Improvement of mitochondria extract from *Saccharomyces cerevisiae* characterization in shotgun proteomics using sheathless capillary electrophoresis coupled to tandem mass spectrometry

Marianne Ibrahim[1], Rabah Gahoual[1], Ludovic Enkler[2], Hubert Becker[2], Johana Chicher[3], Philippe Hammann[3], Yannis-Nicolas François[1], Lauriane Kuhn[3], Emmanuelle Leize-Wagner[1]

Author information:

[1] Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), UDS-CNRS UMR 7140, Université de Strasbourg, 67008 Strasbourg, France.

[2] Unité Mixte de Recherche 7156 Génétique Moléculaire Génomique Microbiologie, Centre National de la Recherche Scientifique, Université de Strasbourg, 67084 Strasbourg, France.

[3] Plateforme Protéomique Strasbourg-Esplanade, Institut de Biologie Moléculaire et Cellulaire, FRC 1589, Centre National de la Recherche Scientifique, Université de Strasbourg, 67084 Strasbourg, France.

Corresponding author: Yannis-N. François, yfrancois@unistra.fr

ABSTRACT

In this work, we describe the characterization of a quantity-limited sample (100ng) of yeast mitochondria by shotgun bottom-up proteomics. Sample characterization was carried out by sheathless capillary electrophoresis, equipped with a high sensitivity porous tip and coupled to tandem mass spectrometry (CESI-MS/MS) and concomitantly with a state-of-art nano flow liquid chromatography coupled to a similar mass spectrometry (MS) system (nanoLC-MS/MS). With single injections, both nanoLC-MS/MS and CESI-MS/MS 60 min-long separation experiments allowed to identify 271 proteins (976 unique peptides) and 300 proteins (1765 unique peptides) respectively, demonstrating a significant specificity and complementarity in identification depending on the physicochemical separation employed. Such complementary, maximizing the number of analytes detected, presents a powerful tool to deepen a biological sample's proteomic characterization. A comprehensive study of the specificity provided by each separating technique was also performed using the different properties of the identified peptides: molecular weight (MW), mass-to-charge ratio (m/z), isoelectric point (pI), sequence coverage or MS/MS spectral quality enabled to determine the contribution of each separation. For example, CESI-MS/MS enables to identify larger peptides and eases the detection of those having extreme pI without impairing spectral quality. The addition of peptides, and therefore proteins identified by both techniques allowed to increase significantly the sequence coverages and then the confidence of characterization. In this study, we also demonstrated that the 2 yeast enolase isoenzymes were both characterized in the CESI-MS/MS dataset. The observation of discriminant proteotypic peptides is facilitated when a high number of precursors with high-quality MS/MS spectra are generated.

52
53
54
55
56
57
58
59   **Introduction**

60

61   In recent years, mass spectrometry-based proteomics was recognized as one of the best

62   tools for identifying proteins with new functions, mapping their interactions in a cellular

63   context and discovering new biomarkers for medical research [1]. In order to maximize the

64   number of analytes successfully ionized, detected and identified in mass spectrometry (MS), a

65   physicochemical separation prior to MS analysis is often applied. Currently, liquid

66   chromatography (LC) is the most widely used separation techniques that can be coupled on-

67   line with a mass spectrometer [2]. In a classical nanoLC coupled to tandem mass spectrometry

68   (nanoLC-MS/MS) experiment, which is the most usual method for proteomic application,

69   analytes are eluted from a reverse-phase column by increasing the organic content of the

70   mobile phase to separate peptides by hydrophobicity. Capillary Zone Electrophoresis (CZE) is

71   another separating technique that has been successfully coupled to electrospray ionization [3]:

72   with the use of an electrical field, ions are separated based on their electrophoretic mobility

73   which is dependent upon the charge of the molecule and the analyte's hydrodynamic radius [4].

74   Capillary electrophoresis (CE) is gaining more and more interest mostly because of recent

75   technical improvements regarding capillary zone electrophoresis–mass spectrometry (CZE-

76   MS) coupling that allow the use of lower flow rates resulting in improved sensitivity [5,6] and

77   also because ion suppression phenomenon is largely reduced compared to a nanoLC-MS

78   analysis [7-9]. Recently, CE–MS has emerged as a highly efficient separation technique [10] and is

79   considered as a powerful technique for the analysis of peptides and proteins [4,11-22].

80   Taking into considerations these observations, an ideal workflow would be the

81   combination of orthogonal separative techniques based on different physiochemical

82   principles. Indeed, as a significant fraction of proteins can escape detection in individual

83   separation approaches [23], crossing over different techniques and their respective benefits

84   would minimize that problem. Consecutive CZE-MS/MS and nanoLC-MS/MS analysis

85   campaigns can be easily imagined on the same samples as further discussed in this study with

86   the example of a quantity-limited sample from a yeast mitochondrial extract.

87   In this study, we used CZE-MS/MS as well as a classical nanoLC configuration to

88   explore the *S. cerevisiae* mitochondrial proteome. CZE-MS/MS experiments were performed

89   using a recently introduced CE-MS interface referred as CESI-MS. New instrumental

90   approaches are often developed and validated on model systems such as the baker's yeast

91   *Saccharomyces cerevisiae,* mainly because many essential processes are conserved between

92   yeast and other organisms of interest like humans [24]. Numerous benefits derive from the fact

93   that yeast is one of the simplest eukaryotes [25]. Because of its importance in molecular biology,

94   it was the first eukaryotic organism for which the genome sequence was completed [26].

95   Mitochondrion is a complex intracellular organelle, crucial for numerous cellular functions

96   like normal cell metabolism [27], cellular energetic [28], maintenance of ion homeostasis [29] and

97   programmed cell death [30,31]. To achieve the best results, a highly purified mitochondrial

98   preparation is mandatory [32,33]. To understand the role of this complex organelle especially in

99   disease, it is important to extensively characterize its protein composition: identifying the

100 whole set of resident proteins within a complex organelle remains a major challenge in cell

101 biology even if innovative technologies are emerging year after year. Previous proteomic

102 analyses on purified mitochondria resulted in a repertoire of 1000 to 1500 different proteins

103 for that organelle, in either yeast or human samples [34,35].

104   In the current work, the aim is to demonstrate the improvement in protein identification

105   by shotgun bottom-up proteomics of mitochondria extract from *Saccharomyces cerevisiae* by

106   using two different physicochemical separations prior to the MS analysis. The study is carried

107   out on a limited quantity (100ng) of sample, roughly representing 1000 proteins.

108   Identifications obtained by CESI-MS/MS were compared to those achieved in nanoLC-

109   MS/MS. Specificity of each separating technique was determined observing different

110   physicochemical properties of the identified peptides: molecular weight (MW), mass-to-

111   charge ratio (*m/z*), isoelectric point (pI), sequence coverage as well as MS/MS spectral

112   quality. Moreover, a cellular localization analysis of all identified proteins was also

113   performed.   Biological significance was finally investigated by comparing the protein

114   identifications from this study to previously published studies. In addition, emphasis was

115   placed onto a challenging area of proteomics research which is isoforms characterization.

116   Taking together, the results presented in this study allowed us to assess the gain of

117   information which can be achieved by two complementary and orthogonal techniques, in

118   terms of peptide metrics, sequence coverage, spectrum quality and isoforms characterization.

119

120   **Experimental**

121   **Materials and reagents.**

122   Chemicals used were of analytical grade or high purity grade and purchased from Sigma-

123   Aldrich (Saint Louis, MO, USA).Water used to prepare buffers, mobile phases and sample

124   solutions was obtained using an ELGA purelab UHQ PS water purification system (Bucks,

125   UK). Dithiothreitol (DTT) and iodoacetamide (IAA) were purchased from Sigma (St. Louis,

126   MO, USA). Trypsin sequencing grade was obtained from Promega (Madison, WI USA). All

127   other reagents and plastic ware were obtained from commercial sources.

128

Isolation of Mitochondria from *S. cerevisiae*

Yeast cells were grown in Yeast Peptone Dextrose (YPD) at 30°C up to $OD_{600nm}=1,2$. Mitochondria were isolated as detailed in the Supplemental Information section. Briefly, spheroplasts have been first generated by the enzymatic digestion of the cells wall by Zymolase 20T (Euromedex) and further broken by homogenization in a glass-Teflon potter. After the removal of cell debris and nuclei by centrifugation ($1,500\times g$, 5 min at 4°C), mitochondria were recovered from the supernatant by additional centrifugations at $12,000\times g$ for 15 min at 4°C, to give a final concentration of 5 mg/ ml. Mitochondrial fraction was treated by 10 strokes in a glass-Teflon potter and loaded onto a four-step sucrose gradient, from which purified mitochondria were recovered from the 60 %/ 32 % interface.

In solution protein digestion

10μg of protein were diluted in 75μL ammonium bicarbonate buffer (50mM) and incubated with 5μL DTT 0.1M for protein denaturation (10min, 95°C). After cooling down, 10μL IAA 0.1M were added and the samples were incubated for 20min at room temperature in the dark to allow alkylation of reduced cysteine residues. For protein digestion, 500ng (1:20) trypsin (Promega, V5111) was added and sample was incubated overnight at room temperature. For nano-LC/MSMS analysis, 3μg of the clear supernatant were transferred into glass inserts and evaporated to dryness using a miVac DNA concentrator (Genevac, NY, USA). The sample waseither reconstituted in 15μL formic acid (0.1% v/v) for nanoLC-MS/MS analyses or in ammonium acetate 50 mM (pH 4.0) for CESI-MS/MS analyses.

NanoLiquid Chromatography

Samples were transferred into glass inserts, compatible with the LC autosampler system (ultra nanoLC-2D plus, Eksigent, UK). Five microliters of each sample were loaded on a

155 nanoFlexcHiP module consisting of a Trap-and-elute jumper (fluidics configuration design), a

156 ChromXP-C18 trap cHiP column (0.5cm x 200µm i.d., 3µm, 120Å, Eksigent) and a

157 ChromXP-C18 analytical cHiP column (15cm x 75µm i.d., 3µm, 120Å, Eksigent). The

158 separation method consisted in a 60 min run at a flow rate of 300 nL/min using a gradient of

159 two solvents: A (99.9% water: 0.1% formic acid) and B (99.9% acetonitrile: 0.1% formic

160 acid). After a 10 min step to trap the peptides on the pre-concentration column at 20µL/min,

161 they were eluted from the analytical column as follows: 0-35 min, 5-40% B; 35-36 min, 40-

162 80% B; 36-40 min, 80% B; 40-41 min, 80-5% B; 41-60 min, 5% B.

163

164 **Capillary Electrophoresis**
165
166 The CE experiments were performed with a CESI8000 capillary electrophoresis system from

167 Sciex separations (Brea, CA) equipped with a temperature controlled autosampler and a

168 power supply able to deliver up to 30 kV. Bare fused-silica capillaries (total length 91 cm;

169 30µm i.d.) with characteristic porous tip on its final 3 cm supplied by Sciex separation

170 allowed electric contact between both electrodes of the CE system. New capillaries were

171 flushed at 75 psi (5.17 bar) for 10min with methanol, then 10min with 0.1M sodium

172 hydroxide, followed 10min with 0.1M hydrochloric acid and water for 20min. Finally, the

173 capillary was flushed 10min at 75 psi with BGE which was acetic acid 10%. Hydrodynamic

174 injection (410 mbar for 1min) corresponding to a total volume of 55 nL of sample injected

175 was used. Separations were performed using an electric field of +219.8 V/cm^{-1}.

176

177 **Mass spectrometry**

178   The LC and CESI systems were coupled to a 5600Triple TOF^{TM} system (AB Sciex, CA,

179 USA) with nanospray source operating in positive ESI mode. Interchangeability of the LC

180 and CESI systems is easy and takes only 20 min. The mass spectrometer was operated in the

181   IDA mode (Information-Dependent Acquisition) and was externally calibrated each 6 hours

182   with 25fmol of an *E.coli* Beta-galactosidase digest. This standard also allowed to regularly

183   check the performances and reproducibility of the mass spectrometer. For the nanoLC-

184   MS/MS, the following ESI source parameters were used: ion spray voltage, 2.3 kV; ion

185   source heater, 150 °C; curtain gas, 22 psi and ion source gas 1,5 psi. While for the CESI-

186   MS/MS and due to the nanospray properties, the ESI source parameters were: ion spray

187   voltage, 1.75 kV (1.0 to 4.0 mm between the MS inlet and capillary tip); ion source heater, 75

188   °C; curtain gas, 5 psi (flow rate < 100 nL/min) and ion source gas 1, 0 psi. The MS survey

189   scan was acquired over a 400-1250 *m/z* mass range with a 250ms accumulation time; whereas

190   the product ion experiments mass range was extended to 100-2000 *m/z* with a 60ms

191   accumulation time. The duty cycle time was therefore maintained below 2s with up to 30

192   precursor ions selected on the MS survey scan to undergo CID fragmentation (Collision

193   Induced Dissociation), while respecting the following criteria: intensity greater than 150

194   counts per second, charge state between 2+ and 5+, not present in the dynamic exclusion list,

195   exclusion after a period of 15s and 2 consecutive MS/MS spectra, and use of a rolling

196   collision energy. Data were acquired with Analyst 1.6 software (AB Sciex) and the resulting

197   .wiff files were primarily evaluated with PeakView 1.2 software (AB Sciex).

198

199   **Data Analysis**

200   For protein identification, the MS/MS raw data files were first converted to mascot generic

201   format (mgf) using the executable MS-DataConverter.exe (AB Sciex). MS data processing

202   was then performed with Proteinscape software (version 3.1, Bruker Daltonics). We used the

203   Mascot algorithm (v2.2, Matrix Science, UK) for database searching, configured with the

204   following parameters: (a) database: Swiss-Prot with *Baker's Yeast* taxonomy (version 2013-

205   01-09, 18860 sequences), (b) enzyme: Trypsin/P, (c) maximum missed cleavages: 3, (d)

variable modifications: acetylation (Protein N-term), oxidation (Met), carbamidomethylation (Cys), (e) peptide mass tolerance: 30 ppm, (f) MS/MS tolerance: 0.5 Da, (g) Instrument: ESI-Quad-TOF. To evaluate the false-positive rate of each analysis, we generated a reversed yeast database using a Pearl script (makeDecoyDB.pl, Bruker): the fusion of this decoy database with the original forward yeast database allowed Proteinscape to validate the data at FDR<1%. Moreover, redundancy was taken into account by Proteinscape (grouping of proteins sharing exactly the same set of peptides) and a manual inspection of the MS/MS fragmentation spectra was done for proteins identified with 1 peptide and a Mascot Score below 50. The same MS/MS raw data files were then submitted to a second database search algorithm, (Paragon, AB Sciex) with the following parameters: (a) database: the same Baker's Yeast taxonomy as described previously for Mascot, (b) enzyme: Trypsin, (c) Cys alkylation: Iodoacetamide, (d) ID focus: biological modifications, (e) search effort: thourough ID, (f) confidence p-value: 0.05, (g) Instrument: Triple-TOF 5600.

To facilitate the biological interpretation of the protein identification lists, the PANTHER classification system (http://pantherdb.org/) has been applied in a first instance on both nanoLC-MS/MS and CESI-MS/MS datasets. A second classification system named iLoc-Euk (http://www.jci-bioinfo.cn/iLoc-Euk) was then used to assess the subcellular localization of each identified protein. Combining both prediction softwares during the final data mining step allowed to strengthen the confidence related to each protein biological significance. Parameters used for both systems are detailed in the Supplemental Information section.

Results

Peptide metrics to highlight CESI-MS/MS and nanoLC-MS/MS approaches.

Consecutively to the development of the ESI ionization source and its application to study biological analytes in MS, technical developments have been achieved in order to

231 perform a separation prior to the MS analysis. Performing a separation before the ionization

232 process allows to analyze samples with an increased complexity and tends to enhance the

233 sensitivity of the signal by limiting the competition effect during ESI ionization [36]. Interfaces

234 allowing CE-MS hyphenation were therefore developed as CE is well suited for the separation

235 of biological analytes such as peptides or proteins while providing high efficiency separation

236 due to the absence of stationary phase. In this work, a recently introduced sheathless CE-MS

237 interface (CESI) has been used. A detailed description of the system has been provided in the

238 literature by Haselberg et al.[13]. Recently, following the improvement of shotgun proteomics

239 with recent mass spectrometer, different research groups have described shotgun proteomic

240 experiments performed by sheathless CE-MS using commercially available or custom-made

241 interfaces [37,38]. One of the major goals of this study was in a first instance to identify a great

242 number of proteins from a relatively low amount of yeast mitochondrial extract. For that

243 purpose, we used either the state-of-art shotgun proteomic configuration with a RP-nanoLC-

244 MS/MS interface, or the promising CESI-MS/MS interface [39-41]. ESI parameters for NanoLC

245 and CESI systems are defined in order to be compatible with nanoESI flowrate (300 nL/min

246 for nanoLC and <100 nL/min for CESI) and totally comparable in terms of performance (See

247 Experimental section). 100ng of material, a quantity which can be typically obtained after a

248 sub-cellular fractionation process on a wide variety of organisms, were injected on both

249 couplings. When considering studies published during the last 5 years, we can notice that

250 starting sample amounts ranging from 1ng to 400ng were often considered, with an average at

251 100ng (Supplemental Table S1). Moreover, as stated by Sun et al in their review published in

252 2014 (Sun et al., 2014), the vast majority of proteomic studies using a nanoLC-MS/MS

253 approach are typically starting with biological samples quantities in the milligram or

254 microgram range. That's why CESI-MS/MS approach appears as a good alternative when the

255 starting material is limited and falling into the midnanogram range.

256   To obtain a simple picture from the two complex MS datasets generated, heat maps were

257   reconstructed using PeakView software. The mass/charge ratio of the peptides (Da) is plotted

258   against the LC elution time or the CZE migration time (Figures 1-A and 1-B). A nanoLC-

259   MS/MS heat map, like the one we obtained for the injection of 100ng of a yeast mitochondrial

260   extract, is generally characterized by a wide cloud of dots, each of them representing a unique

261   MS/MS spectrum related to the fragmentation of a given precursor ion (Figure 1-A). This

262   cloud of dots means that LC conditions are regularly eluting peptides from the reverse-phase

263   column, without introducing any additional phenomenon except the fact that hydrophobic

264   peptides will elute latter than hydrophilic peptides. In contrast, the heat map obtained with the

265   CESI-MS/MS interface displays a very different picture: distinct and successive lines or

266   "strikes" of analyzed peptides can be drawn on the graph (Figure 1-B). This pattern can be

267   explained by the CZE migration process itself as it depends not only on the charge in solution

268   of the analytes but also on their mass-to-charge ratios.
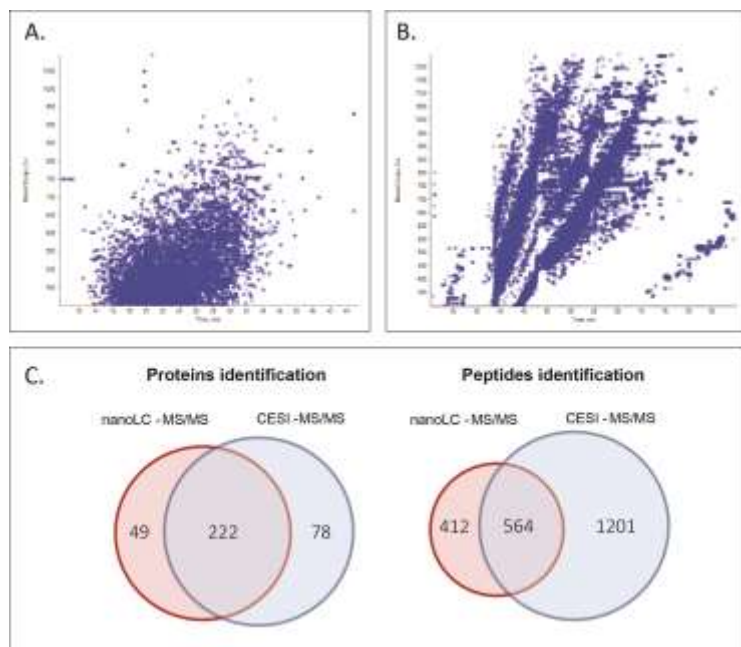
269

270   **Figure 1**. MS data obtained on 100ng of yeast mitochondrial extract. (A) NanoLC-MS/MS
271   injection: heat map, where the mass-to-charge ratio of the peptides (Da) is plotted against the
272   LC elution time (min). Each blue dot is related to a single MS/MS fragmentation spectrum.
273   (B) CESI-MS/MS injection: heat map, where the mass-to-charge ratio of the peptides (Da) is
274   plotted against the CE migration time (min). (C) Number of proteins and peptides identified in
275   either the nanoLC-MS/MS (red) or the CESI-MS/MS (blue) injection.

276   A yeast UniProtKB database was used to identify peptides and proteins and a maximum

277   false-positive rate of 1% was applied to validate both MS datasets. As shown in the Venn

278   diagram (Figure 1-C), 222 proteins were identified with either CESI-MS/MS or nanoLC-

279   MS/MS analysis. In addition, 78 proteins were only identified by CESI-MS/MS while 49

280   other proteins were only identified by nanoLC-MS/MS. Proteins identified by both

281   approaches represent the majority of the proteins identified in the two nanoLC-MSMS and

282   CESI-MSMS analyses, 82% and 74% respectively. This complementarity observed between

283   both couplings also agrees with previous reports [37,42]. Interestingly, CESI-MS/MS coupling

284   allowed to identify 3 times more additional peptides compared to the classical nanoLC-

285   MS/MS configuration (412 versus 1201 peptides), providing already a positive insight into

286   the power of CESI-MS/MS for obtaining good sequence coverages. It must be emphasized

287   that the number of identified proteins for each approach was obtained with a single injection

288   and is not the sum of multiple technical repeats. Among the 222 proteins identified commonly

289   by both CESI-MS/MS and nanoLC-MS/MS, half of them are characterized by shared

290   peptides, as well as additional peptides identified specifically by CE or LC (scenario #2 =

291   EQUIV in Supplemental Figure S1): this scenario leads all the same to an equivalent result

292   which is the identification of a given protein in the biological sample. Interestingly, this

293   observation means that the two orthogonal methods of separation used in this study were

294   identifying peptides bearing the same biophysical properties (564 common peptides,

295   representing 26% of the whole dataset) but that there are more additional peptides only

296   identified by the CE approach (1201) compared to the LC approach (412). This observation

297   is also concordant with the fact that 41 proteins identified by LC are also identified by CE but

298   with a higher number of peptides (scenario #3 = CE+ in Supplemental Figure S1). To note,

299   25% of the 222 common proteins are characterized by a completely different set of peptides

300   (scenario #5 = DIFF in Supplemental Figure S1). This latest observation is nicely showing the

301 usefulness of combining orthogonal separative methods before the mass spectrometry
302 analysis.
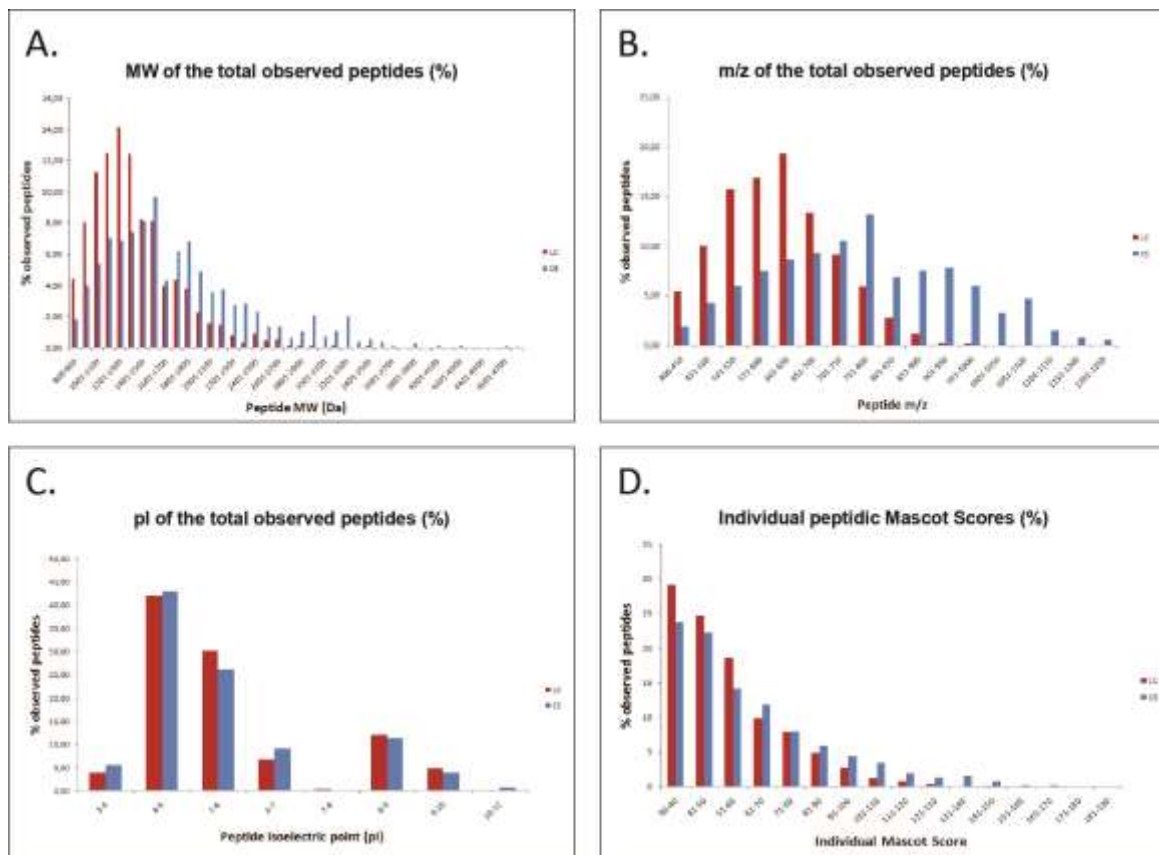
Figure 2. Comparison of peptides molecular weight (A), mass-to-charge (*m/z*) ratio (B), pI
(C) and Mascot score (D) by means of CESI-MS/MS (red) and nanoLC-MS/MS (blue). Data
are obtained from the analysis of 100ng of a mitochondrial yeast tryptic digest on both types
of coupling.

309   To evaluate each type of coupling for identification of peptides bearing specific biophysical

310   properties, we next represented the percentage of observed peptides obtained by each

311   separation against their molecular weights (MW, Figure 2-A), mass-to-charge ratios (*m/z*,

312   Figure 2-B), isoelectric points (pI, Figure 2-C) and Mascot scores (Figure 2-D). The most

313   evident trend arises from the molecular weight distribution pattern: very large peptides were

314   mostly detected with CESI-MS/MS rather than nanoLC-MS/MS. Indeed, only 10 peptides

315   observed by nanoLC-MS/MS have a molecular weight above 2800 Da, whereas this number

316   is reaching 466 peptides in the CESI-MS/MS injection: this might be explained by their

317   irreversible adsorption to either the preconcentration or the chromatographic column, or the

318   difficulty to correctly elute these large peptides. The detection of large peptides by CESI-

319   MS/MS is also a phenomenon visible in the medium mass range: the apexes of both MW

320   distributions are separated by 400 Da, with the CESI-MS/MS data displaying the highest apex

321   and the wider distribution. For example, in the medium mass range from 751 to 1250 Da, 52%

322   of peptides were detected with CESI-MS/MS, *versus* 10% with nanoLC-MS/MS. On the

323   contrary, in the low mass range from 400 to 750 Da, 90% of peptides were detected with

324   nanoLC-MS/MS, *versus* 48% with CESI-MS/MS (Figure 2-A). The same results are obtained

325   when we look at mass-to-charge ratios of the peptides, which are more widely used instead of

326   the global molecular weight (Figure 2-B). In this study, no enrichment towards low MW

327   peptides was observed in the CESI-MS/MS dataset. Interestingly, some other groups also

328   highlighted, on another biological system, the fact that CESI-MS/MS allows to elute peptides

329   of low MW [18]. These peptides are often composed of only a few amino acids, thereby giving

330   them a hydrophilic property. This type of peptides is also more difficult to observe with

331   nanoLC-MS/MS, as they are weakly bound or even unretained on reverse-phase stationary

332   phases.

333   Another important feature to monitor is the pI value from the identified peptides. Here, we

334   used the « Compute MW/pI » tool from Expasy to obtain the isoelectric point values for all

335   the observed peptides, resulting in a bi-modal distribution with unevenly distributed peptidic

336   pI values across the pH scale, with a gap at pH 7-8, and the majority of peptides clustering at

337   pH 4-6 (Figure 2-C). It can be noted that the pI distribution obtained on this *S. cerevisiae*

338   mitochondrial extract is in agreement with previously published data like theoretical 2D-gels

339   reconstructed by Knight et al. [43] on a wide variety of organisms including the total yeast

340   proteome as well as sub-cellular yeast fractions. Results on pI distribution obtained in this

341   study on a mitochondrial yeast extract are relatively similar. Slight differences can be

342   observed all the same for the extreme pI values, especially extremely basic peptides (pI>10)

343   or peptides with more acidic property (pI<4), as already shown by other research groups [37].

344   This phenomenon is potentially a result of the decreased ion suppression at very low flow

345   rates and could be of particular interest for phosphopeptides detection without any enrichment

346   techniques [42].
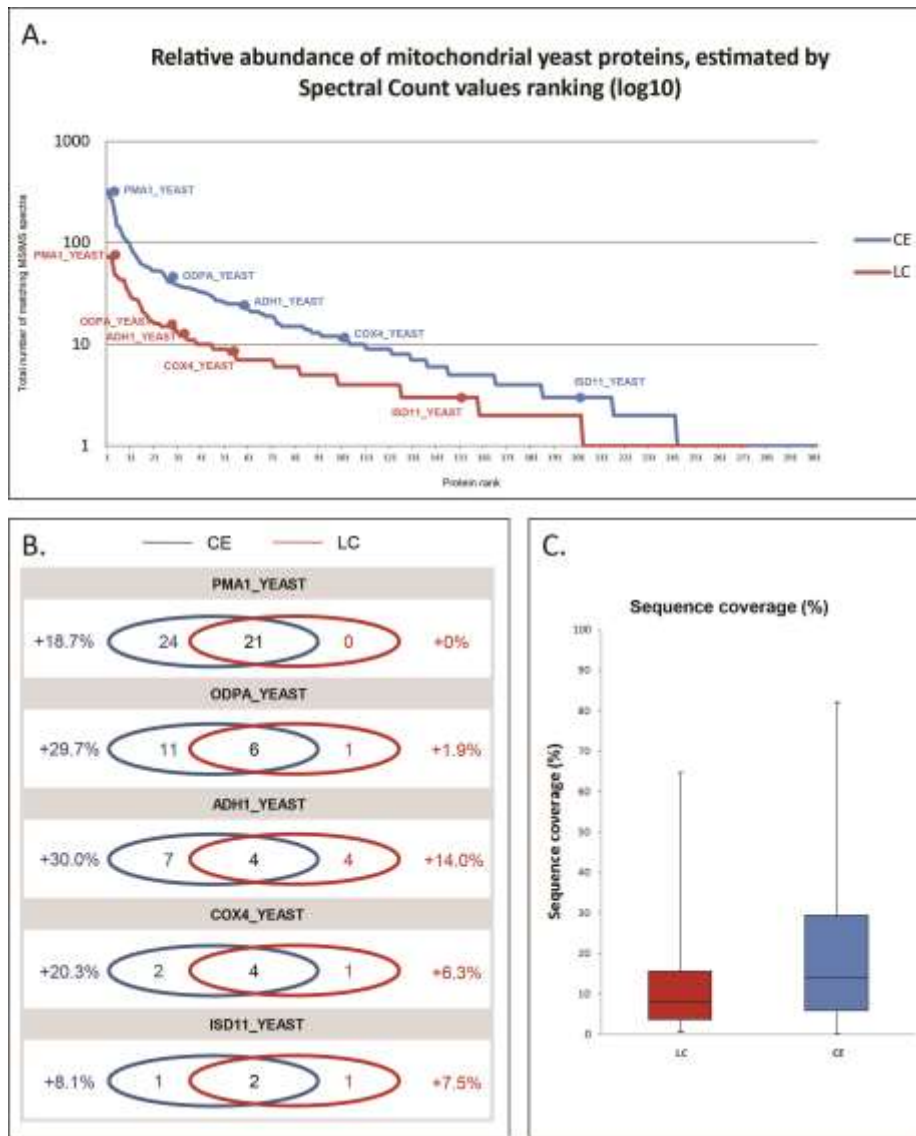
348   **Figure 3.** Effect on the global protein coverage when comparing the CESI-MS/MS (blue) and
349   the nanoLC-MS/MS (red) datasets. (A) Proteins are ranked according to the total number of
350   MS/MS spectra matching to each sequence (log scale) and 5 proteins, distributed all along the
351   dynamic range, were further investigated. (B) Comparison of the sequence coverage from the
352   5 previously chosen proteins: shared and specific peptides are indicated in each Venn
353   diagram, as well as the total sequence coverage increase (in %). (C) Boxplot showing the
354   sequence coverage distributions for both datasets when considering all the identified proteins.

Discussions

Increasing sequence coverage versus spectrum quality.

Maximal sequence coverage is recommended to increase protein identification confidence but coverage should not be won to the detriment of the spectral quality. As this parameter can be evaluated with the Mascot Score obtained for each spectrum, we next represented the Mascot Score distributions obtained on the CESI-MS/MS and nanoLC-MS/MS injections of 100ng mitochondrial extract (Figure 2-D). Interestingly, there are more peptides with a score lower than 61 identified by the nanoLC-MS/MS approach. The CESI-MS/MS coupling provides more peptides with a score upper than 61: these high quality MS/MS spectra are representing 39.8% of the whole CESI dataset and 27.7% of the whole LC dataset. As for the median Mascot Score, it shows a slight increase from 48.3 to 52.8, for the LC or CE separations respectively (Supplemental Figure S4-A). A wider distribution is also observed for the CESI-MS/MS dataset, and this observation is more noticeable when considering the sub-set of peptides composed of the LC- and CE-specific features (Supplemental Figure S4-A). This Mascot Scores comparison is thus reflecting the overall good spectral quality obtained for both types of coupling.

As Wang et al [39] did previously while using the Sequest XCorr, we decided to use the Mascot score, which is probably the most widely-used database search algorithm by the proteomic community. Moreover, to strengthen these data, we also used a second dataset search algorithm, namely Paragon (ProteinPilot, AB Sciex): relevant observations can thus be conducted by combining 2 different algorithms, as recommended by proteomics guidelines. Paragon algorithm returns the percentage of MS/MS spectra above a given confidence threshold: we can observe that there are always more fragmentation spectra above this threshold in the CESI-MS/MS dataset compared to the nanoLC-MS/MS dataset, whether the confidence threshold is set (up to 99%, Supplemental Figure S4-C).
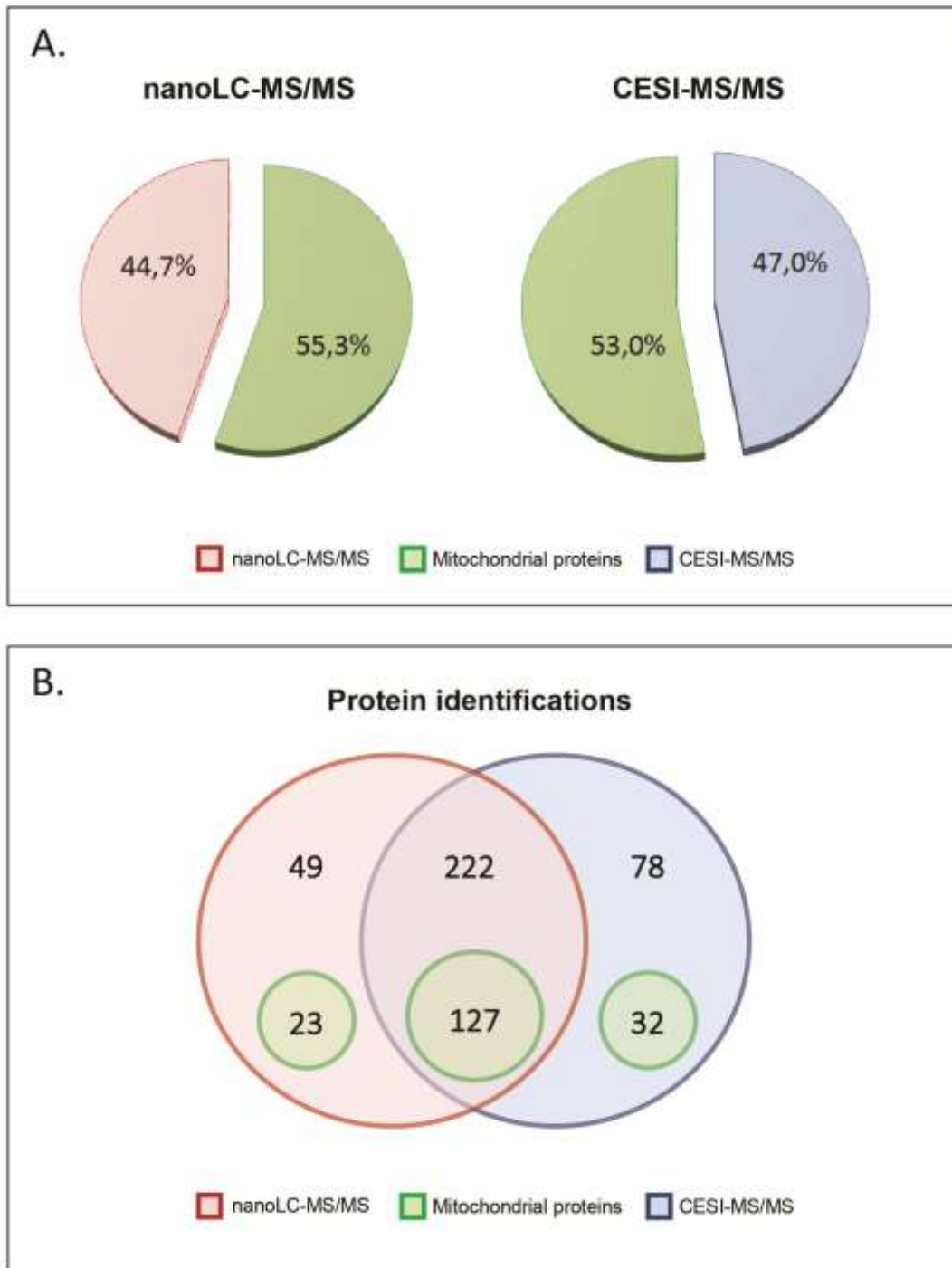
Figure 4. Venn diagrams indicating the distribution of mitochondrial proteins identified with Panther and iLoc-Euk databases using nanoLC-MS/MS and CESI-MS/MS approaches. Distribution of proteins between mitochondrion and other sub-cellular compartments using the nanoLC-MS/MS dataset (A) or the CESI-MS/MS dataset (B). (C) displays the overlap between the global protein dataset and the mitochondrial proteins using the two approaches.

There is a particular interest in the development of alternative technologies that could improve sequence coverages, to secure more confident identifications, especially for low-

390   abundant proteins often characterized by a single or a few peptides. To study the usefulness of

391   CESI-MS/MS to increase sequence coverage, we first ranked the identified proteins according

392   to the total number of MS/MS fragmentation spectra matching to each of them. While plotting

393   the resulting Spectral Count value as a function of the protein rank, we were able to select a

394   panel of 5 different proteins which were homogeneously distributed all along the dynamic

395   range from both datasets (Figure 3-A). When specifically inspecting the peptides distribution

396   from this sub-set of proteins, we can observe that sequence coverage is indeed increased in

397   the CESI-MS/MS analysis compared to the nanoLC-MS/MS analysis (Figure 3-B). While the

398   effect on the sequence coverage increase is drastically observable on high-abundant proteins

399   like PMA1, ODPA and ADH1, a more reasonable impact is observed for mid- and low-

400   abundant proteins like COX4 and ISD11. We further considered the whole set of proteins and

401   thus represented the distribution of the global protein sequence coverage depending on the

402   type of coupling (Figure 3-C) and observed the same trend. The example of the yeast plasma

403   membrane ATPase PMA1 can illustrate the sequence coverage increase that was observed in

404   this yeast mitochondrial dataset (Supplemental Figure S3-A). 24.8% sequence coverage was

405   reported for PMA1 by nanoLC-MS/MS with 21 identified peptides, while sequence coverage

406   from CESI-MSMS was 43.5% with 45 identified peptides (Supplemental Figure S3-A).

407   Interestingly, the 21 peptides identified by nanoLC-MS/MS were all identified by CESI-

408   MS/MS (annotated in red on Supplemental Figure S3-A). Moreover, among the additional 24

409   peptides identified only by CESI-MS/MS, it can be noted that 7 of them have a large mass-to-

410   charge ratio ($m/z > 1000$, highlighted in yellow on Supplemental Figure S3-A): these 5 unique

411   peptides at [175-215], [216-252], [386-414], [482-508] and [584-615] are already

412   representing 17.7% of the whole PMA1 sequence (17678 Da over 99619 Da for the full

413   length protein). For the 21 peptides identified by both separative techniques, 17 of them have

414   a CE Mascot Score higher than the corresponding LC Mascot Score (highlighted in grey in

415   Supplemental Figure S3-A). To note, the MS/MS fragmentation spectra of the 5 largest

416   peptides are of good quality as suggested by their average Mascot Score (84.6, Supplemental

417   Figure S4-B) and have acidic pI values (average pI = 4.25). Such very large peptides are

418   heavily contributing to increase the total sequence coverage. Interestingly, as suggested by

419   other works, CESI-MS/MS is also able to identify small peptides that aren't by nanoLC-

420   MS/MS. For example, the C-terminal part from PMA1 was covered in the CESI-MS/MS

421   analysis with the doubly-charged peptide [910-918] being sequenced at m/z=529.7611, as

422   well as 4 other small peptides (highlighted in grey on Supplemental Figure S3-A, peptides

423   [272-278], [380-385], [429-435] and [436-442]). Similarly, when investigating the 4 other

424   proteins distributed all along the dynamic range from the yeast mitochondrial sample (Figure

425   3-A), the same sequence coverage increase is observed (Supplemental Figures S3-B to S3-E).

426   ### *S.cerevisiae* mitochondrial proteins identification

427   Biological interpretation of the data is of paramount importance for assessing the

428   strength of mass spectrometric and proteomic results, and validation using orthogonal

429   techniques is often requested to publish MS data. In this study, a focus was established on the

430   yeast mitochondrial sub-proteome. Indeed, the determination of a protein's localization is

431   useful for biologists because it is linked to its cellular function [44]. It can also pinpoint some

432   molecular functions to specific organelles [45]. Thus, in order to perform a high-throughput

433   cellular localization analysis of all identified proteins, two protein classification systems were

434   used: PANTHER [46] and iLoc-Euk [47]. With the nanoLC-MS/MS approach (Figure 4-A), 55.3%

435   of proteins have been identified and clustered as mitochondrial proteins while 53% of them

436   were classified as mitochondrial proteins in the CESI-MS/MS dataset (Figure 4-B). In Figure

437   4-B, we can see that 127 mitochondrial proteins overlap between the two methods. Among

438   the 349 proteins identified with both methods, 52.1% are known mitochondrial proteins, while

439   the rest of them have been reported to be located into other subcellular compartments.

440
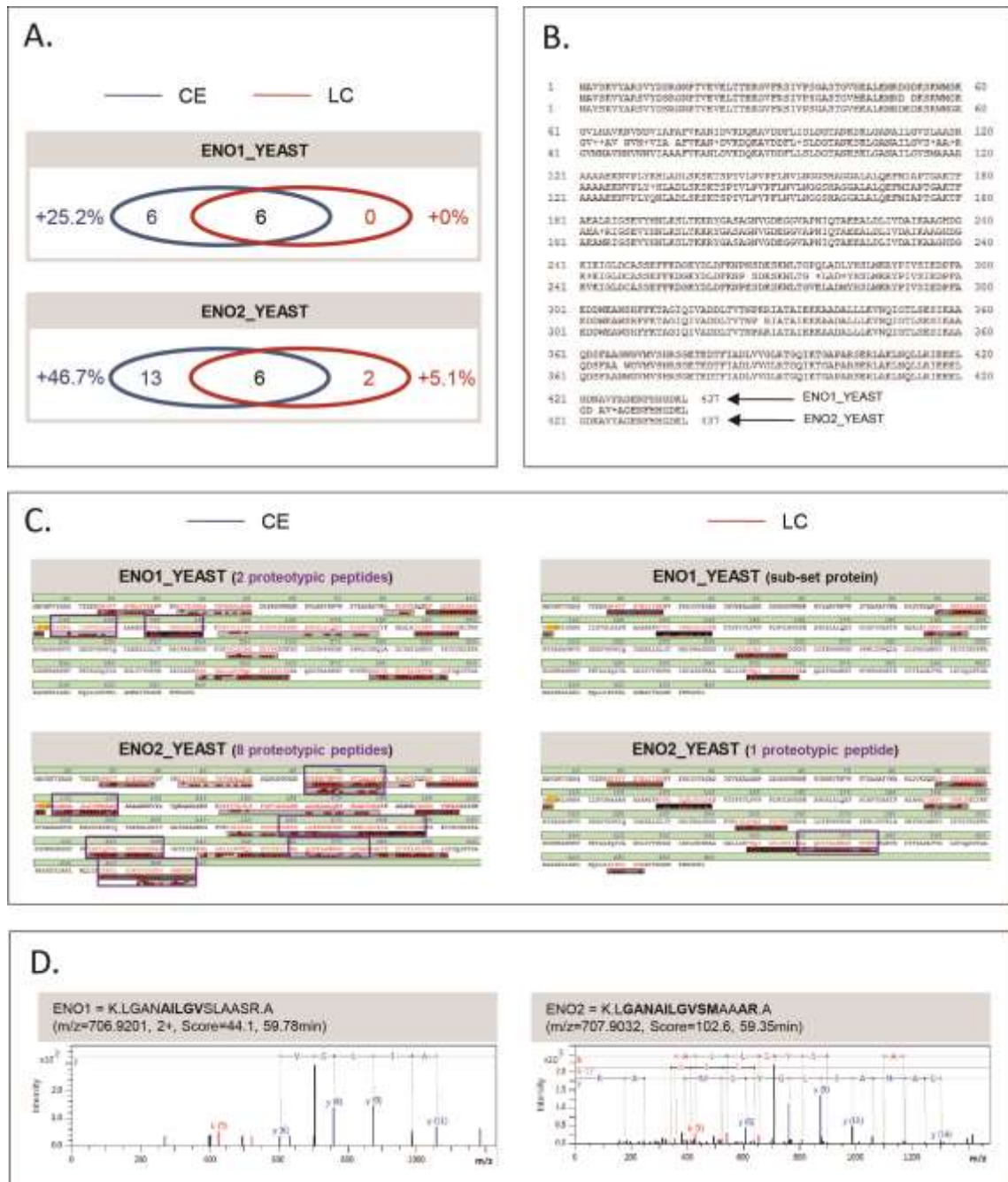
Figure 5. How is it possible to distinguish 2 protein isoforms using discriminant peptides and an increased sequence coverage: ENO1 and ENO2 proteins as an example. (A) Comparison of the sequence coverage for each isoform: shared and specific peptides are indicated in the Venn diagrams, as well as the total sequence coverage increase (in %). (B) BLAST-P alignment between both ENO isoforms. (C) Detailed sequence coverage on both isoforms obtained by either the CESI-MS/MS or the nanoLC-MS/MS analysis: identified amino acids are shown with a red label, whereas the grey box below the corresponding peptide is indicating the MS/MS spectrum quality (a red box is attributed under each amino acid if it was seen with either a b- or a y-ion, represented in the upper and lower section of the grey box respectively). Proteotypic peptides allowing the distinction between both isoforms in each dataset are indicated with purple boxes. (D) Focus on 2 proteotypic peptides matching on ENO1 and ENO2 sequences: a difference of 2 amino acids in the composition of a peptide is easily distinguishable when inspecting the MS/MS fragmentation spectra.

454

Identifying 100% of the proteins of an entire organelle from only one proteomic experiment remains a big challenge. Some proteins may escape from detection because of the technology used to separate the analytes upstream from the mass spectrometer itself: this effect is often emphasized when analytes are of extreme biophysical properties (like hydrophilic or very small peptides). To evaluate the pertinence of our proteomic dataset compared to the set of mitochondrial proteins already seen by coworkers, we selected 16 studies dealing with yeast mitochondria published between 1997 and 2014 [23,33,35,48-60]. Accession numbers were homogeneously converted and aligned: many proteins thus appeared to be well-characterized by the community as seen by the frequency of observation, whereas a sub-set of 23 proteins was only identified in our study. Among them, an intriguing protein named ENO1 was further investigated because it was only identified in the CESI-MS/MS dataset with 41 MS/MS spectra. Enolases are essential glycolytic enzymes that catalyse the interconversion of 2-phosphoglycerate to phosphoenolpyruvate. They are present under two isoforms, named A and B, in *Saccharmomyces ceravisia*e [61]. It is very important to describe a biological sample as exhaustively as possible, thus including isoforms characterization by mass spectrometry: this is possible by the identification of discriminant proteotypic peptides matching on each isoform sequence. Figure 5-A displays the number of distinct peptides identified for each enolase isoform using either the CE or LC approach: isoform B (ENO2) was covered by a higher number of peptides even if the 2 proteins have slightly the same molecular weight. 95% of identity is achieved on 495 residues when aligning the two protein sequences with Blast-P (98% of positive matches, Figure 5-B), rendering these isoforms difficult to distinguish. While investigating the detailed sequences of the matching peptides using the Sequence viewer from Proteinscape software, we can conclude that ENO1 (isoform A) was not validated in the LC dataset because its 6 peptides were an exact sub-set from the 8

479   peptides matching on the ENO2 sequence (ENO1 is a so-called sub-set protein). Whereas in

480   the CE dataset, 2 proteotypic peptides were identified for the ENO1 sequence and 8

481   proteotypic peptides were specifically matching on the ENO2 sequence (Figure 5-C). To note,

482   MS/MS fragmentation spectra of good quality were obtained for these proteotypic peptides

483   (Figure 5-D) and allow to distinguish two peptide sequences with close amino acids

484   composition. Likewise, the CE results also allow to observe two yeast glyceraldehyde-3-

485   phosphate dehydrogenase isoforms (G3P2 and G3P3). The CESI-MS/MS injection was

486   finally the only analysis in which we were able to distinguish the two yeast enolase isoforms

487   with a good confidence. Moreover, we also clustered the protein identifications in terms of

488   protein families, to see if specific families were identified by one of the two approaches. The

489   CESI-MS/MS dataset was seen to increase the number of members from four protein families,

490   related to mitochondrial protein import (TIM), ergosterol biosynthetic process (ERG),

491   mitochondrial electron transport (QCR) and ATP synthesis (ATP) (Supplemental Figure S5-

492   A). Proteins belonging to the TIM, QCR and ATP families are located in the mitochondrial

493   inner membrane whereas proteins from the QCR family are related to cell membrane, where

494   sterols are targeted. Given the importance of membrane proteins in various cellular processes,

495   as well as their roles in diseases, it is important that this class of proteins be better studied.

496   But identifying and characterizing proteins embedded in membranes still remains a challenge

497   in proteomics due to difficulties during the solubilization step. With the complementary

498   identifications obtained by combining orthogonal separative techniques, a better chance is

499   given to membrane proteins and therefore hydrophobic related peptides to be detected and

500   identified by MS. Exploring the 2 datasets with the same type of clusterization also allows to

501   identify protein families that are covered by CESI-MS/MS and nanoLC-MS/MS in an

502   equivalent manner, with members specifically identified by only 1 of the 2 techniques: this is

503   the case for the 40S and 60S ribosomal proteins family (Supplemental Figure S5-B).

504

505

506   **Conclusion**

507   To compare correctly datasets and to assess an eventual complementarity between CESI-

508   MS/MS and nanoLC-MS/MS techniques, we used a large panel of peptide properties: MW,

509   m/z, pI, Mascot Score, % peptides per protein and % sequence coverage. Based on the results

510   presented in this study, when a low quantity of yeast mitochondrial extract is analyzed, CESI-

511   MS/MS enables the identification of more peptides than nanoLC-MS/MS, and thus a higher

512   protein sequence coverage. This improvement is consistent with the actual proteomics

513   guidelines, as drastic procedures must be employed to assess the maximum of confidency on

514   protein identification. Moreover, one of the more visible trend is concerning the peptide

515   metric related to the MW of the observed peptides: indeed, more peptides having a molecular

516   weight above 2000 Da have been detected by CESI-MS/MS, whereas only some of them

517   falling into the same mass range have been identified by nanoLC-MS/MS. In parallel to the

518   increase of the number of large peptides, the peptide charaterizing metric concerning the pI

519   distribution was also investigated: our results indicate that more extreme pI values (pI<4 and

520   pI>10) are covered by CESI-MS/MS rather than nanoLC-MS/MS. These observations are

521   leaving the door open for the detection of peptides carrying particular biophysical properties,

522   like phosphopeptides, without the need of enrichment techniques. In this study, the evaluation

523   of the overall spectral quality from the CESI-MS/MS and nanoLC-MS/MS data was

524   performed by considering the Mascot score distributions and the Paragon confidence

525   thresholds. Results are suggesting that increasing the sequence coverage has no detrimental

526   impact on spectrum quality.

527   A challenging area of research in proteomics concerns isoforms charaterization. There are

528   several explanations for protein isoforms, among which multiple gene copies (allele

529   variation), alternative splicing, post-translational modifications (PTM) or degradation

530   products. In this study, we demonstrated that the 2 yeast enolase isoenzymes were both

531   charaterized in the CESI-MS/MS dataset. The observation of discriminant proteotypic

532   peptides is facilitated when a high number of precursors with high-quality MS/MS spectra are

533   generated. Traditional approaches are combining different proteolytic enzymes or are

534   benefiting from the integration of bottum-up proteomics with top-down and middle-down

535   approaches. A combination of orthogonal separative techniques, coupled online to the mass

536   spectrometer, also appears to be a good alternative to decipher a complex proteome.

537   The evaluation of both techniques using a large panel of peptide metrics allow us to say

538   that they possess complementary properties for peptide and protein identification in

539   mitochondria isolated from cultured *S. cerevisiae*. Moreover, the power of integrating two

540   orthogonal separative technologies and two protein classification systems offers new

541   promising opportunities to researchers working in the mitochondrial field of yeasts and other

542   organisms, and more broadly in subcellular proteomics.

543
544
545
546
547

554
555
556
557
558
559
560
561

References

1. Aebersold, R., Mann, M.; Mass spectrometry-based proteomics; *Nature,* (2003); 422: 198-207.

2. Smith, R.D., Olivares, J.A., Nguyen, N.T., Udseth, H.R.; Capillary zone electrophoresis mass-spectrometry using an electrospray ionisation interface; *Analytical Chemistry,* (1988); 60: 436-441.

3. Mann, M., Wilm, M.; Electrospray mass-spectrometry for protein characterization; *Trends in Biochemical Science,* (1995); 20: 219-224.

4. Ramautar, R., Heemskerk, A.A.M., Hensbergen, P.J., Deelder, A.M., Busnel, J.M., Mayboroda, O.A.; CE-MS for proteomics: Advances in interface development and application; *Journal of Proteomics,* (2012); 75: 3814-3828.

5. Hjerten, S.; High-performance electrophoresis - The elcectrophoretic conterpart of high-performance liquid-chromatography; *Journal of Chromatography,* (1983); 270: 1-6.

6. Jorgenson, J.W., Lukacs, K.D.; Zone electrophoresis in open-tubular glass-capillaries; *Analytical Chemistry,* (1981); 53: 1298-1302.

7. Busnel, J.M., Schoenmaker, B., Ramautar, R., Carrasco-Pancorbo, A., Ratnayake, C., Feitelson, J.S., Chapman, J.D., Deelder, A.M., Mayboroda, O.A.; High Capacity Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry: Coupling a Porous Sheathless Interface with Transient-Isotachophoresis; *Analytical Chemistry,* (2010); 82: 9476-9483.

8. Gahoual, R., Busnel, J.M., Wolff, P., Francois, Y.N., Leize-Wagner, E.; Novel sheathless CE-MS interface as an original and powerful infusion platform for nanoESI study: from intact proteins to high molecular mass noncovalent complexes; *Analytical and Bioanalytical Chemistry,* (2014); 406: 1029-1038.

9. Haselberg, R., de Jong, G.J., Somsen, G.W.; Low-Flow Sheathless Capillary Electrophoresis-Mass Spectrometry for Sensitive Glycoform Profiling of Intact Pharmaceutical Proteins; *Analytical Chemistry,* (2013); 85: 2289-2296.

10. Heemskerk, A.A.M., Busnel, J.M., Schoenmaker, B., Derks, R.J.E., Klychnikov, O., Hensbergen, P.J., Deelder, A.M., Mayboroda, O.A.; Ultra-Low Flow Electrospray Ionization-Mass Spectrometry for Improved Ionization Efficiency in Phosphoproteomics; *Analytical Chemistry,* (2012); 84: 4552-4559.

11. Geiger, M., Hogerton, A.L., Bowser, M.T.; Capillary Electrophoresis; *Analytical Chemistry,* (2012); 84: 577-596.

12. Fonslow, B.R., Yates, J.R.; Capillary electrophoresis applied to proteomic analysis; *Journal of Separation Science,* (2009); 32: 1175-1188.

13. Haselberg, R., de Jong, G.J., Somsen, G.W.; Capillary electrophoresis-mass spectrometry for the analysis of intact proteins 2007-2010; *Electrophoresis,* (2011); 32: 66-82.

14. Gahoual, R., Burr, A., Busnel, J.M., Kuhn, L., Hammann, P., Beck, A., Franccois, Y.N., Leize-Wagner, E.; Rapid and multi-level characterization of trastuzumab using sheathless capillary electrophoresis-tandem mass spectrometry; *Mabs,* (2013); 5: 479-490.

15. Wojcik, R., Dada, O.O., Sadilek, M., Dovichi, N.J.; Simplified capillary electrophoresis nanospray sheath-flow interface for high efficiency and sensitive peptide analysis; *Rapid Communications in Mass Sepctrometry,* (2010); 24: 2554-2560.

16. Zhu, G.J., Sun, L.L., Yang, P., Dovichi, N.J.; On-line amino acid-based capillary isoelectric focusing-ESI-MS/MS for protein digests analysis; *Analytica Chimica Acta,* (2012); 750: 207-211.

17. Sun, L.L., Zhu, G.J., Li, Y.H., Wojcik, R., Yang, P., Dovichi, N.J.; CZE-ESI-MS/MS system for analysis of subnanogram amounts of tryptic digests of a cellular homogenate; *Proteomics,* (2012); 12: 3013-3019.

18. Faserl, K., Sarg, B., Kremser, L., Lindner, H.H.; Optimization and Evaluation of a Sheathless Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry Platform for Peptide Analysis: Comparison to Liquid Chromatography-Electrospray Ionization Mass Spectrometry; *Analytical Chemistry,* (2011); 83: 7297-7305.

612 19. Sun, L.L., Zhu, G.J., Zhao, Y.M., Yan, X.J., Mou, S., Dovichi, N.J.; Ultrasensitive and Fast
613 Bottom-up Analysis of Femtogram Amounts of Complex Proteome Digests; *Angewandte Chemie-*
614 *Internation Edition,* (2013); 52: 13661-13664.

615 20. Sun, L., Hebert, A.S., Yan, X., Zhao, Y., Westphall, M.S., Rush, M.J.P., Zhu, G., Champion,
616 M.M., Coon, J.J., Dovichi, N.J.; Over 10 000 Peptide Identifications from the HeLa Proteome by Using
617 Single-Shot Capillary Zone Electrophoresis Combined with Tandem Mass Spectrometry; *Angewandte*
618 *Chemie-Internation Edition,* (2014); 53: 13931-13933.

619 21. Han, X., Wang, Y., Aslanian, A., Bern, M., Lavallee-Adam, M., Yates, J.R.; Sheathless Capillary
620 Electrophoresis-Tandem Mass Spectrometry for Top-Down Characterization of Pyrococcus furiosus
621 Proteins on a Proteome Scale; *Analytical Chemistry,* (2014); 86: 11006-11012.

622 22. Faserl, K., Kremser, L., Müller, M., Teis, D., Lindner, H.H.; Quantitative Proteomics Using
623 Ultralow Flow Capillary Electrophoresis–Mass Spectrometry; *Analytical Chemistrty,* (2015); 87: 4633-
624 4640.

625 23. Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H.E., Schonfisch, B.,
626 Perschil, I., Chacinska, A., Guiard, B., Rehling, P., Pfanner, N., Meisinger, C.; The proteome of
627 Saccharomyces cerevisiae mitochondria; *Proceedings of the National Academy of Sciences of the*
628 *United States of America,* (2003); 100: 13207-13212.

629 24. Muller, B., Grossniklaus, U.; Model organisms - A historical perspective; *Journal of*
630 *Proteomics,* (2010); 73: 2054-2063.

631 25. Castrillo, J.O., Oliver, S.G.; Yeast as a touchstone in post-genomic research: Strategies for
632 integrative analysis in functional genomics; *Journal of Biochemistry and Molecular Biology,* (2004);
633 37: 93-106.

634 26. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F.,
635 Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin,
636 H., Oliver, S.G.; Life with 6000 genes; *Science,* (1996); 274: 546-552.

637 27. Pereira, C.V., Moreira, A.C., Pereira, S.P., Machado, N.G., Carvalho, F.S., Sardao, V.A., Oliveira,
638 P.J.; Investigating drug-induced mitochondrial toxicity: a biosensor to increase drug savety?; *Current*
639 *Drug Safety,* (2009); 4: 34-54.

640 28. Benard, G., Rossignol, R.; Ultrastructure of the mitochondrion and its bearing on function and
641 bioenergetics; *Antioxidants & Redox Signaling,* (2008); 10: 1313-1342.

642 29. Yi, M.Q., Weaver, D., Hajnoczky, G.; Control of mitochondrial motility and distribution by the
643 calcium signal: a homeostatic circuit; *Journal of Cell Biology* (2004); 167: 661-672.

644 30. Zamsami, N., Kroemer, G.; Apoptosis-condensed matter in cell death; *Nature,* (1999); 401:
645 127-128.

646 31. Moyle, J., Mitchell, P.; Active-inactive state transitions of mithochondrial atpase molecules
647 influenced by mg2+, anion and aurovertin; *Febs Letters,* (1975); 56: 55-61.

648 32. Gaucher, S.P., Taylor, S.W., Fahy, E., Zhang, B., Warnock, D.E., Ghosh, S.S., Gibson, B.W.;
649 Expanded coverage of the human heart mitochondrial proteome using multidimensional liquid
650 chromatography coupled with tandem mass spectrometry; *Journal of Proteome Research,* (2004); 3:
651 495-505.

652 33. Prokisch, H., Scharfe, C., Camp, D.G., Xiao, W.Z., David, L., Andreoli, C., Monroe, M.E., Moore,
653 R.J., Gritsenko, M.A., Kozany, C,. Hixson, K.K., Mottaz, H.M., Zischka, H., Ueffing, M., Herman, Z.S.,
654 Davis, R.W., Meitinger, T., Oefner, P.J., Smith, R.D., Steinmetz, L.M.; Integrative analysis of the
655 mitochondrial proteome in yeast; *PLoS Biology,* (2004); 2: 795-804.

656 34. Meisinger, C., Sickmann, A., Pfanner, N.; The mitochondrial proteom: From inventory to
657 function; *Cell,* (2008); 134: 22-24.

658 35. Reinders, J., Zahedi, R.P., Pfanner, N., Meisinger, C., Sickmann, A.; The complete yeast
659 mitochondrial proteome: Multidimensional separation techniques for mitochondrial proteomics;
660 *Journal of Proteome Research,* (2006); 5: 1543-1554.

661 36. Tang, K.Q., Page, J.S., Smith, R.D.; Charge competition and the linear dynamic range of
662 detection in electrospray ionization mass spectrometry; *Journal of the American Society for Mass*
663 *Spectrometry,* (2004); 15: 1416-1423.

664   37.   Zhu, G.J., Sun, L.L., Yan, X.J., Dovichi, N.J.; Single-Shot Proteomics Using Capillary Zone
665   Electrophoresis-Electrospray Ionization-Tandem Mass Spectrometry with Production of More than 1
666   250 Escherichia coli Peptide Identifications in a 50 min Separation; *Analytical Chemistry,* (2013); 85:
667   2569-2573.
668   38.   Moini, M.; Simplifying CE-MS operation. 2. Interfacing low-flow separation techniques to
669   mass spectrometry using a porous tip; *Analytical Chemistry* (2007); 79: 4241-4246.
670   39.   Wang, Y., Fonslow, B.R., Wong, C.C.L., Nakorchevsky, A., Yates, J.R.; Improving the
671   Comprehensiveness and Sensitivity of Sheath less Capillary Electrophoresis-Tandem Mass
672   Spectrometry for Proteomic Analysis; *Analytical Chemistry* (2012); 84: 8505-8513.
673   40.   Gahoual, R., Busnel, J.M., Beck, A., François, Y.N., Leize-Wagner, E.; Full Antibody Primary
674   Structure and Microvariant Characterization in a Single Injection Using Transient Isotachophoresis
675   and Sheathless Capillary Electrophoresis–Tandem Mass Spectrometry; *Analytical Chemistry,* (2014);
676   86: 9074-9081.
677   41.   Gahoual, R., Biacchi, M., Chicher, J., Kuhn, L., Hammann, P., Beck, A., Leize-Wagner, E.,
678   Francois, Y.N.; Monoclonal antibodies biosimilarity assessment using transient isotachophoresis
679   capillary zone electrophoresis-tandem mass spectrometry; *Mabs,* (2014); 6: 1464-1473.
680   42.   Sarg, B., Faserl, K., Kremser, L., Halfinger, B., Sebastiano, R., Lindner, H.H.; Comparing and
681   Combining Capillary Electrophoresis Electrospray Ionization Mass Spectrometry and Nano-Liquid
682   Chromatography Electrospray Ionization Mass Spectrometry for the Characterization of Post-
683   translationally Modified Histones; *Molecural & Cellular Proteomics,* (2013); 12: 2640-2656.
684   43.   Knight, C.G., Kassen, R., Hebestreit, H., Rainey, P.B.; Global analysis of predicted proteomes:
685   Functional adaptation of physical properties; *Proceedings of the National Academy of Sciences of the
686   United States of America,* (2004); 101: 8390-8395.
687   44.   Dreger, M.; Subcellular proteomics *Mass Spectrometry Review,* (2003); 22: 27-56.
688   45.   L Lilley, K.S., Dupree, P.; Mant organelle proteomics; *Current Opinion in Plant Biology,* (2007);
689   10: 594-599.
690   46.   Mi, H.Y., Muruganujan, A., Thomas, P.D.; PANTHER in 2013: modeling the evolution of gene
691   function, and other gene attributes, in the context of phylogenetic trees; *Nucleic Acids Research,*
692   (2013); 41: D377-D386.
693   47.   Chou, K.C., Wu, Z.C., Xiao, X.A.; iLoc-Euk: A Multi-Label Classifier for Predicting the Subcellular
694   Localization of Singleplex and Multiplex Eukaryotic Proteins; *Plos One,* (2011); 6: 146-153.
695   48.   Renvoise, M., Bonhomme, L., Davanture, M., Valot, B., Zivy, M., Lemaire, C.; Quantitative
696   variations of the mitochondrial proteome and phosphoproteome during fermentative and
697   respiratory growth in Saccharomyces cerevisiae; *Journal of Proteomics,* (2014); 106: 140-150.
698   49.   Ben-Menachem, R., Tal, M., Shadur, T., Pines, O.; A third of the yeast mitochondrial
699   proteome is dual localized: A question of evolution; *Proteomics,* (2011); 11: 4468-4476.
700   50.   Zahedi, R.P., Sickmann, A., Boehm, A.M., Winkler, C., Zufall, N., Schonfisch, B., Guiard, B.,
701   Pfanner, N., Meisinger, C.; Proteomic analysis of the yeast mitochondrial outer membrane reveals
702   accumulation of a subclass of preproteins; *Molecular Biology of the Cell,* (2006); 17: 1436-1450.
703   51.   Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., O'Shea, E.K.;
704   Global analysis of protein localization in budding yeast; *Nature,* (2003); 425: 686-691.
705   52.   Ohlmeier, S., Kastaniotis, A.J., Hiltunen, J.K., Bergmann, U.; The yeast mitochondrial
706   proteome, a study of fermentative and respiratory growth; *Journal of Biological Chemistry,* (2004);
707   279: 3956-3979.
708   53.   Reijans, M., Lascaris, R., Groeneger, A.O., Wittenberg, A., Wesselink, E., van Oeveren, J., de
709   Wit, E., Boorsma, A., Voetdijk, B., van der Spek, H., Grivell, L.A., Simons, G.; Quantitative comparison
710   of cDNA-AFLP, microarrays, and GeneChip expression data in Saccharomyces cerevisiae; *Genomics,*
711   (2003); 82: 606-618.
712   54.   Marc, P., Margeot, A., Devaux, F., Blugeon, C., Corral-Debrinski, M., Jacq, C.; Genome-wide
713   analysis of mRNAs targeted to yeast mitochondria; *Embo Reports,* (2002); 3: 159-164.
714   55.   Snyder, M., Kumar, A.; Yeast genomics: past, present, and future promise; *Functional &
715   Integrative Genomics,* (2002); 2: 135-137.

716 56. Dimmer, K.S., Fritz, S., Fuchs, F., Messerschmitt, M., Weinbach, N., Neupert, W.,
717 Westermann, B.; Genetic basis of mitochondrial function and morphology in Saccharomyces
718 cerevisiae; *Molecular Biology of the Cell,* (2002); 13: 847-853.
719 57. Steinmetz, L.M., Scharfe, C., Deutschbauer, A.M., Mokranjac, D., Herman, Z.S., Jones, T., Chu,
720 A.M., Giaever, G., Prokisch, H., Oefner, P.J., Davis, R.W.; Systematic screen for human disease genes
721 in yeast; *Nature Genetics,* (2002); 31: 400-404.
722 58. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P.; Comparative
723 assessment of large-scale data sets of protein-protein interactions; *Nature,* (2002); 417: 399-403.
724 59. Pflieger, D., Le Caer, J.P., Lemaire, C., Bernard, B.A., Dujardin, G., Rossier, J.; Systematic
725 identification of mitochondrial proteins by LC-MS/MS; *Analytical Chemistry* (2002); 74: 2400-2406.
726 60. DeRisi, J.L., Iyer V.R., Brown, P.O.; Exploring the metabolic and genetic control of gene
727 expression on a genomic scale; *Science,* (1997); 278: 680-686.
728 61. Brandina, I., Graham, J., Lemaitre-Guillier, C., Entelis, N., Krasheninnkov, I., Sweetlove, L.,
729 Tarassov, I., Martin, R.P.; Enolase takes part in a macromolecular complex associated to
730 mitochondria in yeast; *Biochimica et Biophysica Acta,* (2006); 1757: 1217-1228.

731

732
733

734