; Superti-Furga, G. Protein complexes and proteome organization from yeast to man. *Curr. Opin. Chem. Biol.* **2003**, *7* (1), 21–7.
(20) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W.; Figeys, D.; Tyers, M. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **2002**, *415* (6868), 180–3.
(21) Taylor, S. W.; Fahy, E.; Ghosh, S. S. Global organellar proteomics. *Trends Biotechnol.* **2003**, *21* (2), 82–8.
(22) Ho, E.; Hayen, A.; Wilkins, M. R. Characterisation of organellar proteomes: a guide to subcellular proteomic fractionation and analysis. *Proteomics* **2006**, *6* (21), 5746–57.
(23) Gauthier, D. J.; Lazure, C. Complementary methods to assist subcellular fractionation in organellar proteomics. *Expert Rev. Proteomics* **2008**, *5* (4), 603–17.
(24) Au, C. E.; Bell, A. W.; Gilchrist, A.; Hiding, J.; Nilsson, T.; Bergeron, J. J. Organellar proteomics to create the cell map. *Curr. Opin. Cell Biol.* **2007**, *19* (4), 376–85.
(25) Mosley, A. L.; Florens, L.; Wen, Z.; Washburn, M. P. A label free quantitative proteomic analysis of the Saccharomyces cerevisiae nucleus. *J. Proteomics* **2009**, *72* (1), 110–20.
(26) Krijgsveld, J.; Gauci, S.; Dormeyer, W.; Heck, A. J. In-gel isoelectric focusing of peptides as a tool for improved protein identification. *J. Proteome Res.* **2006**, *5* (7), 1721–30.
(27) Reinders, J.; Zahedi, R. P.; Pfanner, N.; Meisinger, C.; Sickmann, A. Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. *J. Proteome Res.* **2006**, *5* (7), 1543–54.
(28) Essader, A. S.; Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr. A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. *Proteomics* **2005**, *5* (1), 24–34.
(29) Lemeer, S.; Pinkse, M. W.; Mohammed, S.; van Breukelen, B.; den Hertog, J.; Slijper, M.; Heck, A. J. Online automated in vivo zebrafish phosphoproteomics: from large-scale analysis down to a single embryo. *J. Proteome Res.* **2008**, *7* (4), 1555–64.
(30) Yaneva, M.; Tempst, P. Isolation and mass spectrometry of specific DNA binding proteins. *Methods Mol. Biol.* **2006**, *338*, 291–303.
(31) Cargile, B. J.; Sevinsky, J. R.; Essader, A. S.; Stephenson, J. L., Jr.; Bundy, J. L. Immobilized pH gradient isoelectric focusing as a first-dimension separation in shotgun proteomics. *J. Biomol. Tech.* **2005**, *16* (3), 181–9.
(32) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–7.
(33) Wilson-Grady, J. T.; Villen, J.; Gygi, S. P. Phosphoproteome analysis of fission yeast. *J. Proteome Res.* **2008**, *7* (3), 1088–97.
(34) Wolters, D. A.; Washburn, M. P.; Yates, J. R., III. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73* (23), 5683–90.
(35) Hung, C. W.; Kubler, D.; Lehmann, W. D. pI-based phosphopeptide enrichment combined with nanoESI-MS. *Electrophoresis* **2007**, *28* (12), 2044–52.
(36) Heller, M.; Ye, M.; Michel, P. E.; Morier, P.; Stalder, D.; Junger, M. A.; Aebersold, R.; Reymond, F.; Rossier, J. S. Added value for tandem mass spectrometry shotgun proteomics data validation through isoelectric focusing of peptides. *J. Proteome Res.* **2005**, *4* (6), 2273–82.
(37) Thomas, M. C.; Chiang, C. M. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **2006**, *41* (3), 105–78.
(38) Rout, M. P.; Kilmartin, J. V. Components of the yeast spindle and spindle pole body. *J. Cell Biol.* **1990**, *111* (5 Pt. 1), 1913–27.
(39) Rout, M. P.; Kilmartin, J. V. *Preparation of Yeast Spindle Pole Bodies*; Academic Press: London, 1994; Vol. 1, pp 605−612.
(40) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **2003**, *75* (3), 663–70.
(41) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74* (20), 5383–92.
(42) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75* (17), 4646–58.
(43) Gauci, S.; van Breukelen, B.; Lemeer, S. M.; Krijgsveld, J.; Heck, A. J. A versatile peptide pI calculator for phosphorylated and N-terminal acetylated peptides experimentally tested using peptide isoelectric focusing. *Proteomics* 2008.
(44) Cargile, B. J.; Talley, D. L.; Stephenson, J. L., Jr. Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides. *Electrophoresis* **2004**, *25* (6), 936–45.
(45) Flores, O.; Lu, H.; Reinberg, D. Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J. Biol. Chem.* **1992**, *267* (4), 2786–93.
(46) Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. Global analysis of protein expression in yeast. *Nature* **2003**, *425* (6959), 737–41.
(47) Pontremoli, S.; Melloni, E.; Michetti, M.; Salamino, F.; Sparatore, B.; Horecker, B. L. Limited proteolysis of liver aldolase and fructose 1,6-bisphosphatase by lysosomal proteinases: effect on complex formation. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79* (8), 2451–4.
(48) Zhang, N.; Lu, X.; Legerski, R. J. Partial reconstitution of human interstrand cross-link repair in vitro: characterization of the roles of RPA and PCNA. *Biochem. Biophys. Res. Commun.* **2003**, *309* (1), 71–8.
(49) Chifflot, S.; Sommer, P.; Hartmann, D.; Stussi-Garaud, C.; Hirth, L. Replication of alfalfa mosaic virus RNA: evidence for a soluble replicase in healthy and infected tobacco leaves. *Virology* **1980**, *100* (1), 91–100.
(50) Sanders, S. L.; Jennings, J.; Canutescu, A.; Link, A. J.; Weil, P. A. Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. *Mol. Cell. Biol.* **2002**, *22* (13), 4723–38.
(51) Andersen, J. S.; Lam, Y. W.; Leung, A. K.; Ong, S. E.; Lyon, C. E.; Lamond, A. I.; Mann, M. Nucleolar proteome dynamics. *Nature* **2005**, *433* (7021), 77–83.
(52) Reinders, J.; Sickmann, A. Proteomics of yeast mitochondria. *Methods Mol. Biol.* **2007**, *372*, 543–57.

(53) Buhr, N.; Carapito, C.; Schaeffer, C.; Kieffer, E.; Van Dorsselaer, A.; Viville, S. Nuclear proteome analysis of undifferentiated mouse embryonic stem and germ cells. *Electrophoresis* **2008**, *29* (11), 2381–90.
(54) Andersen, J. S.; Lyon, C. E.; Fox, A. H.; Leung, A. K.; Lam, Y. W.; Steen, H.; Mann, M.; Lamond, A. I. Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **2002**, *12* (1), 1–11.
(55) Coute, Y.; Burgess, J. A.; Diaz, J. J.; Chichester, C.; Lisacek, F.; Greco, A.; Sanchez, J. C. Deciphering the human nucleolar proteome. *Mass Spectrom. Rev.* **2006**, *25* (2), 215–34.
(56) Boisvert, F. M.; van Koningsbruggen, S.; Navascues, J.; Lamond, A. I. The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (7), 574–85.
(57) Leung, A. K.; Trinkle-Mulcahy, L.; Lam, Y. W.; Andersen, J. S.; Mann, M.; Lamond, A. I. NOPdb: Nucleolar Proteome Database. *Nucleic Acids Res.* **2006**, *34* (Database issue), D218−20.
(58) Strasser, K.; Masuda, S.; Mason, P.; Pfannstiel, J.; Oppizzi, M.; Rodriguez-Navarro, S.; Rondon, A. G.; Aguilera, A.; Struhl, K.; Reed, R.; Hurt, E. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **2002**, *417* (6886), 304–8.
(59) Dreyfuss, G.; Kim, V. N.; Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* **2002**, *3* (3), 195–205.
(60) Kruger, T.; Zentgraf, H.; Scheer, U. Intranucleolar sites of ribosome biogenesis defined by the localization of early binding ribosomal proteins. *J. Cell Biol.* **2007**, *177* (4), 573–8.
(61) Lam, Y. W.; Lamond, A. I.; Mann, M.; Andersen, J. S. Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Curr. Biol.* **2007**, *17* (9), 749–60.
(62) Zemp, I.; Kutay, U. Nuclear export and cytoplasmic maturation of ribosomal subunits. *FEBS Lett.* **2007**, *581* (15), 2783–93.
(63) von Mikecz, A.; Chen, M.; Rockel, T.; Scharf, A. The nuclear ubiquitin-proteasome system: visualization of proteasomes, protein aggregates, and proteolysis in the cell nucleus. *Methods Mol. Biol.* **2008**, *463*, 191–202.
(64) Rockel, T. D.; Stuhlmann, D.; von Mikecz, A. Proteasomes degrade proteins in focal subdomains of the human cell nucleus. *J. Cell Sci.* **2005**, *118* (Pt. 22), 5231–42.
(65) Joly, E. C.; Tremblay, E.; Tanguay, R. M.; Wu, Y.; Bibor-Hardy, V. TRiC-P5, a novel TCP1-related protein, is localized in the cytoplasm and in the nuclear matrix. *J. Cell Sci.* **1994**, *107* (Pt. 10), 2851–9.
(66) de Lanerolle, P.; Johnson, T.; Hofmann, W. A. Actin and myosin I in the nucleus: what next. *Nat. Struct. Mol. Biol.* **2005**, *12* (9), 742–6.
(67) Olave, I. A.; Reck-Peterson, S. L.; Crabtree, G. R. Nuclear actin and actin-related proteins in chromatin remodeling. *Annu. Rev. Biochem.* **2002**, *71*, 755–81.
(68) van der Knaap, J. A.; Kumar, B. R.; Moshkin, Y. M.; Langenberg, K.; Krijgsveld, J.; Heck, A. J.; Karch, F.; Verrijzer, C. P. GMP synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. *Mol. Cell* **2005**, *17* (5), 695–707.
(69) Kvint, K.; Uhler, J. P.; Taschner, M. J.; Sigurdsson, S.; Erdjument-Bromage, H.; Tempst, P.; Svejstrup, J. Q. Reversal of RNA polymerase II ubiquitylation by the ubiquitin protease Ubp3. *Mol. Cell* **2008**, *30* (4), 498–506.
(70) Metzger, E.; Yin, N.; Wissmann, M.; Kunowska, N.; Fischer, K.; Friedrichs, N.; Patnaik, D.; Higgins, J. M.; Potier, N.; Scheidtmann, K. H.; Buettner, R.; Schule, R. Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation. *Nat. Cell Biol.* **2008**, *10* (1), 53–60.
(71) Cho, Y. H.; Yoo, S. D.; Sheen, J. Regulatory functions of nuclear hexokinase1 complex in glucose signaling. *Cell* **2006**, *127* (3), 579–89.
(72) Mukhtar, M. H.; Payne, V. A.; Arden, C.; Harbottle, A.; Khan, S.; Lange, A. J.; Agius, L. Inhibition of glucokinase translocation by AMP-activated protein kinase is associated with phosphorylation of both GKRP and 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. *Am. J. Physiol.: Regul., Integr. Comp. Physiol.* **2008**, *294* (3), R766−74.
(73) Li, H.; Tsang, C. K.; Watkins, M.; Bertram, P. G.; Zheng, X. F. Nutrient regulates Tor1 nuclear localization and association with rDNA promoter. *Nature* **2006**, *442* (7106), 1058–61.
(74) Tsang, C. K.; Zheng, X. F. TOR-in(g) the nucleus. *Cell Cycle* **2007**, *6* (1), 25–9.
(75) Wang, X.; Hoekstra, M. F.; DeMaggio, A. J.; Dhillon, N.; Vancura, A.; Kuret, J.; Johnston, G. C.; Singer, R. A.; Prenylated isoforms of yeast casein kinase, I. including the novel Yck3p, suppress the gcs1 blockage of cell proliferation from stationary phase. *Mol. Cell. Biol.* **1996**, *16* (10), 5375–85.
(76) Poon, I. K.; Jans, D. A. Regulation of nuclear transport: central role in development and transformation. *Traffic* **2005**, *6* (3), 173–86.
(77) Nakielny, S.; Dreyfuss, G. Transport of proteins and RNAs in and out of the nucleus. *Cell* **1999**, *99* (7), 677–90.
(78) Fried, H.; Kutay, U. Nucleocytoplasmic transport: taking an inventory. *Cell. Mol. Life Sci.* **2003**, *60* (8), 1659–88.

PR9000948