

Rapid Peptide in-Solution Isoelectric Focusing Fractionation for Deep Proteome Analysis

Mastrobuoni G, Zasada C, Bindel F, Aeberhard L and Kempa S*

Integrative proteomics and metabolomics, Berlin Institute for Medical Systems Biology at the Max Delbrück Center for Molecular Medicine, Robert-Roessle-Strasse 10, 13125 Berlin, Germany

Abstract

The interplay between resolution, accuracy, sensitivity and speed of the mass spectrometer, as well as the complexity of the peptide mixture in relation to chromatographic separation and the analysis time finally determines the number of identified proteins within a 'shotgun' proteomic study. The improvement of one of these parameters can enhance the quality of the proteome analysis. Here we evaluated the technique of in-solution isoelectric focusing (IEF) for pre-fractionation of tryptic peptides prior to LC-MS/MS-based proteomics analysis. In-solution IEF turned out to be a simple and fast method for peptide fractionation prior to LC-MS analysis. By adapting the experimental procedures, this approach enabled identification of more than 44,000 peptides belonging to 5,800 proteins in less than 48 working hours, from protein extraction until the end of LC-MS analysis.

This technique was applied successfully to analyze the proteomes of mammalian cells and different model organisms, without additional efforts or special technical equipment. The in-solution IEF of peptides is very robust and can be applied in combination with different extraction procedures. The high number of identified peptides using a standard LC-MS system led to average protein coverage of 25%. Such a high average number of identified peptides per protein improved the discrimination of protein species as isoforms or splice variants. Thus, in solution IEF is a fast and robust alternative to gel-based proteomics or other gel free fractionation techniques upstream to LC-MS/MS analysis. The reduction of processing time and the high performance of this technique can speed up deep proteomics analyses significantly.

Keywords: LC-MS/MS; isoelectric focusing; peptide fractionation; shotgun proteomics; high resolution mass spectrometry; LTQ-Velos Orbitrap; SILAC; Microrotofor

Introduction

Technical advancements of mass spectrometry instrumentation are continuously improving the sensitivity, resolution, accuracy and the dynamic range of MS-based proteomics analyses. Thus, the number of identified peptides, and subsequently proteins, determined by the untargeted "shotgun" proteomics approach constantly increases [1,2].

The identification of tens of thousands of peptides and thousands of proteins can be routinely achieved with an Orbitrap mass spectrometer coupled with nano High Performance Liquid Chromatography (nHPLC) within a few hours of analysis time. Recently, using an exceptional chromatographic setup, including ultra-high performance liquid chromatography and ultra-low nano flow rates, Thakur et al. reported the identification of more than 4,500 protein and 30,000 peptides from human cell line in a single 8h chromatographic run [3].

However, the complexity and dynamic range of the protein sample, especially of higher organisms, is a challenging task and thus only a fraction of the proteome can be detected and identified [4,5]. The most abundant proteins are always identified in shotgun proteomic experiments; lower abundance proteins are often not reproducibly identified and quantified from experiment to experiment.

One possibility to tackle this issue and to increase protein coverage is fractionating the sample before protein digestion. This can be achieved by organelle fractionation and other protein fractionation methods. A common approach is the so called GeLC-MS/MS approach [6,7], where the protein sample is loaded on a Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE), the lane is cut in several slices, subsequently proteins are *in gel* digested [8] and the resulting peptide mixtures are analyzed on the Liquid Chromatography-Mass Spectrometry (LC-MS) system.

The advantage of the GeLC-MS/MS method is that the information about the relative protein molecular weight for each identified protein is retained; the drawbacks are that only limited amount of protein can be separated, the low yield of the digestion process and the susceptibility to contaminations, beside the laboriousness of the protocol.

An alternative solution is fractionating the peptide mixture after protein digestion, in order to reduce the complexity of the sample prior to LC-MS/MS analysis. A possible approach makes use of the multidimensional protein identification technology (MudPIT) [9,10]; with this approach the peptide sample is first loaded on a strong cation exchange column, from where it is stepwise eluted and directly separated on a reverse-phase column before mass spectrometer detector. The entire process is completely automated but it requires a dedicated LC system and a time-consuming setup. Alternatively, the peptide sample can be fractionated by isoelectric-focusing (IEF); available systems make use of immobilized pH gradients (IPG) and allow the separation of the sample in up to 24 fractions before the LC-MS analysis [11,12]. In-solution IEF allows instead a faster fractionation; although it has been since long time used for protein fractionation [13-15] its application for

*Corresponding author: Kempa S. Integrative proteomics and metabolomics, Berlin Institute for Medical Systems Biology at the Max Delbrück Center for Molecular Medicine, Robert-Roessle-Strasse 10, 13125 Berlin, Germany; Tel: +49 (0) 3094063033; E-mail: Stefan.kempa@mdc-berlin.de

Received September 01, 2014; Accepted September 18, 2014; Published September 26, 2014

Citation: Mastrobuoni G, Zasada C, Bindel F, Aeberhard L, Kempa S (2014) Rapid Peptide in-Solution Isoelectric Focusing Fractionation for Deep Proteome Analysis. J Chromatogr Sep Tech 5: 240. doi:10.4172/2157-7064.1000240

Copyright: © 2014 Mastrobuoni G, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

peptide fraction was limited to simple proteomes [16] or was used for phospho-peptide enrichment [17].

In our laboratory, we were interested in a fractionation technique which allows a simple and fast fractionation of peptides without limitation for the amount of input material and which enables high identification rates. Here we applied for the first time the in-solution isoelectric focusing to highly complex peptide mixtures and we systematically compared the performances and the results of this technique with those obtained with other published methods. We applied this technique to separate peptide mixtures from various organisms of different complexity and performed deep proteome analyses. The applied technique worked fast and robustly and the amounts of input material could be adjusted without negative impact on the overall performances. Furthermore, our established workflow allowed deep proteome analyses within 48 hours from protein extraction till end of LC-MS/MS analyses. The analyses resulted in high numbers of identified proteins with remarkable sequence coverage. The workflow was also already used to validate the *de novo* assembled transcriptome of planarian [18] or for high proteome coverage in *C. elegans* [19].

Material and Methods

HEK cells and yeast culture and protein extraction

HEK293 cells were cultivated in Dulbecco's Modified Eagle Medium (DMEM [Invitrogen]) supplemented with 10% fetal bovine serum with 2 mM Glutamine and 0.9 g/L glucose. Isotopically labeled HEK293 cells were grown in DMEM supplemented with dialyzed fetal bovine serum and L-¹³C₆¹⁵N₄-arginine (Arg10) and L-¹³C₆¹⁵N₂-lysine (Lys8) replacing the natural amino acids. Heavy cells were cultured in the Stable Isotope Labeling by Amino acids in Cell culture (SILAC) medium approximately for seven generation to reach complete labeling. Confluent cultures were trypsinized and 1×10⁶ cells were pelleted by centrifugation.

Heavy and light cells pellets were mixed and lysed in urea buffer (8 M Urea, 100 mM Tris-HCl, pH 8.3) and briefly sonicated. Cell debris was removed by centrifugation (14000g, 5 min). Protein concentration was then measured by Bradford colorimetric assay [20].

Yeast strain FY-4 was cultivated in standard yeast Synthetic Defined (SD) medium (Sigma), cells were filtered on nylon membrane (1µm pore size, 3M) and cells were disrupted using a CryoMill device (Retsch). Proteins were resolved using urea containing buffer.

Protein concentration was then measured by Bradford colorimetric assay [20] before enzymatic digestion.

In-solution digestion

Disulfide bridges were then reduced in dithiothreitol (DTT) 2mM for 30 minutes at 25°C and successively free cysteines were alkylated in 11 mM iodoacetamide for 20 minutes at room temperature in the darkness. LysC digestion was performed by adding LysC (Wako) in a ratio 1:40 (w/w) to the sample and incubating it for 18 hours under gentle shaking at 30°C. After LysC digestion, the samples were diluted 3 times with 50 mM ammonium bicarbonate solution, 7 µL of immobilized trypsin (Applied Biosystems) were added and samples were incubated 4 hours under rotation at 30°C. Digestion was stopped by acidification with 10 µL of trifluoroacetic acid and removal of trypsin beads by centrifugation.

The resulting peptide mixtures were loaded on Empore cartridges (3M) following the instructions of the manufacturer and eluted with

70% acetonitrile.

In-solution isoelectric focusing IEF

After partial removal by evaporation of the acetonitrile used for the elution from the Empore cartridges, the peptide mixture was diluted up to 2.5 mL with MilliQ water and 150 µL of ampholyte solution (pH range 3-10, 40% w/w, Bio-Rad) were added. The sample was then loaded into the focusing chamber of the Microrotofor device. Isoelectric focusing was performed following the manufacturer instruction. Briefly, constant 1 W power current was applied, while the maximum allowed voltage and current were set to 500V and 10 mA, respectively. After reaching stable voltage (~2.5 h time), the current was applied for further 30 minutes before collecting the fractionated peptides from the focusing chamber.

The resulting fractions were desalted on STAGE Tips (max 15 µg per StageTip), dried and reconstituted to 25 µL of 0.5 % acetic acid in water [21].

LC-MS analysis

5 microliters of the desalted peptides were injected on a LC-MS/MS system (Agilent 1200 [Agilent Technologies] and LTQ-Orbitrap Velos [Thermo]). Each analysis was run in duplicate. For the chromatographic separation, a linear binary gradient, ranging from 5% to 40% of organic buffer (80% acetonitrile, 20% water, 0.1% formic acid), in 155 or 240 minutes was used. As aqueous solvent, 5% acetonitrile in water with 0.1% formic acid was used. 20 cm long capillary (75 µm inner diameter), packed in-house with 3 µm C18 beads (ReprosilPur C18 AQ, Dr. Maisch), was used as chromatographic column. At one end of the capillary a nanospray tip was generated using a laser puller (P-2000 Laser Based Micropipette Puller, Sutter Instruments), allowing fretless packing.

The nanospray source of the mass spectrometer was operated with a spray voltage of 1.9 kV and with an ion transfer tube temperature of 260°C. Data were acquired in data dependent mode, with one survey MS scan in the Orbitrap mass analyzer (resolution 60000 at m/z 400), followed by up to 20 MS/MS scans in the ion trap on the most intense ions (intensity threshold=500 counts). Once selected for fragmentation, ions were excluded from further selection for the next 30 seconds, in order to increase the number of new sequencing events.

Data processing and analysis

Raw data were analyzed using the MaxQuant proteomics pipeline v1.1.1.36 and the built-in Andromeda search engine [22,23] with the International Protein Index Human version 3.71 database or the Saccharomyces Genome Database, version from 5 Jan 2010. Carbamidomethylation of cysteines was chosen as fixed modification, oxidation of methionine and acetylation of N-terminus were chosen as variable modifications. 2 missed cleavage site were allowed and peptide tolerance was set to 7 ppm. The search engine peptide assignments were filtered at 1% false discovery rate at both the peptide and protein level. The 'match between runs' feature was not enabled, 'second peptide' feature was enabled, while other parameters were left as default. For SILAC samples, two ratio counts were set as threshold for quantification.

Data analysis was performed using custom tools in Microsoft Excel and R!. Gene Ontology (GO) analysis was performed using David tool [24]. Enrichment of specific GO terms in the urea digestion/in-solution IEF dataset was calculated using the entire human genome as background; enrichment GO term enrichment was calculated using the entire human genome as background.



Figure 1: Schematic view on the workflow and timeline of the urea digestion/in-solution IEF protocol.

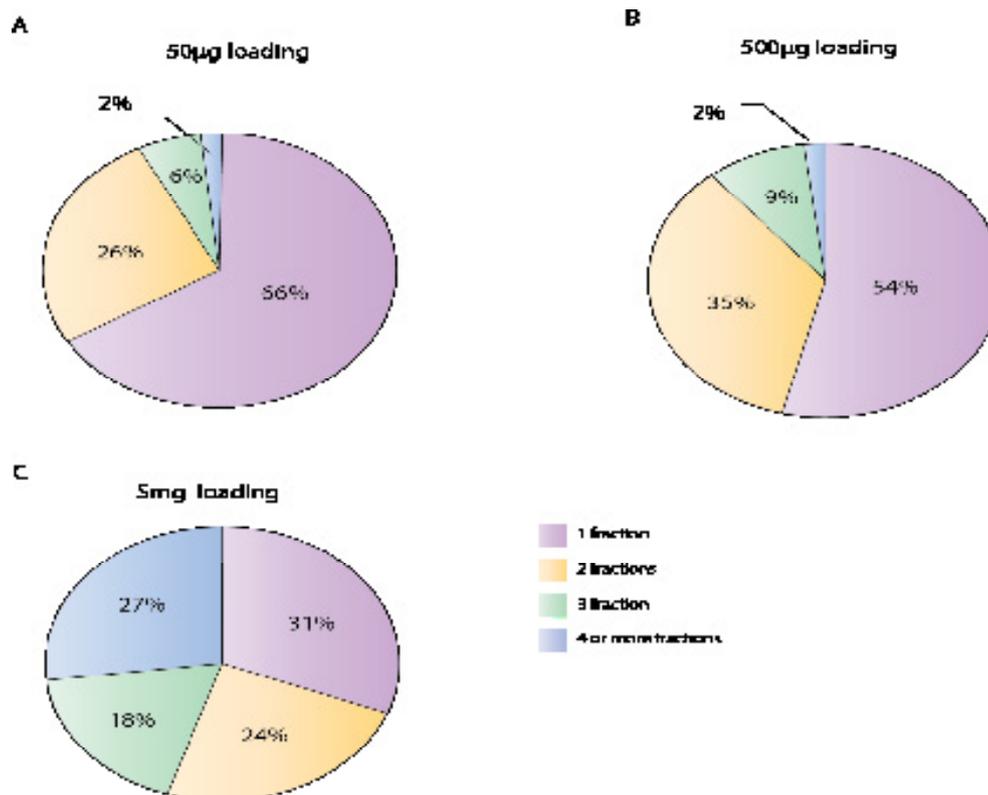


Figure 2: Influence of loaded sample amount on the focusing performance. 50µg (A) and 500µg (B) of HEK cells or 5 mg of yeast protein digests (C) were fractionated on the in-solution IEF device. Percentages represent the fraction of peptides identified in only one well or in two, three or four and more fractions.

Theoretical isoelectric point of identified peptides was calculated by a Perl script using the available tools in the BioPerl package (<http://search.cpan.org/dist/BioPerl-1.6.1/>). Default pK values for the charged aminoacids were used.

Results and discussion

In contrast to classical IPG isoelectric focusing techniques, which are time consuming and require higher applied voltages (500-4000V and 24hours separation time), in-solution IEF proceeds with lower voltages and finishes within 3 hours. Furthermore, the amount of input material can vary from few micrograms to several milligrams, offering a wide range of applications. The time required for the entire workflow we describe here is considerably reduced compared to GeLC-MS because the starting material is directly digested in solution and then loaded on the device for fractionation; the resulting fractions are purified using Stage Tips [21] and are ready for LC-MS/MS analysis (Figure 1).

To investigate the performance of the in-solution IEF technique we analyzed SILAC-labeled HEK293 cells and yeast proteomes and evaluated the number of identified proteins and peptides, sequence

coverage and focusing efficiency using different amounts of peptides.

Evaluation of fractionation efficiency

To investigate the influence of sample amount on the focusing efficiency, we performed three separations using different amount of input spanning on two orders of magnitude. In particular we fractionated 50 and 500 µg of HEK cell digest or 5 mg of yeast digest.

As expected and already observed for other isoelectric focusing techniques, increasing the loading corresponded to a decreased peptide focusing [12].

Using 50 µg of sample, 66% of peptides were 'perfectly' focused, being identified in only one of the ten fractions, while additional 26% were found in two fractions (Figure 2A and supplementary Table 1). A ten-fold increase of the loaded sample did not significantly compromise the focusing efficiency, with 54% of peptides perfectly focused and other 35% found in two adjacent fractions (Figure 2B and supplementary table 2). Only in the case of an extreme sample loading (5 mg) the focusing performance dropped down, with only 31% of

	Replicate 1	Replicate 2	Combined
Peptides	44742	43931	50672
Proteins (1 unique peptide)	5884	5846	6150
Proteins (2 peptides*)	5072	5036	5834

In the combined column are reported the numbers of protein and peptides identified at least in one of the two runs. (*at least one unique peptide)

Table 1: Protein and peptide identification after in solution IEF of 500µg of HEK cells peptides.

	155 min gradient	240 min gradient	Combined
Peptides	17335	19115	22353
Proteins (1 unique peptide)	3276	3579	3782
Proteins (2 peptides*)	2525	2778	3384

In the combined column are reported the numbers of protein and peptides identified at least in one of the two runs. (*at least one unique peptide)

Table 2: Protein and peptide identification after in solution IEF of 50µg of HEK cells peptides.

	Replicate 1
Peptides	22102
Proteins (1 unique peptide)	3226
Proteins (2 peptides)	2769

(*at least one unique peptide)

Table 3: Protein and peptide identification after in solution IEF of 5 mg of yeast peptides.

perfectly focused peptides (Figure 2C and supplementary table 3). An *in silico* calculation of the isoelectric point of the identified peptides showed a good correlation ($R^2= 0.86$), with the expected pH of the fractions in which they were identified, supporting the effectiveness of the fractionation (supplementary Figure1).

Since the in-solution pH gradient cannot be as accurate as an immobilized pH gradient and is susceptible of diffusion phenomena, the resolving power of in-solution IEF is consequently lower than IPG-IEF, but the resulting fractionation still can be considered satisfactory for the downstream LC-MS/MS analysis. We compared then our results with those obtained with IPG-IEF reported by Hubner et al. [12].

In their work peptides were separated into 24 fractions; for small sample loadings (50 µg) the focusing efficiency of the IPG system is superior, with 99% of peptides focused in one or two fractions. However, the in-solution IEF is less sensitive to increase of sample loadings. The focusing results are better than those observed with the immobilized gradient (89% versus 54% focused in one or two fractions) when 500 µg of peptides were separated. Only the separation of 5 mg of peptides led to a clear decrease of the focusing efficiency, but resulting still comparable with those of the IPG-IEF system with 500 µg loaded (55% of peptides focused in one or two fractions for both methods).

Evaluation of HEK cells proteome coverage

The fractions from the in-solution IEF of 500 µg of peptides were analyzed within one day of measurement time (10 fractions, 155 minutes LC-MS/MS analysis for each fraction). This analysis resulted in the identification of 44,742 distinct peptides mapping to 5,884 proteins with at least one unique peptide. In average 7.6 peptides per protein were identified resulting in 24.9% sequence coverage.

Repetition of the analysis of each fraction led to a modest increase in the number of the identified peptides and proteins, suggesting that already a single analysis covers most of the detectable proteins with this approach (Table 1 and supplementary Table 4).

The use of reduced amounts of peptide material resulted in lower

proteome coverage, even if a better focusing performance could be achieved. Using 50 µg of peptides led to identification of 3,276 proteins and 17,335 peptides, with an average of 5.2 peptides per protein and sequence coverage of 19%.

In order to test whether the use of a longer LC gradient may lead to a significant increase of protein identification, we repeated the analysis using a 240 minutes gradient. This analysis identified only 10% more proteins and the combination with the 155 minutes gradient analyses yielded only another 5% increase (Table 2 and supplementary Table 5). Instead, considering peptide identification the increase was 28%, with direct impact on the number of proteins identified with 2 or more peptides. In fact, in the combined dataset this number increased by 34%.

These results thus suggest that the single analysis of the in-solution IEF fractions on a 155 minutes gradient can already cover most of peptides detectable by the mass spectrometer; use of a longer gradient or replicate injection does not increase significantly the protein identifications, but enhances sequence coverage.

Protein isoforms and splicing variants can play a crucial role in regulating the normal activity of the cell. High peptide coverage is then fundamental for the discrimination of proteins that share a large part of the sequence. Interestingly, in our dataset we could detect an isoform of pyruvate kinase (Uniprot ID Q504U3) that currently has been observed only at transcript level. This finding is intriguing because this isoform constitute a shortened version of the isoform PKM1 (Uniprot ID P1618-2), which normally is expressed in organs that are strongly dependent upon a high rate of energy regeneration, such as muscle and brain. PKM1 presence was not reported in HEK cells and actually we could not identify any of its unique peptides. Since the protein Q504U3 was identified by a single unique peptide, the MS/MS spectrum was validated manually (supplementary Figure 2).

In addition to that, high peptide identification improves also the quantification through SILAC technology. In our experiments we could quantify with a minimum of two Heavy/Light counts up to 92% of the identified proteins (supplementary Table 6).

Recently, Wiśniewski et al. published the largest human proteome dataset obtained within a single experiment, using a detergent-based filter-aided sample preparation (FASP) and IPG-IEF fractionation [25]. Although the reported workflow differs from ours and we analyzed a SILAC labeled sample, resulting in doubled complexity, the results constitute an excellent term of comparison to further evaluate the urea extraction/in-solution IEF procedure and the quality of the results.

Considering the proteins identified in the combined dataset of the 500 µg HEK cell sample, we could find that the FASP dataset and in-solution IEF dataset overlap substantially (Figure 3B), whereas a substantial large proportion of identified peptides (45% of the FASP dataset and 56% of the in-solution IEF one) is specific for one of the two workflows (Figure 3A). Examination of the peptide sequences identified by only one of the two methods does not reveal a significant difference in the amino acid composition. On the contrary, the average and median lengths of peptides identified only in the FASP dataset are longer than those found only in the in solution IEF (14.6 versus 12.3 amino acids), while the peptides in common have an intermediate length of 13.4 amino acids. This may be due to the higher hydrophobicity of longer peptides, which are better covered by the FASP protocol.

Gene Ontology analysis of the identified proteins in the FASP dataset shows a significant enrichment of membrane proteins [25], while the proteins only identified by urea extraction/in-solution IEF

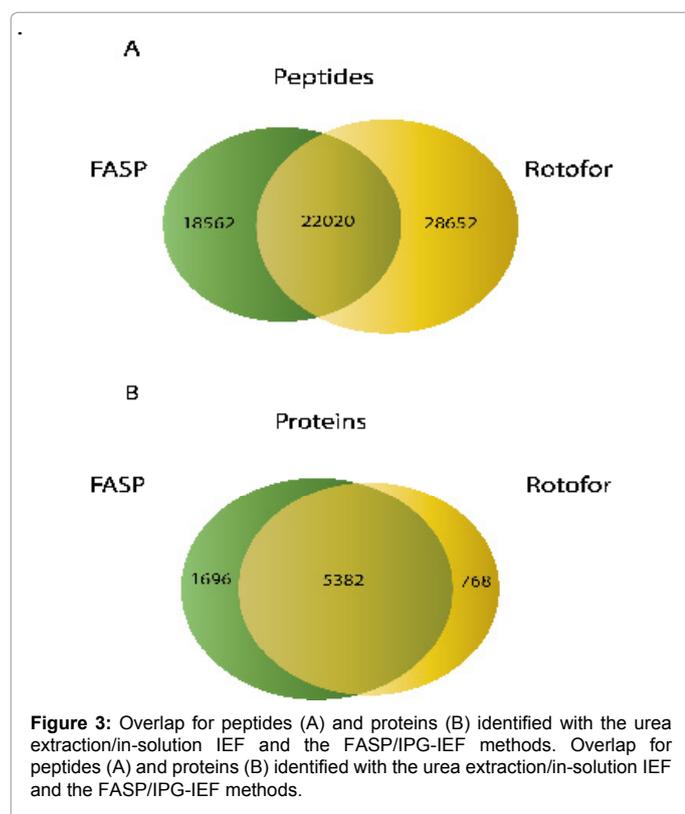


Figure 3: Overlap for peptides (A) and proteins (B) identified with the urea extraction/in-solution IEF and the FASP/IPG-IEF methods. Overlap for peptides (A) and proteins (B) identified with the urea extraction/in-solution IEF and the FASP/IPG-IEF methods.

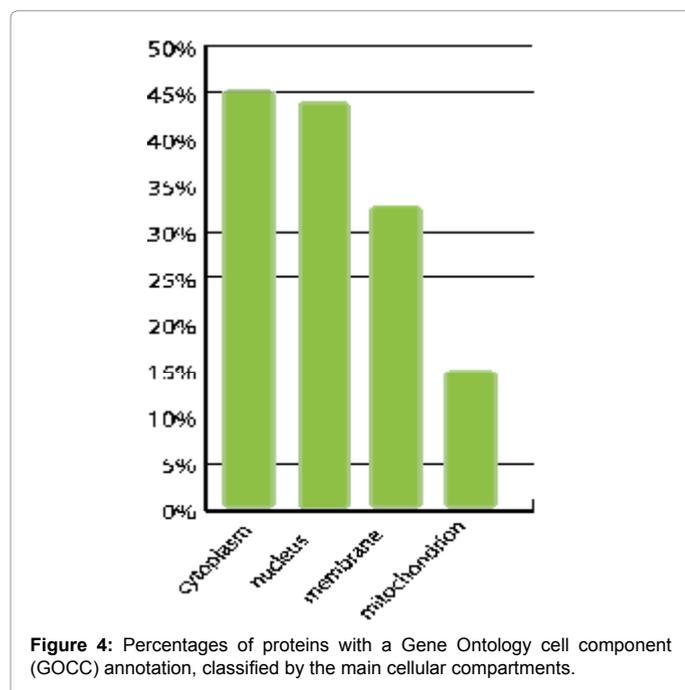


Figure 4: Percentages of proteins with a Gene Ontology cell component (GOCC) annotation, classified by the main cellular compartments.

show an enrichment in intracellular proteins. However, a significant presence of integral membrane proteins could be detected proving that the method, even without detergents, can still be used to detect hydrophobic proteins (Figure 4 and supplementary Table7).

Finally, in solution IEF can be applied with other than urea extraction protocols, as for example the FASP protein extraction technique.

Evaluation of yeast proteome coverage

To evaluate the robustness of the in solution IEF protocol, we fractionated 5 mg of yeast peptides and analyzed the resulting fractions on 240 minutes LC-gradient.

Using such a large sample amount the fractionation was completed after 3 hours as for lower sample loadings. Analyzing the ten fractions we could identify 22,102 different peptides mapping to 3,226 distinct proteins (Table 3 and supplementary Table 8).

In frame of a recent comprehensive yeast proteome analysis, that involved extensive peptide fractionation in 24 fractions and triplicate analyses [26], 3,987 proteins could be identified. This dataset is 23% larger than the one obtained with in solution IEF fractionation, but the latter required a fourth of the analysis time. To evaluate the dynamic range of our analysis, we considered the expression levels of yeast proteins reported by Ghaemmaghami et al. [27] that were later confirmed by mass spectrometry [28].

In our dataset most of the proteins expressed with more than 100,000 copies per cell were present; more interestingly, 85 out of 236 (36%) proteins expressed at less than 250 copies per cell were detected in our analysis. Thus, a dynamic range of 5 orders of magnitude could be achieved with this approach.

The abundance of three proteins reported with <128 copies per cell by Ghaemmaghami et al. [27] (YKR031C, YGL006W, YNR067C) was confirmed by Single Reaction Monitoring (SRM) approach [28]. In our analysis we could detect two of those proteins (YGL006W, YNR067C); interestingly only one unique peptide for YNR067C was identified, while for YGL006W 17 unique peptides were found. Such high coverage of YGL006W despite its low abundance was also observed by Thakur et al. [3]. Furthermore, we could detect 245 out of 1718 proteins that were not detected by Ghaemmaghami et al. [27], showing that urea extraction/in-solution IEF/LC-MS/MS approach can overlap with and complement other techniques.

Functional annotation of identified HEK cells proteins

The functional annotation of large protein dataset is of extreme utility for monitoring the global changes within cellular pathways. Here we used the pathway database of Kyoto Encyclopedia of Genes and Genomes (KEGG) to evaluate the functional information content of our dataset. Notably, only 14 pathways were not represented in our dataset and for 221 pathways at least two proteins were present (supplementary Table 10). In total 42% of the represented pathways have a coverage equal or higher than 50%, including major metabolic processes such as oxidative phosphorylation, tricarboxylic acid (TCA) cycle (supplementary Figures 3 and 4) and purine metabolism, major molecular machineries such as spliceosome (supplementary Figure 5), ribosome and the DNA replication machinery. Also signaling pathways such as mTOR and phosphatidylinositol signaling pathways were largely covered (57% and 46% coverage, respectively).

Furthermore, the GeneOntology analysis of the identified proteins did not suggest a major bias toward proteins from different cellular compartments (Figure 4). For example, the proportion of proteins with 'membrane' annotation is above 30%, close to results obtained with detergent-based extraction procedures [25].

Interestingly, the enrichment analysis of GO terms for biological process shows a significant enrichment in proteins involved in various RNA processing activities, while the enrichment analysis of GO terms for molecular function shows enrichment for RNA and nucleotide binding proteins (supplementary Table 9). This could be explained with

high efficiency of the urea extraction methods for nuclear and cytosolic proteins.

Conclusions

High proteome coverage is still a challenging task and requires laborious sample preparation, expensive instrumentation, special instrumental setup and long working times. For that reason, we developed a simple, fast and robust workflow for sample preparation and peptide fractionation. Using urea extraction and in solution IEF, we could identify more than 5,800 proteins with an average sequence coverage of 25% in just 48 hours including all experimental procedures.

In-solution IEF evidenced to be an excellent solution for peptide fractionation prior to the LC-MS/MS analysis. It performs robustly, especially when large amounts of peptides are loaded. In addition, the focusing is faster than IPG-IEF and is completed in less than 3 hours.

Furthermore, this technique can be coupled with different preprocessing methods (data not shown) and can be used also as enrichment step for phospho-peptides [17] since it can be applied with several mg of starting material.

The established workflow reduces time and working steps and allows deep proteome analyses with a high depth and sequence coverage using a normal nanoscale liquid chromatography coupled to an Orbitrap Velos system.

The chosen protein extraction can influence the detected fraction of the proteome; urea extraction works excellently with hydrophilic proteins. In our HEK293 cell dataset we could observe enrichment in nuclear, cytosolic and RNA binding proteins. However, the presence of a large proportion of membrane proteins, as well as membrane-associated complexes such as proteins of oxidative phosphorylation, suggests the absence of any major bias against hydrophobic and membrane proteins. With the applied strategy, nearly all the annotated enzymatic pathways could be detected and a large proportion extensively covered.

Comparison of our data with published results from SDS extraction [25] and IPG-IEF [12] showed comparable proteome coverage. Moreover, the overlap of the identified peptides and proteins suggests that a deeper coverage of the proteome can be obtained by combining different techniques. Reduction of sample complexity allowed protein detection over several orders of magnitude, similarly to the results obtained by targeted proteomic approaches.

In our group, this fractionation strategy has been already successfully applied to samples from different organisms, yielding high proteome coverage in short time [18,19,29].

We believe that in solution-IEF will be a valid alternative method for peptide fractionation. It is straightforward to think that several improvements in the future will increase the proteome coverage achievable with this method. The use of longer columns and UPLC systems will enhance the chromatographic resolution and increase proteome coverage (data not shown), while improvement of the MS instrumentation, e.g. Orbitrap analyzer with higher resolution [30] or new instrumentation, as the Q Exactive [31], will allow the detection of a higher number of peptides.

Acknowledgement

We thank Julia Diesbach for her excellent technical assistance. This research was funded by the Federal Ministry for Education and Research (BMBF), the HepatomaSys project and the Senate of Berlin, Berlin, Germany.

Author contributions

G.M. and S.K. designed the project. F.B. and C.Z. prepared part of the HEK cells samples, L.A. prepared the yeast sample. G.M. setup the entire protocol, prepared part of the HEK cells samples, performed the LC-MS/MS measurements and analyzed the data. G.M. and S.K. wrote the manuscript. S.K. supervised the project.

References

1. Cox J, Mann M (2007) Is proteomics the new genomics? *Cell* 130: 395-398.
2. Beck M, Claassen M, Aebersold R (2011) *Comprehensive proteomics*. *Curr Opin Biotechnol* 22: 3-8.
3. Thakur SS, Geiger T, Chatterjee B, Bandilla P, Froehlich F, et al. (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without pre-fractionation. *Mol Cell Proteomics* 10.
4. Liu H, Sadygov RG, Yates JR III (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76: 4193-4201.
5. Michalski A, Cox J, Mann M (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* 10: 1785-1793.
6. Blagoev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 22: 1139-1145.
7. de Godoy LM, Olsen JV, de Souza GA, Li G, Mortensen P, et al. (2006) Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol* 7: R50.
8. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1: 2856-2860.
9. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2: 43-50.
10. Washburn MP, Wolters D, Yates JR 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19: 242-247.
11. Horth P, Miller CA, Preckel T, Wenz C (2006) Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol Cell Proteomics* 5: 1968-1974.
12. Hubner NC, Ren S, Mann M (2008) Peptide separation with immobilized pl strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* 8: 4862-4872.
13. Bier M (1998) Recycling isoelectric focusing and isotachopheresis. *Electrophoresis*, 19: 1057-1063.
14. Gazzana G, Borlak J (2007) Improved method for proteome mapping of the liver by 2-DE MALDI-TOF MS. *J Proteome Res* 6: 3143-3151.
15. Thorsell A, Portelius E, Blennow K, Westman-Brinkmalm A (2007) Evaluation of sample fractionation using micro-scale liquid-phase isoelectric focusing on mass spectrometric identification and quantitation of proteins in a SILAC experiment. *Rapid Commun Mass Spectrom* 21: 771-778.
16. Xiao Z, Conrads TP, Lucas DA, Janini GM, Schaefer CF, et al. (2004) Direct ampholyte-free liquid-phase isoelectric peptide focusing: application to the human serum proteome. *Electrophoresis* 25: 128-133.
17. Rogers LD, Fang Y, Foster LJ (2010) An integrated global strategy for cell lysis, fractionation, enrichment and mass spectrometric analysis of phosphorylated peptides. *Mol Biosyst* 6: 822-829.
18. Adamidi C, Wang Y, Gruen D, Mastrobuoni G, You X, et al. (2011) De novo assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics. *Genome research* 21: 1193-1200.
19. Jungkamp AC, Stoeckius M, Mecenas D, Gruen D, Mastrobuoni G, et al. (2011) In vivo and transcriptome-wide identification of RNA binding protein target sites. *Molecular cell* 44: 828-840.
20. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* 72: 248-254

21. Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 75: 663-670.
22. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367-1372.
23. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, et al. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10: 1794-1805.
24. Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1-13.
25. Wisniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6: 359-U360.
26. de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, et al. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 455: 1251-1254.
27. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425: 737-741.
28. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R (2009) Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* 138: 795-806.
29. Mastrobuoni G, Irgang S, Pietzke M, Assmus HE, Wenzel M, et al. (2012) Proteome dynamics and early salt stress response of the photosynthetic organism *Chlamydomonas reinhardtii*. *BMC genomics* 13: 215.
30. Makarov A, Denisov E, Lange O (2009) Performance evaluation of a high-field Orbitrap mass analyzer. *J Am Soc Mass Spectrom* 20: 1391-1396.
31. Michalski A, Damoc E, Hauschild JP, Lange O, Wiegand A, et al. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* 10.