Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis
35 (2004) 895–904
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Fractionation of peptides and identification of proteins from *Saccharomyces cerevisiae* in proteomics with the use of reversed-phase capillary liquid chromatography and p*I*-based approach

Tomasz Bączek*

Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, Gen. J. Hallera 107, 80-416 Gdańsk, Poland

Received 19 January 2003; accepted 26 February 2004

Available online 10 May 2004

Abstract

The aim of the work was to explore the identification of proteins from *Saccharomyces cerevisiae* using combined capillary reversed-phase liquid chromatography (RPLC) and in-solution isoelectric focusing (sIEF) for fractionation of peptides prior to mass spectrometry analysis. That method was proved to be the alternative separation method for complex mixtures of protein tryptic digests in proteomics. Analysis of the identification of peptides was performed with the use of electrospray ionization-ion trap tandem mass spectrometry (ESI-IT-MS/MS). First, the sIEF fractionation was carried out prior to separation and mass spectrometry identification by nano-LC/ESI-MS/MS instrument. The proposed approach based on sIEF and nano-LC/ESI-MS/MS analysis was proved to be an efficient and accurate alternative fractionation method of complex protein digests and can be considered as the useful tool for identification of proteins. Moreover, analytical information from that approach can be considered as the additional source of database matching constraint and can be valuable tool for analytical and bioinformatics studies of peptides fractionation in proteomics. Based on the MS/MS results obtained with ESI-IT-MS/MS instrument, 851 proteins from *S. cerevisiae* were identified. However, after careful analysis of the data reduction in number of proteins to 126 was obtained. Those results are discussed and interpreted in the view of the evaluation method used.

Keywords: Yeast; Peptides; Proteomics; Two-dimensional separations

1. Introduction

Completion of the Human Genome Project makes a better position for understanding of biological functions of organisms [1,2]. However, these studies provide a limited view of cellular processes. Nowadays products of the genes—proteins and a comprehensive analysis and characterization of all expressed proteins called proteomics are the point of the interest. However, the complexity of proteome analysis has been noted as a result of the continual change in concentration of the protein in a cell and their multiple forms due to post-translational modifications. Indeed,

* Tel.: +48-58-3493260; fax: +48-58-3493262.
E-mail address: tbaczek@amg.gda.pl (T. Bączek).

0731-7085/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2004.02.032

while there is 30,000–40,000 of genes in human cells, there may be 10–20 times more proteins to be examined [1]. The issue of a proteome identification is more difficult because of the wide dynamic concentration range and the diversity in protein properties, the lack of an amplification procedure for proteins and their varied biological functions. Today, the most widely used procedure for analyzing complex protein mixtures is two-dimensional gel electrophoresis [3–6]. While this approach has achieved the highest resolving power of any method to date, the method suffers from a number of reasons [7,8]. But it is well known that efficient separation is required prior to mass spectrometry analysis and bioinformatics database searching enabling the correct identification of proteins [9,10]. The high-resolution separation techniques like multidimensional chromatography (including especially ion exchange chromatography (IEC) with reversed-phase liquid chromatography (RPLC) [11–16], but also size-exclusion chromatography (SEC) with RPLC [17], and RPLC with capillary zone electrophoresis (CZE) [18]) coupled to mass spectrometry instruments are currently intensively developed and tested. Liquid chromatography is primarily used to separate mixtures of peptides in proteomics. On the other hand, mass spectrometry is the method of choice for the identification of peptides and proteins. Liquid chromatography coupled to tandem mass spectrometry (LC/MS/MS) is widely used to identify the components of protein complexes and subcellular compartments. This technique is capable to identify hundreds of peptides [12,14,17]. Two-dimensional chromatography (LC/LC) enables already many thousands of peptides to be identified through improved resolution. Link et al. used such kind of an approach to identify the components of yeast and human ribosomes [12], and Washburn et al. performed an analysis of the whole yeast cells [15]. Also alternative strategies, including in-solution isoelectric focusing (sIEF) [19–23] and capillary isoelectric focusing (cIEF) [24,25] as well as chromatofocusing [25–28] and pH-gradient reversed-phase HPLC [29], are studied and developed, and can have the potential value for both protein and peptide identifications in proteome analysis. In the presented work two-dimensional separation system comprising sIEF and RPLC was used.

The main goal of that work was to explore two-dimensional separation of peptides from baker's yeast (*Saccharomyces cerevisiae*) proteins digest using RPLC and sIEF method based on polyacrylamide gel membranes with immobilines technology. Utility of that approach as the alternative separation method for complex mixtures of protein tryptic digest was tested and examined. The potential of that approach for the evaluation of proteomics data and identification of proteins from *S. cerevisiae* was also demonstrated.

2. Experimental

2.1. Materials

2.1.1. Reagents

Bovine serum albumin (BSA) was purchased from Sigma–Aldrich (St. Louis, MO) and used without further purification. Ammonium persulfate, *N*,*N*,*N*′,*N*′-tetramethylethylenediamine (TEMED), γ-methacryloxypropyltrimethoxysilane, trypsin, ammonium bicarbonate, dithiothreitol (DTT), iodoacetamide (IAA), trifluoroacetic acid (TFA), YPD broth and protease inhibitor cocktail for fungal and yeast cells were also obtained from Sigma–Aldrich (St. Louis, MO). Immobiline II p*K* 3.6, 4.6, 6.2, 7.0, 8.5 and 9.3 were products of Amersham Biosciences (Piscataway, NJ). Acrylamide and methylenebisacrylamide, urea and tris(hydroxymethyl)aminomethan were from Pharmacia Biotech (Uppsala, Sweden). IEF anode and cathode buffers, ion exchange membranes and PowerPac 3000 power supply were obtained from Bio-Rad (Hercules, CA).

2.2. Methods

2.2.1. Yeast growing and preparation of the yeast proteins sample

Yeast *S. cerevisiae* strain YPH499 was obtained from American Type Culture Collection (Manassas, VA). Growing in YPD broth (containing yeast extract, peptone and dextrose) at 30 °C was performed to a density of 6×10^8 cells/ml. Next, harvesting of the cells by centrifugation at $3000 \times g$ for 10 min at 4 °C was done and after decanting of the supernatant pellets were allowed to drain and then wet weight of

the pellets was determined. Lysis of the cells was performed using YeastBuster Protein Extraction Reagent (Novagen, Madison, WI) [30] according to the instructions of the manufacturer with the addition of protease inhibitors for fungal and yeast cells. Concentration of the proteins was determined with the use of bicinchoninic acid protein assay kit (Sigma–Aldrich, St. Louis, MO) and equaled about 7 mg/ml.

2.2.2. Protein digestion

BSA and yeast proteins digestion was performed with the addition of RapiGest SF reagent (Waters, Milford, MA) according to the instructions of the manufacturer to enhance in-solution enzymatic digestion of proteins. Briefly, liophilized to dry protein pellets were first suspended in the RapiGest SF dissolved previously in 50 mM ammonium bicarbonate (to give 0.2% of RapiGest solution after the addition of DTT and IAA solutions) and vortex. After addition of DTT to a final concentration of 5 mM, sample was heated at 60 °C for 30 min. After denaturation, the mixture was allowed to cool and IAA was added to a final concentration of 15 mM, and placed in the dark for 30 min at room temperature. Then trypsin was added at an enzyme:protein ratio of 1:50 (w/w) and sample was incubated at 37 °C for 90 min.

2.2.3. In-solution isoelectric focusing (sIEF) fractionation

sIEF device was designed and made according to [21]. sIEF fractionation of yeast proteins digest and BSA digest used additionally to calibrate and optimize the fractionation of yeast peptides were performed into twelve fractions separated by polyacrylamide gel membranes. The pH values of the 11 polyacrylamide gel membranes were 4.00, 4.20, 4.36, 4.50, 5.21, 5.83, 5.99, 6.40, 8.47, 8.75 and 9.74, respectively. The following voltage program was set: 100 V for 30 min, 200 V for 30 min, 500 V for 60 min, 1000 V for 60 min and 2000 V until the completion of the sIEF process evaluated on the basis of the current value lower than 200 μA was achieved. For the sIEF analysis a samples with 0.1 mg/ml of BSA digest and 0.7 mg/ml of soluble fraction of proteins from yeast cells was utilized. After focusing, 12 fractions were simultaneously transferred using a 12-channel digital pipette (Labnet International, Woodbridge, NJ) into 0.2 ml tubes.

2.2.4. Nanocapillary reversed-phase liquid chromatography

LC–MS analysis of sIEF fractions of yeast proteins digest sample was performed with the use of UltiMate Capillary/Nano LC System (Dionex, San Francisco, CA) coupled to the LCQ Deca XP ESI-IT-MS/MS (ThermoFinnigan, San Jose, CA). PepMap C18 column (3 μm, 100 Å, 75 μm i.d. × 150 mm) with PepMap μ-precolumn (300 μm i.d. × 1 mm, packed with 5 μm C18 100) both from Dionex (San Francisco, CA) were used. Gradient elution was performed with a solvent A (water with the addition of 2% acetonitrile and 0.1% formic acid), and solvent B (water with the addition of 85% acetonitrile, 5% isopropanol and 0.1% formic acid). The gradient was 5–35% B in 85 min, followed by 35–90% B in 10 min, and finally, 90% B for another 5 min. Flow rate was 300 nl/min.

2.2.5. ESI-IT-MS/MS

Both MS and MS/MS data were obtained by tandem electrospray ion trap mass spectrometry (ESI-IT-MS/MS) (LCQ Deca XP, ThermoFinnigan, San Jose, CA). On-line ESI was carried out in the positive-ion mode, with the ESI voltage typically set at 0.5–1.4 kV, and the heated inlet capillary maintained at 160 °C. A full MS scan between 400 and 2000 *m/z* was performed by three MS/MS scans between 150 and 2000 *m/z* for the three most intense ions of the MS scan. The relative collision energy was established at 35% with an activation time of 30 ms and dynamic exclusion was done with a repeat count of 2 and a repeat duration of 1 min, with a 3 min exclusion duration window. The activation time was set at 30 ms.

2.2.6. Data analysis

The database of proteins for *S. cerevisiae* was taken from European Bioinformatics Institute [31]. Proteins were theoretically digested with the use of macro program written for Microsoft Excel. Isoelectric points of the peptides were calculated using p*K* values for amino acids [32]. In the case of yeast proteins digest sample TurboSequest software (Thermo Finnigan, San Jose, CA) was used for the searching of the database. The main searching criteria applied in TurboSequest were based on the paper by Peng et al. [14]. Spectra for singly charged peptides with a cross-correlation (X_{corr}) score to a tryptic peptide >2.0, doubly charged tryptic

peptides with X_{corr} at least 1.5 and triple charged tryptic peptides with X_{corr} above 3.3 were accepted. All accepted results had a ΔC_n of more than 0.08. Oxidation of methionine and carboxyamidomethylation of cysteine was only considered as variable modification and fixed modification, respectively. Proteins identified on four or fewer peptides were manually confirmed according to the rules recommended by Link et al. [12]. When five or more peptides were identified from the protein, manual validation was performed for at least one peptide.

3. Results and discussion

In the case of such a complex sample like yeast proteins digest sIEF fractionation combined with RPLC separations as two-dimensional system was utilized. The focused peptides in all chambers of the sIEF device were separated and analyzed by nano-LC/ESI-MS/MS. The pH values for sIEF polyacrylamide gel membranes were designed specially with respect to such complex peptide mixture fractionation. pH values of each polyacrylamide gel membrane were chosen to fractionate peptides in equal number among twelve chambers in sIEF device. This simulation was performed with the use of 60,000 peptides, which were derived by the theoretical tryptic digest of the yeast proteins (Fig. 1a). For comparison, the distribution of peptides without the optimization is presented in Fig. 1b. Based on the MS/MS results obtained with nanoLC-ESI-IT-MS/MS instrument, 851 proteins were identified, with 163 protein identifications based on single peptide (Table 1). After manual interpretation of MS/MS spectra according to the rules recommended by Link et al. [12], 542 proteins were finally identified from the single gradient runs performed for all twelve sIEF fractions. Among

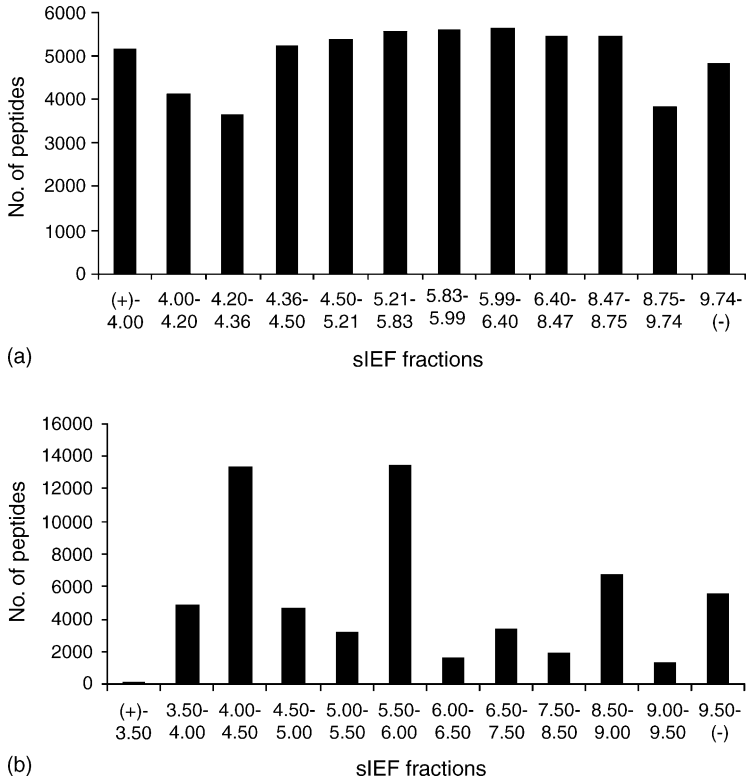

Fig. 1. The distribution of peptides performed with the use of 60,000 peptides, which were derived by the theoretical tryptic digest of the yeast proteins: (a) with optimization of the pH values for polyacrylamide gel membranes; (b) without the optimization of the pH values for polyacrylamide gel membranes.

Table 1
Number of proteins identified with the use of differently considered constraints

No.	Constraints used	No. of all identified proteins	No. of proteins identified on the basis of more than single peptide	Proteins identified on the basis of more than single peptide (%)
1	Tryptic peptide with: X_{corr} > 2.0 (singly charged), X_{corr} > 1.5 (doubly charged), X_{corr} > 3.3 (triple charged), ΔC_n > 0.08	851	163	19.2
2	As (1) and after manual interpretation of MS/MS spectra	542	93	17.2
3	As (2) and after using of sIEF-based p*I* constraint	187	50	26.7
4	As (2) but (1) using without any enzyme fixed	188	67	35.6
5	As (3) but (1) using without any enzyme fixed	126	50	39.7

that number of proteins, 93 of them were identified on the basis of more than one peptide.

Additionally, the distribution of the peptides according to the frequency of their occurrence in one or more sIEF fractions was checked with the use of proteins identifications based on more than one peptide. More than 60% of the identified peptides were found in only ones IEF chamber (Table 2).

As noted earlier, among 542 proteins identified only 17.2% of them were identified on the basis of more than one peptide. One of the reasons of such a situation can be just the complexity of the sample and not sufficiently performed separation of peptides. In fact, after sIEF fractionation, 12 fractions were collected, what can be not satisfactory prior to reversed-phase separation in the case of such the complex sample. Hence, a probability to match only single peptide for any protein increases. On the other hand, despite of the manual interpretation of MS/MS spectra, a number of proteins identified with single peptide can be treated as false positives. Therefore, p*I* values for peptides were next considered as the additional identification constraint. Taking into account potential inaccuracy among theoretically designed pH ranges for each chamber and experimentally achieved ones, BSA fractionation was carried out. Fractionation of BSA digest sample was performed using polyacrylamide gel membranes with pH values designed for yeast proteins digest sample, i.e., 4.00, 4.20, 4.36, 4.50, 5.21, 5.83, 5.99, 6.40, 8.47, 8.75 and 9.74, respectively. The evaluation of peptides distribution in the certain chambers regards the number of individual peptides identified was done. The pH ranges considered for that research were estimated. It was done on the basis of the average standard deviation value (S.D. = 0.12) obtained for relationship between pI_{calc} and pI_{exp} values for peptides identified for BSA with the use of MS/MS spectra and found in one chamber. Now, individual pH of the membrane for each chamber was considered as ±0.12. Using p*I* as the additional constraint it was possible to identify now 187 proteins (Table 1) with 50 of them with the use of more than single peptide (26.7% of all proteins identified).

Another method to confirm the goodness of matching proteins using TurboSequest software comprise the use of the comparison of database searching results obtained with an unrestricted enzyme mode with the specifically restricted enzyme mode [33]. Set of 187 identified proteins was evaluated with that method providing reduction in number of proteins to 126. However, it should be noted that no one protein

Table 2
The distribution of the peptides according to the frequency of their occurrence in one or more fractions performed on the basis of peptides from proteins identified on more than one peptide

No. of fractions with the same peptide	No. of peptides	Peptides (%)
1	228	61.8
2	59	16.0
3	26	7.0
4	24	6.5
5	18	4.9
6	7	1.9
7	2	0.5
8	2	0.5
9	3	0.8
10	0	0.0
11	0	0.0
12	0	0.0
Sum	369	100.0

Table 3
All identified proteins using sIEF combined with nanoRPLC-ESI-IT-MS/MS

SWISS-PROT accession number	Protein	Codon bias	Based on more than a single peptide (M) or based on a single peptide (S)
P00359	Glyceraldehyde 3-phosphate dehydrogenase 3	0.926	M
P06169	Pyruvate decarboxylase isozyme 1	0.914	M
P00358	Glyceraldehyde 3-phosphate dehydrogenase 2	0.900	M
P00925	Enolase 2	0.895	M
P00924	Enolase 1	0.888	M
P14540	Fructose-bisphosphate aldolase	0.888	M
P00549	Pyruvate kinase 1	0.882	M
P02994	Elongation factor 1-alpha	0.877	M
P00330	Alcohol dehydrogenase I	0.846	M
P00560	Phosphoglycerate kinase	0.843	M
P00360	Glyceraldehyde 3-phosphate dehydrogenase 1	0.842	M
P11484	Heat shock protein SSB1	0.833	M
P00950	Phosphoglycerate mutase 1	0.824	M
P06168	Ketol-acid reductoisomerase, mitochondrial precursor	0.823	M
P00942	Triosephosphate isomerase	0.820	M
P10592	Heat shock protein SSA2	0.815	M
P40150	Heat shock protein SSB2	0.808	M
P32324	Elongation factor 2	0.795	M
P10081	Eukaryotic initiation factor 4A	0.772	M
P10591	Heat shock protein SSA1	0.760	M
P02579	Actin	0.734	M
P14832	Peptidyl-prolyl *cis–trans* isomerase	0.714	M
P22943	12 kDa heat shock protein	0.703	M
P38720	6-Phosphogluconate dehydrogenase, decarboxylating 1	0.676	M
P04807	Hexokinase B	0.674	M
P00331	Alcohol dehydrogenase II	0.647	M
P00817	Inorganic pyrophosphatase	0.647	M
P20081	FK506-binding protein 1	0.641	M
P54115	Magnesium-activated aldehyde dehydrogenase, cytosolic	0.620	M
P38879	EGD2 protein	0.617	M
P32449	Phospho-2-dehydro-3-deoxyheptonate aldolase, tyrosine-inhibited	0.608	M
P12398	Heat shock protein SSC1, mitochondrial precursor	0.602	M
P04806	Hexokinase A	0.534	M
P19414	Aconitate hydratase, mitochondrial precursor	0.513	M
P00445	Superoxide dismutase [Cu–Zn]	0.494	M
P08524	Farnesyl pyrophosphate synthetase	0.452	M
P40495	Hypothetical 40.1 kDa protein in SGA1-KTR7 intergenic region	0.427	M
P15992	Heat shock protein 26	0.398	M
P32582	Serine sulfhydrase	0.395	M
O94073	Cystathionine beta-synthase	0.395	M
P10963	Phosphoenolpyruvate carboxykinase [ATP]	0.379	M
P07264	3-Isopropylmalate dehydratase	0.364	M
P07244	Bifunctional purine biosynthetic protein ADE5,7	0.328	M
P26637	Leucyl-tRNA synthetase, cytoplasmic	0.299	M
P39676	Flavohemoprotein	0.295	M
P17709	Glucokinase	0.208	M
P22202	Heat shock protein SSA4	0.176	M
P09435	Heat shock protein SSA3	0.155	M
P16547	Mitochondrial outer membrane 45 kDa protein	0.146	M

Table 3 (*Continued*)

SWISS-PROT accession number	Protein	Codon bias	Based on more than a single peptide (M) or based on a single peptide (S)
P47068	Myosin tail region-interacting protein MTI1	0.071	M
P16521	Elongation factor 3A	0.806	S
P39015	Suppressor protein MPT4	0.804	S
P23301	Eukaryotic translation initiation factor 5A-2	0.783	S
P34760	Thiol-specific antioxidant protein	0.780	S
P16467	Pyruvate decarboxylase isozyme 2	0.765	S
P12709	Glucose-6-phosphate isomerase	0.729	S
P41940	Mannose-1-phosphate guanyltransferase	0.684	S
P04451	60S ribosomal protein L23	0.679	S
P37291	Serine hydroxymethyltransferase, cytosolic	0.672	S
Q02642	BTF3 homolog EGD1	0.662	S
P39954	Adenosylhomocysteinase	0.662	S
P15019	Transaldolase	0.631	S
P41277	(DL)-Glycerol-3-phosphatase 1	0.613	S
P10659	*S*-Adenosylmethionine synthetase 1	0.590	S
P19358	*S*-Adenosylmethionine synthetase 2	0.588	S
P07170	Adenylate kinase cytosolic	0.579	S
P32589	Heat shock protein homolog SSE1	0.572	S
P31373	Cystathionine gamma-lyase	0.548	S
P04147	Polyadenylate-binding protein, cytoplasmic and nuclear	0.544	S
P23641	Mitochondrial phosphate carrier protein	0.532	S
Q03558	NADPH dehydrogenase 2	0.511	S
P22803	Thioredoxin I	0.501	S
Q12207	Non-classical export protein 2	0.488	S
P22768	Argininosuccinate synthase	0.487	S
P16474	78 kDa glucose-regulated protein homolog precursor	0.481	S
P16140	Vacuolar ATP synthase subunit B	0.474	S
P15180	Lysyl-tRNA synthetase, cytoplasmic	0.468	S
P04802	Aspartyl-tRNA synthetase, cytoplasmic	0.464	S
P41338	Acetyl-CoA acetyltransferase	0.464	S
P38910	10 kDa heat shock protein, mitochondrial	0.452	S
P38891	Branched-chain amino acid aminotransferase, mitochondrial precursor	0.421	S
P01095	Protease B inhibitors 2 and 1	0.421	S
P21954	Isocitrate dehydrogenase [NADP], mitochondrial precursor	0.408	S
P29509	Thioredoxin reductase 1	0.403	S
P38999	Saccharopine dehydrogenase [NADP+, L-glutamate forming]	0.403	S
P00931	Tryptophan synthase	0.401	S
P38009	Bifunctional purine biosynthesis protein ADE17	0.377	S
Q05911	Adenylosuccinate lyase	0.374	S
Q07478	Probable ATP-dependent RNA helicase SUB2	0.370	S
P46367	Potassium-activated aldehyde dehydrogenase, mitochondrial precursor	0.368	S
P17423	Homoserine kinase	0.364	S
P36139	Hypothetical 31.2 kDa protein in GAP1-NAP1 intergenic region	0.364	S
P02557	Tubulin beta chain	0.359	S
P06208	2-Isopropylmalate synthase	0.347	S
P25087	Sterol 24-C-methyltransferase	0.340	S
P40106	(DL)-Glycerol-3-phosphatase 2	0.340	S
Q00711	Succinate dehydrogenase [ubiquinone] flavoprotein subunit	0.314	S
P00815	Histidine biosynthesis trifunctional protein	0.314	S
P32316	Acetyl-CoA hydrolase	0.311	S

Table 3 (*Continued*)

SWISS-PROT accession number	Protein	Codon bias	Based on more than a single peptide (M) or based on a single peptide (S)
P14020	Dolichol-phosphate mannosyltransferase	0.308	S
P38113	Alcohol dehydrogenase V	0.304	S
P04037	Cytochrome *c* oxidase polypeptide IV, mitochondrial precursor	0.299	S
P35732	Hypothetical 84.0 kDa protein in NUP120-CSE4 intergenic region	0.298	S
P27616	Phosphoribosylamidoimidazole-succinocarboxamide synthase	0.292	S
P53598	Succinyl-CoA ligase [GDP-forming] alpha-chain	0.290	S
P36136	Hypothetical 31.0 kDa protein in GAP1-NAP1 intergenic region	0.288	S
P33330	Phosphoserine aminotransferase	0.279	S
P25373	Glutaredoxin 1	0.272	S
Q04869	Hypothetical 38.2 kDa protein in PRE5-FET4 intergenic region	0.262	S
P30952	Malate synthase 1, glyoxysomal	0.259	S
P03965	Carbamoyl-phosphate synthase, arginine-specific, large chain	0.256	S
P53912	Hypothetical 41.2 kDa protein in FPR1-TOM22 intergenic region	0.245	S
P17695	Glutaredoxin	0.241	S
P49367	Homoaconitase, mitochondrial precursor	0.232	S
P39533	Putative aconitase in PRP21-UBP12 intergenic region	0.230	S
Q99383	Nuclear polyadenylated RNA-binding protein NAB4	0.225	S
P00175	Cytochrome B2, mitochondrial precursor	0.224	S
P28272	Dihydroorotate dehydrogenase	0.209	S
P26263	Pyruvate decarboxylase isozyme 3	0.183	S
P47052	Probable succinate dehydrogenase [ubiquinone] flavoprotein subunit 2	0.176	S
Q07844	Hypothetical 93.1 kDa protein YLL034C	0.138	S
P32457	Cell division control protein 3	0.107	S
Q08645	Folylpolyglutamate synthase	0.090	S
P15202	Catalase A	0.065	S
P08004	Chitin synthase 1	0.060	S
P40893	Hypothetical 22.0 kDa protein in HXT11-HXT8 intergenic region	0.019	S

identified on the basis of more than single peptide was discarded. Hence, proteins with more than one peptide comprised now almost 40% (Table 1). Moreover, using several constraints including p*I* value, it is more probable that hits based on only single peptide are significant. The same searching procedure but without using sIEF-based p*I* constraint was able to provide 188 identified proteins with 67 proteins based on single peptide (35.6% of all proteins). Interestingly and more importantly, when searching additionally that set of proteins (188 identified proteins with 67 proteins based on single peptide) with the use of p*I*-based approach, the same number of proteins was finally identified as in the previously used sequence of constraints (126 identified proteins with 50 proteins based on more than single peptide, Table 1).

All identified proteins with the use of all constraints used and proposed were also checked regards to their abundance. The codon bias information was taken from the Saccharomyces Genome Database [34]. Considering the value of codon bias for all proteins identified, it was found that they represent rather the whole range of the abundance with the average value of 0.490 (Table 3). However, considering separately proteins identified on the basis of single peptide and more than one peptide it is seen in Table 4 that in that first case more proteins are in the lower range of the codon bias value (usually between 0.2 and 0.5) and in that latter case they are in the higher range of the codon bias value (with the highest frequency between 0.8 and 0.9). Five low abundance proteins with codon bias values below 0.1 were also identified. For example, myosin tail region-interacting protein MTI1 (P47068)

Table 4
The distribution of the proteins according to the frequency of their occurrence regards the values of the codon bias

The range of the codon bias value	Proteins identified on the basis of more than a single peptide		Proteins identified on the basis of a single peptide	
	No. of proteins	Proteins (%)	No. of proteins	Proteins (%)
0.0–0.1	1	2.0	4	5.3
0.1–0.2	3	6.0	4	5.3
0.2–0.3	3	6.0	17	22.4
0.3–0.4	6	12.0	15	19.7
0.4–0.5	3	6.0	14	18.4
0.5–0.6	2	4.0	9	11.8
0.6–0.7	9	18.0	7	9.2
0.7–0.8	6	12.0	4	5.3
0.8–0.9	15	30.0	2	2.6
0.9–1.0	2	4.0	0	0.0
Sum	50	100.0	76	100.0

has the codon bias value of 0.071, and folylpolyglutamate synthase (Q08645), the codon bias value 0.090 were detected and identified on the basis of two peptides and one peptide, respectively. All identified proteins are listed in Table 3 using codon bias values for sorting, along with their names and SWISS-PROT accession numbers.

4. Conclusions

The proposed approach based on RPLC and sIEF method was proved to be useful as an alternative fractionation method for *S. cerevisiae* protein digest and can be considered as useful two-dimensional proteomics separation approach. sIEF fractionation combined with reversed-phase nanocapillary liquid chromatography separations enabled the evaluation of yeast proteome. Besides the separation aspects, the approach proposed can be considered as the additional source of database matching constraint used in the evaluation process of proteomics data. Using p*I* values calculated for identified peptides it was possible to check the validity of the database searching considering the occurrence of peptides in the appropriate sIEF fraction.

Acknowledgements

The author is very grateful to the Foundation for Polish Science for the support during the course of this research.

References

[1] International Human Genome Sequencing Consortium, Nature 409 (2001) 860–921.
[2] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, et al., Science 291 (2001) 1304–1351.
[3] R. Joubert, J.-M. Strub, S. Zugmeyer, D. Kobi, N. Carte, A. van Dorsselaer, H. Boucherie, L. Jaquet-Gutfreund, Electrophoresis 22 (2001) 2969–2982.
[4] M. Perrot, F. Sagliocco, T. Mini, C. Monribot, U. Schneider, A. Shevchenko, M. Mann, P. Jeno, H. Boucherie, Electrophoresis 20 (1999) 2280–2298.
[5] M. Poutanen, L. Salusjarvi, L. Ruohonen, M. Penttila, N. Kalkkinen, Rapid Commun. Mass Spectrom. 15 (2001) 1685–1692.
[6] L. Salusjarvi, M. Poutanen, J.-P. Pitkanen, H. Koivistoinen, A. Aristidou, N. Kalkkinen, L. Ruohonen, M. Penttila, Yeast 20 (2003) 295–314.
[7] D.C. Liebler, Introduction to Proteomics, Humana Press, Totowa, NJ, 2002.
[8] A. Pandey, M. Mann, Nature 405 (2000) 837–846.
[9] T. Wehr, LCGC N. Am. 20 (2002) 954–962.
[10] T. Bączek, A. Buciñski, A.R. Ivanov, R. Kaliszan, Anal. Chem. 76 (2004) 1726–1732.
[11] M.T. Davis, J. Beierle, E.T. Bures, M.D. McGinley, J. Mort, J.H. Robinson, C.S. Spahr, W. Yu, R. Luthy, S.D. Patterson, J. Chromatogr. B 752 (2001) 281–291.
[12] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates Jr., Nat. Biotechnol. 17 (1999) 676–682.
[13] G.J. Opiteck, K.C. Lewis, J.W. Jorgenson, R.J. Anderegg, Anal. Chem. 69 (1997) 1518–1524.
[14] J. Peng, J.E. Elias, C.C. Thoreen, L.J. Licklider, S.P. Gygi, J. Proteome Res. 2 (2003) 43–50.
[15] M.P. Washburn, D. Wolters, J.R. Yates Jr., Nat. Biotechnol. 19 (2001) 242–247.

[16] K. Wagner, T. Miliotis, G. Marko-Varga, R. Bischoff, K.K. Unger, Anal. Chem. 74 (2002) 809–820.
[17] G.J. Opiteck, J.W. Jorgenson, R.J. Anderegg, Anal. Chem. 69 (1997) 2283–2291.
[18] K.C. Lewis, G.J. Opiteck, J.W. Jorgenson, D.M. Sheeley, J. Am. Mass Spectrom. 8 (1997) 495–500.
[19] B. Herbert, P.G. Righetti, Electrophoresis 21 (2000) 3639–3648.
[20] P.E. Michel, F. Reymond, I.L. Arnaud, J. Josserand, H.H. Girault, J.S. Rossier, Electrophoresis 24 (2003) 2–11.
[21] A. Tan, A. Pashkova, L. Zang, F. Foret, B.L. Karger, Electrophoresis 23 (2002) 3599–3607.
[22] X. Zuo, D.W. Speicher, Proteomics 2 (2002) 58–68.
[23] T. Bączek, J. Pharm. Biomed. Anal. 34 (2004) 851–860.
[24] Y. Shen, S.J. Berger, G.A. Anderson, R.D. Smith, Anal. Chem. 72 (2000) 2154–2159.
[25] Y. Shen, F. Xiang, T.D. Veenstra, E.N. Fung, R.D. Smith, Anal. Chem. 71 (1999) 5348–5353.
[26] D.M. Lubman, M.T. Kachman, H. Wang, S. Gong, F. Yan, R.L. Hamler, K.A. O'Neil, K. Zhu, N.S. Buchanan, T.J. Barder, J. Chromatogr. B 782 (2002) 183–196.
[27] X. Kang, D.D. Frey, Anal. Chem. 74 (2002) 1038–1045.
[28] X. Kang, D.D. Frey, J. Chromatogr. A 991 (2003) 117–128.
[29] R. Kaliszan, P. Wiczling, M.J. Markuszewski, Anal. Chem. 76 (2004) 749–760.
[30] D. Drott, S. Bahairi, A. Grabski, inNovations 15 (2002) 14–16.
[31] http://www.ebi.ac.uk/.
[32] B. Bjellqvist, B. Basse, E. Ilsen, J.E. Celis, Electrophoresis 15 (1994) 529–539.
[33] http://www.enovatia.com/sequest/.
[34] http://genome-www4.stanford.edu/Saccharomyces/SGD/.